

# **ANALISIS KELOMPOK DENGAN ALGORITMA *FUZZY CLUSTERING***

Studi Kasus Pengelompokan Desa di Provinsi Kalimantan Timur

(*CLUSTER ANALYSIS WITH FUZZY CLUSTERING ALGORITHM*  
*Case Study Grouping Villages in Kalimantan Timur Province*)

**Rudy Ramadani Syoer<sup>1</sup>, Yudi Wahyudin<sup>2</sup>**

Statistisi Ahli Madya, BPS Provinsi Kalimantan Timur<sup>1, 2</sup>

[ramadani@bps.go.id](mailto:ramadani@bps.go.id), [yudi.wahyudin@bps.go.id](mailto:yudi.wahyudin@bps.go.id)

## **ABSTRAK**

Pengelompokan desa selama ini telah dilakukan oleh Kemendes dengan Indeks Desa Membangun dan Indeks Pembangunan Desa oleh BPS dengan menggunakan data hasil Survei Potensi Desa. Dalam prosesnya, variabel-variabel ini akan dilakukan *scoring* kemudian dikelompokkan menurut jumlah *score* total masing-masing desa. Penentuan *score* dari masing-masing variabel tentu memerlukan dasar teori yang harus dapat diperbandingkan secara matematis. Pada penelitian ini, pengelompokan desa menggunakan teknik statistik multivariat di mana nilai masing-masing variabel secara langsung digunakan dalam penghitungan tanpa melalui proses *scoring*. Metode pengelompokan yang digunakan adalah *fuzzy clustering analysis* dengan Algoritma *fuzzy Fuzzy C-means* (FCM). Penelitian ini mengkaji pengelompokan yang paling optimal dari desa/kelurahan di Provinsi Kalimantan Timur, berdasarkan variabel-variabel terkait sarana/prasarana desa dan kondisi sosial ekonomi penduduknya. Berdasarkan fungsi objektif, waktu komputasi dan rasio simpangan baku kelompok yang minimum serta melalui penghitungan beberapa indeks validitas kelompok, jumlah kelompok optimal yang dapat terbentuk adalah sebanyak empat kelompok.

**Kata kunci** : Pengelompokan wilayah, *Fuzzy C-means clustering*, Analisis Faktor, Indeks Validitas Kelompok

## **ABSTRACT**

*The Ministry of Villages and BPS group villages based on the Village Building Index and Village Development Index using data from the Village Potential Survey. In the process, these variables will be scored and then grouped according to the total score of each village. Determining the score of each variable certainly requires a theoretical basis that must be compared mathematically. In this study, the grouping of villages uses a multivariate statistical technique in which the value of each variable is directly used in the calculation without going through a scoring process. The clustering method used is fuzzy clustering analysis with fuzzy C-means (FCM) algorithm. This study examines the most optimal grouping of villages in East Kalimantan Province, based on variables related to village facilities/infrastructure and socio-economic conditions of the population. Based on the objective function, computation time, minimum standard deviation ratio of the groups, and several group validity indices, the optimal number of groups that can be formed is four groups.*

**Keywords** : *Grouping Regions, Fuzzy C-means clustering, Factor Analysis, Cluster Validity Index*

## PENDAHULUAN

Analisis pengelompokan adalah salah satu analisis untuk menentukan kelompok berdasarkan kesamaan karakteristik yang digunakan untuk menyelesaikan permasalahan dalam beberapa disiplin ilmu. Maxwell dan Pryor (2014) menerapkan analisis *clustering* dalam bidang sosial untuk mengelompokkan status sosial masyarakat berdasarkan persamaan dan perbedaan data lintas budaya. Wang *et al.* (2005) meneliti tentang penemuan metode baru dalam dunia kedokteran untuk meningkatkan akurasi diagnosa. Analisis kelompok dimulai dari metode hierarki yang secara garis besar membentuk sebuah *tree* diagram dalam menentukan kelompok. Kemudian berlanjut ke metode nonhierarki dikenal dengan cara partisi, yang menentukan jumlah kelompok disesuaikan dengan tujuan penelitian. Perkembangan lebih lanjut dengan mempertimbangkan tingkat keanggotaan himpunan *fuzzy* sebagai dasar pembobotan yang memungkinkan objek untuk bergabung ke setiap kelompok yang ada disebut *fuzzy clustering* (Bezdek, 1981). Teknik paling umum digunakan secara luas adalah algoritma *Fuzzy C-means* (Dunn, 1975). Dengan teknik ini, objek akan cenderung menjadi anggota suatu kelompok yang memiliki derajat keanggotaan tertinggi terhadap kelompoknya. Perluasan untuk model *Fuzzy C-means* yaitu dengan menghitung jarak antar objek dengan pusat kelompok menggunakan rumus *Mahalanobis* (Gustafson & Kessel, 1978) dan model *Gath-Geva* dimana algoritmanya menggunakan formula jarak dengan adanya tambahan pada *fuzzy covariance* matriksnya (Gath & Geva, 1989).

Penelitian menggunakan *fuzzy clustering* pernah dilakukan sebelumnya oleh Munaf (2009) yang meneliti masalah pengelompokan kabupaten/kota di Jawa Timur sebagai dasar perencanaan pembiayaan kegiatan pendataan lapangan menggunakan algoritma *Fuzzy C-means* dengan modifikasi formula jarak. Syoer dan Mashuri (2014) melakukan pengelompokan desa/kelurahan di Kalimantan Timur untuk identifikasi daerah maju atau tertinggal memanfaatkan algoritma *Fuzzy C-means* dan *Gath-Geva clustering*.

Penelitian ini bermaksud mengelompokkan wilayah desa/kelurahan menggunakan algoritma *Fuzzy C-means clustering* di seluruh wilayah Provinsi Kalimantan Timur menggunakan data Potensi Desa (Podes) 2018 berdasarkan karakteristik desa-desa di Kalimantan Timur yang menunjukkan tingkat kemajuan atau ketertinggalan pembangunan di wilayah desa. Penelitian ini merekomendasikan jumlah kelompok yang tepat berdasarkan beberapa indeks validitas kelompok sehingga dapat mengelompokkan wilayah desa sesuai klasifikasinya. Pengelompokan desa selama ini dilakukan oleh Kementerian Desa, Pembangunan Desa Tertinggal dan Transmigrasi (Kemendes) dengan Indeks Desa Membangun (IDM) dan melalui Indeks Pembangunan Desa (IPD) oleh Badan Pusat Statistik (BPS), yang mengelompokkan desa-desa berdasarkan *scoring*, sedangkan penelitian ini menggunakan pendekatan berbeda melalui teknik statistik multivariat, yang belum pernah diteliti sebelumnya untuk kasus sejenis.

## METODE

Analisis kelompok merupakan suatu teknik statistik multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan kesamaan karakteristik yang dimilikinya (MacQueen, 1967). Dua metode paling umum dalam algoritma kelompok adalah metode hierarki dan metode nonhierarki. Analisis pengelompokan yang terdapat dalam perangkat lunak MATLAB, menggunakan jarak *Euclidean* sebagai *default* ukuran jaraknya, dimana ukuran jarak ini hanya bisa mendeteksi kelompok dengan susunan *spherical*, yang didefinisikan sebagai berikut:

$$D_{ik} = D(\mathbf{x}_k, \mathbf{v}_i) = \sum_{i=1}^c \|\mathbf{x}_k - \mathbf{v}_i\|^2 = \sqrt{\sum_{i=1}^c (\mathbf{x}_k - \mathbf{v}_i)^2} \dots\dots\dots (1)$$

Untuk mendeteksi kelompok dengan susunan selain bentuk *spherical*, dikembangkan beberapa formula jarak, antara lain Gustafson & Kessel (1978) menggunakan formula jarak *Mahalanobis* untuk bentuk *hyperellipsoidal* serta Gath-Geva (1989) yang menggunakan formula jarak *Mahalanobis* dengan menambahkan batasan pada *fuzzy covariance* matriksnya.

Tujuan *clustering* adalah untuk mempartisi himpunan data  $\mathbf{X}$  ke dalam  $c$  buah kelompok. Asumsi  $c$  yang diketahui, didasarkan pada pengetahuan sebelumnya yang sifat-sifatnya sebagai berikut:

$$u_{ik} \in 0,1, \sum_{i=1}^c u_{ik} = 1, \quad 1 \leq k \leq N, \quad 0 < \sum_{k=1}^N u_{ik} < N = 1, \quad 1 \leq i \leq c, \quad 2 \leq c \leq N, \dots\dots\dots(2)$$

Nilai derajat keanggotaan,  $u_{ik}$ , adalah 0 atau 1, data secara tegas dinyatakan sebagai anggota kelompok yang satu dan tidak menjadi anggota kelompok lainnya (contohnya *K-means*). Tetapi adakalanya objek mungkin berada dalam dua atau lebih partisi, sehingga perlu dilakukan pembobotan berdasarkan tingkat keanggotaan himpunan *fuzzy*-nya (Janos Abonyi & Szeifert, 2003). Dengan cara ini, maka algoritma pembentukan kelompok menjadi lebih fleksibel. Konsep ini disebut *fuzzy clustering* (Valente de Oliveira & Pedrycz, 2007). Berbeda dengan cara tegas, dengan cara *fuzzy* ini masing-masing data item diberikan nilai kemungkinan untuk bisa bergabung ke setiap kelompok yang ada, yang memungkinkan  $u_{ik}$  mencapai nilai riil  $\{0,1\}$ .  $N \times c$  matriks  $\mathbf{U} = [u_{ik}]$  merepresentasikan partisi *fuzzy*, yang diberikan oleh kondisi:  $u_{ik} \in [0,1], \quad 1 \leq i \leq c, \quad 1 \leq k \leq N$ .

**Algoritma Fuzzy C-means (FCM)**

Pada tahun 1965, Prof. Lotfi A. Zadeh dari Universitas California di Barkeley memperkenalkan konsep tentang himpunan *fuzzy* (Zadeh, 1965). Metode FCM mengalokasikan kembali data ke dalam kelompok memanfaatkan teori *fuzzy*, mempergunakan variabel *membership function*. FCM yang diusulkan oleh Bezdek (1981), memperkenalkan suatu variabel  $m$  yang merupakan *weighting exponent* dari *membership function*. Variabel ini dapat mengubah besaran pengaruh dari *membership function*,  $u_{ik}$ . Sedangkan  $m$  mempunyai wilayah nilai lebih besar dari 1 ( $m > 1$ ). Tidak ada ketentuan yang jelas berapa besar nilai  $m$  yang optimal dalam melakukan proses optimasi suatu permasalahan *clustering*, nilai  $m$  yang umum digunakan adalah 2 (Johnson, R.A. and Wichern, 2007). Sedangkan  $u_{ik}$ , mempunyai wilayah nilai  $0 \leq u_{ik} \leq 1$ , dengan syarat  $1 < m < \infty, \quad 1 \leq u_{ik} \leq 1, \quad \sum_{i=1}^c u_{ik} = 1$ . *Membership function* dan *objective function* formula jarak FCM dirumuskan sebagai berikut:

$$u_{ik} = \sum_{j=1}^c \left[ \frac{D(\mathbf{x}_k, \mathbf{v}_i)}{D(\mathbf{x}_k, \mathbf{v}_j)} \right]^{-2/(m-1)} \dots\dots\dots(3)$$

$$J(\mathbf{X} : \mathbf{U}, \mathbf{V}) = \sum_{i=1}^c \sum_{k=1}^N (u_{ik})^m D(\mathbf{x}_k, \mathbf{v}_i)^2 \dots\dots\dots(4)$$

Algoritma FCM *clustering* untuk pengelompokkan data secara umum dirumuskan dengan (Balazs Balasko, 2013) :

1. Menentukan jumlah kelompok ( $c$ ). Pada tahap awal  $c=2$ .
2. Menentukan inisiasi awal matriks partisi  $\mathbf{U}$  secara *random* dan menghitung *centroid* atau  $\mathbf{v}_i$  dari masing-masing kelompok.
3. Menghitung *distance measure* atau formula jarak, yang merupakan *square euclidean distance* atau jarak *euclidean* kuadrat sesuai persamaan (1).
4. Menghitung nilai fungsi keanggotaan data di masing-masing kelompok atau meng-*update* matriks partisi  $\mathbf{U}$  dengan persamaan (3) serta mengalokasikan datanya ke *centroid* terdekat dan menghitung *centroid* baru.
5. Kembali ke langkah 2, apabila perubahan nilai fungsi keanggotaan data masih di atas nilai *threshold* ( $\varepsilon$ ) yang ditentukan oleh  $|\mathbf{U}^t - \mathbf{U}^{t-1}| < \varepsilon$ , atau apabila perubahan nilai fungsi objektif yang dihitung berdasarkan persamaan (4) masih di atas nilai *threshold* yang ditentukan. Nilai *threshold* adalah suatu nilai yang sangat kecil mendekati 0, (misal 0,000001).

### Evaluasi Hasil Pengelompokkan

Untuk melihat kinerja 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 *et al.*, 1996). Formulasnya sebagai berikut:

$$S_W = \frac{1}{c} \sum_{i=1}^c S_c, S_B = \sqrt{\frac{1}{c-1} \sum_{i=1}^c (\bar{X}_c - \bar{X})^2} \dots\dots\dots (5)$$

Semakin kecil nilai  $S_W$  dan semakin besar nilai  $S_B$ , maka algoritma 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.

### Indeks Validitas Cluster

Penentuan jumlah kelompok yang optimal dalam metode pengelompokkan nonhierarki, dapat dicoba beberapa buah kelompok atau disingkat  $c$ , kemudian digunakan besaran indeks validitas kelompok untuk menilai besaran  $c$  yang terbaik atau optimal. Dalam papernya, Kim et al. (2003) menyajikan beberapa indeks validitas kelompok tradisional, diantaranya seperti yang diusulkan oleh Bezdek (1981) dan Xie and Beni (1991). Beberapa skala pengukuran validitas lainnya yang berbeda telah diusulkan dalam beberapa literatur seperti *partition index* (SC), *separation index* (S), dan indeks Dunn, walaupun tidak ada satupun yang sempurna (János Abonyi et al., 2007), antara lain: *Partition Coefficient* (PC) dan *Classification Entropy* (CE); *Partition Index* (PI), *Separation Index* (SI), Indeks Xie Beni (XB), dan Indeks Dunn.

### 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 bertujuan untuk mendapatkan sejumlah kecil faktor yang memiliki sifat berikut (Johnson, R.A. and Wichern, 2007):

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

Model analisis faktor dalam notasi matriks adalah sebagai berikut:

$$\underset{(px1)}{\mathbf{X}} - \underset{(px1)}{\boldsymbol{\mu}} = \underset{(pxm)}{\mathbf{L}} \underset{(mx1)}{\mathbf{F}} + \underset{(px1)}{\boldsymbol{\varepsilon}} \dots\dots\dots (6)$$

Dengan asumsi :  $E(\mathbf{F}) = 0, cov(\mathbf{F}) = \mathbf{I}, E(\boldsymbol{\varepsilon}) = 0, cov(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}$ ,  $\boldsymbol{\Psi}$  adalah matriks diagonal.

### FCM Toolbox

Perangkat lunak MATLAB dapat digunakan sebagai alat bantu untuk pembuatan program aplikasi berupa *toolbox* dan semacamnya menggunakan bantuan *Mathworks toolbox*. Paket program dan fungsi ini sudah terintegrasi dalam MATLAB, menggunakan bantuan *m-files* yang berguna dalam pembuatan *macro* program. Paket pemrograman ini dipilih karena fasilitas yang dimilikinya telah menyediakan suatu fungsi *fuzzy toolbox* yang sesuai dengan tujuan penelitian ini, yaitu penggunaan metode FCM. Berdasarkan *Mathworks Toolbox* yang dapat diunduh secara gratis pada laman MATLAB CENTRAL, situs resmi dari The MathWorks Inc., produsen pembuat *software* MATLAB, menyediakan *toolbox* untuk *Fuzzy*

*C-means* (FCM) termasuk pula *cluster validity*-nya. Penggunaan algoritma FCM dalam tulisan ini, didasarkan pada *Clustering and Data Analysis Toolbox* yang diterbitkan oleh János Abonyi et al. (2007).

### **Tahapan Analisis Data**

Data yang digunakan dalam penelitian ini adalah hasil pendataan Potensi Desa (PODES) Provinsi Kalimantan Timur tahun 2018. Perapihan data dilakukan dengan *software IBM Statistical Package for Social Science (SPSS) Statistics 23 for windows*. Sedangkan pengolahan data dan algoritmanya dibuat menggunakan *software MATLAB R2021a*. Variabel diidentifikasi menjadi dua faktor yang diusahakan semaksimal mungkin mewakili tingkat kemajuan dan perkembangan pembangunan desa yang dicerminkan dari indikator penyusun Indeks Pembangunan Desa (Badan Pusat Statistik, 2019) yang terdiri dari 4 Dimensi yaitu Dimensi Pelayanan Dasar, Kondisi Infrastruktur, Aksesibilitas dan Pelayanan Umum. Setelah diidentifikasi menjadi sebanyak 38 variabel. Tahapan Analisis Data, dimulai dari mengkaji hasil pengelompokkan, yaitu dengan:

- a. Melakukan perapihan data yaitu menilai perlu tidaknya melakukan standarisasi data berdasarkan deskripsi data, kemudian melakukan analisis faktor, untuk mereduksi variabel yang ada sehingga dihasilkan *factor score*, dan dipilih yang sesuai dengan kebutuhan penelitian (kasus).
- b. Menyajikan data dalam matriks berukuran  $N \times n$ .
- c. Meng-*input* data dan parameter, dari skrip MATLAB.
- d. Menggunakan metode pengelompokkan nonhierarki dengan menerapkan algoritma FCM *clustering*. Tahapannya secara rinci adalah sebagai berikut:
  1. Menentukan jumlah kelompok ( $c$ ), dimulai dari  $c=2$ .
  2. Menentukan tingkat ke-*fuzzy*-an hasil pengelompokkan, yaitu  $m=2$  (*default*).
  3. Inisiasi awal matriks  $\mathbf{U}$  secara random.
  4. Menghitung pusat kelompok dan memperbaharui matriks  $\mathbf{U}$  sesuai persamaan (1) dengan formula jarak sesuai algoritma FCM.
  5. Membandingkan keanggotaan dalam matriks  $\mathbf{U}$ , jika  $|\mathbf{U} - \mathbf{U}^{l-1}| < \varepsilon$  maka kondisinya dianggap sudah konvergen, sehingga iterasinya dihentikan. Jika tidak, maka kembali ke langkah d.4. Penelitian ini menggunakan nilai  $\varepsilon$  sebesar 0,000001.
  6. Kembali ke langkah d.1, dengan nilai  $c=3,4,5$  dan seterusnya sampai dengan  $c=12$ .Berdasarkan pada langkah d, akan didapatkan ukuran-ukuran hasil pengelompokkan seperti jumlah iterasi, nilai fungsi objektif, nilai pusat kelompok, fungsi atau derajat keanggotaan akhir, rasio  $S_W/S_B$ , dan waktu komputasi yang diperlukan beserta indeks validitas kelompoknya. Kemudian berdasarkan beberapa kriteria yang dilihat dari ukuran-ukuran hasil pengelompokkan, dapat diputuskan atau direkomendasikan jumlah kelompok optimal, dengan melihat indeks validitas *clustering*. Jumlah kelompok optimal umumnya diberikan pada saat indeks mencapai kondisi nilai minimum pertama di lembah pertama yang didapatkan.

### **PEMBAHASAN**

Sebelum melakukan analisis *cluster* dan menjalankan algoritma FCM, terlebih dahulu dilakukan uji terhadap datanya untuk memastikan bahwa metode multivariat dapat digunakan pada data, seperti kemungkinan data hilang atau *missing* dan *data outlier*, terhadap variabel yang akan digunakan, memanfaatkan perangkat lunak SPSS. Berdasarkan deskripsi data, maka datanya tidak perlu untuk distandarasi dan ditransformasi ke dalam bentuk *zscore* karena tidak terlalu bervariasi (Santosa, 2014). Dalam algoritma FCM *clustering* ini diterapkan proses normalisasi data, sehingga nilainya hanya akan berada dalam *range* 0 sampai 1.

Kemudian, setelah melalui serangkaian uji data dan dianggap layak untuk diolah dengan metode statistik multivariat, maka selanjutnya dilakukan analisis faktor yang bertujuan untuk mereduksi variabel-variabel kedalam objek penelitian sehingga bisa digunakan dalam analisis *clustering*. Selanjutnya, variabel-variabel diringkas kedalam variabel baru, yang disebut faktor, dimana faktor-

faktor tersebut mencerminkan dan mewakili variabel-variabel aslinya. Hal ini dilakukan, karena algoritma dan *toolbox clustering* yang digunakan hanya memerlukan dua variabel untuk keperluan plot yang akan menampilkan grafik hasil pengelompokkan, dan diset agar dapat memunculkan kontur dalam format dua dimensi.

Melalui analisis faktor, didapatkan matriks *Anti-Image*, yang menunjukkan nilai *Measure of Sampling Adequacy* (MSA) di atas 0,5, berarti semua variabel bisa diprediksi dan dapat dijelaskan lebih lanjut, dan berdasarkan nilai KMO sebesar 0,917 dan nilai signifikansi Barlett's test sebesar 0,000, dinyatakan bahwa semua variabel layak untuk dianalisa dan dapat diolah lebih lanjut, jumlahnya sebanyak 37 buah variabel. Sedangkan total keragaman atau varians yang dapat dijelaskan dan berdasarkan grafik *scree plot* dengan kriteria nilai eigen di atas 1, terdapat 8 komponen utama atau 8 *factor score* yang terbentuk dengan total keragaman mencapai 61,161 persen (Tabel 1). Dari 8 komponen tersebut mewakili variabel penelitian sebanyak 33 variabel, dan telah jelas masuk ke dalam faktor mana.

**Tabel 1.** Total Varians yang Dapat Dijelaskan

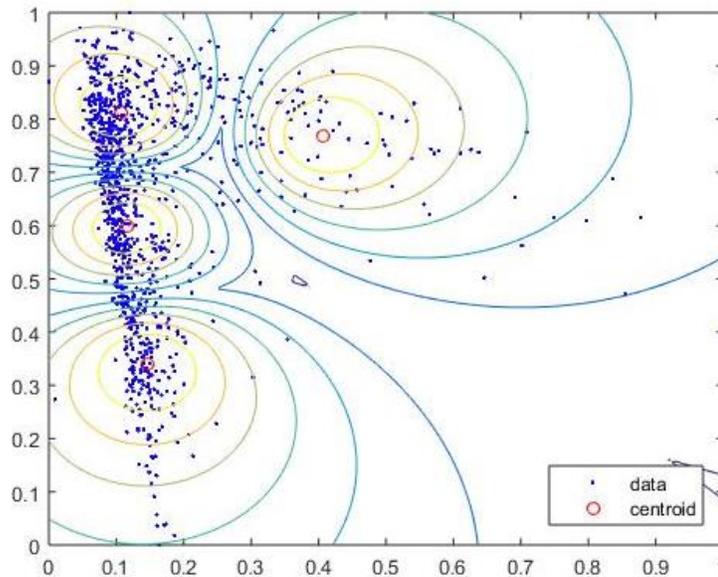
Komponen	Inisial Eigenvalues		
	Total	Persentase Varians	Persentase Kumulatif
1	10,771	29,111	29,111
2	3,763	10,170	39,280
3	1,590	4,297	43,577
4	1,552	4,196	47,773
5	1,503	4,062	51,834
6	1,253	3,387	55,221
7	1,134	3,064	58,285
8	1,064	2,876	61,161

Setelah mendapatkan 8 *factor score*, dipilih kombinasi antara *factor score* 1 dengan *factor score* 2, dikarenakan hasil kombinasi *factor score* tersebut mewakili variabel-variabel penelitian secara mayoritas. Faktor 1 dan 2, berdasarkan komponennya mewakili sebanyak 22 variabel, yaitu 12 variabel di faktor 1 dan 10 variabel di faktor 2 (Lampiran 1).

### **Proses Clustering**

Untuk melakukan analisis pengelompokkan terhadap desa/kelurahan di Provinsi Kalimantan Timur, dilakukan *running program m-files* (berupa seperangkat *toolbox clustering*) melalui perangkat lunak MATLAB

Dengan menggunakan algoritma FCM, dilakukan pengelompokkan desa/kelurahan dimulai dari jumlah kelompok sebanyak 2 sampai dengan 12, dengan inisiasi derajat keanggotaan (*fuzzifier*) sebesar 2 (secara *default*) dan batas toleransi yang sama (0,000001) untuk setiap jumlah kelompok serta maksimum iterasi sebanyak 1000 kali. Untuk jumlah kelompok atau  $c$ , dari  $c=2$  sampai dengan  $c=12$ , menunjukkan nilai fungsi objektif yang terus menurun (Lampiran 2). Waktu komputasi mengalami fluktuasi, dan cenderung memerlukan waktu yang lebih lama saat melakukan proses komputasi pada jumlah *cluster* di atas 2. Pada *cluster*=2, waktu komputasi yang dibutuhkan 1,2672 detik, sedangkan *cluster*=4 memerlukan waktu 1,3453 detik, atau 0,0781 detik lebih lama, demikian untuk jumlah kelompok=5 sampai 12, menunjukkan waktu komputasi yang cenderung lebih lama.



**Gambar 1.** Plot Hasil Pengelompokkan dengan Algoritma FCM untuk  $c=4$

Dengan melihat kesamaan di dalam kelompok dan perbedaan dengan kelompok yang lain, maka rasio  $S_W/S_B$  yang kecil akan menunjukkan ketepatan hasil pengelompokkan dan kinerja algoritma metode pengelompokkan yang lebih baik, yang dinyatakan dalam  $S_W$  yang minimum dan  $S_B$  yang maksimum. Berdasarkan kriteria ini, rasio  $S_W/S_B$  yang diperoleh terlihat cukup kecil pada semua jumlah kelompok. Pada saat  $c=2$ , nilai rasio  $S_W/S_B$  sebesar 0,4240, dan terus mengecil untuk nilai  $c$  diatas 2, nilai rasio  $S_W/S_B$  hanya sebesar 0,1944 untuk  $c=12$ .

***Rekomendasi Jumlah Cluster Optimal***

Penentuan jumlah kelompok yang terbaik atau optimal, dilakukan setelah proses *clustering* selesai dan menghasilkan beberapa indeks validitas. Jumlah kelompok optimal umumnya ditunjukkan pada saat indeks validitas kelompok mencapai kondisi nilai minimum pertama di lembah pertama yang didapatkan. Nilai CE yang minimal sebesar 0,3036, yaitu pada saat jumlah kelompok sebanyak 2. Untuk PI, nilai minimal pertama di lembah pada saat jumlah kelompok sebanyak 11, sedangkan untuk SI, nilai minimal pertama di lembah pertama sama halnya dengan CE yaitu berada pada jumlah kelompok 2. Adapun untuk indeks Xie Beni (XB) dan indeks Dunn (DI), nilai indeks validitas minimal pertama di lembah pertama sama-sama berada pada jumlah kelompok 4 yaitu sebesar 8,2507 dan 0,0025.

Untuk menentukan berapa jumlah kelompok, referensi dari beberapa indeks validitas bisa dijadikan acuan, termasuk juga dapat menyesuaikan dengan keperluan dan tujuan penelitian (Zadeh, 1965), dimana tujuannya adalah untuk mengelompokkan desa-desa sesuai dengan dimensi kemajuan/ketertinggalan desa yang ditinjau dari tingkat kemajuan dan perkembangan pembangunan desa sehingga dapat mewakili 4 dimensi Indeks Pembangunan Desa. Berdasarkan hasil dari beberapa indeks validitas kelompok tersebut dan dengan mengevaluasi hasil analisis kelompoknya secara kuantitatif, didapatkan jumlah kelompok yang optimum adalah sebanyak 4 kelompok (Lampiran 3).

***Interpretasi Hasil Pengelompokkan***

Berdasarkan pembahasan sebelumnya, jumlah kelompok yang dirasa cukup untuk menjelaskan keragaman dan karakteristik kelompok data, dan dianggap ideal untuk menggolongkan karakteristik desa di Provinsi Kalimantan Timur adalah sebanyak 4. Kelompok atau *cluster* 1 terdiri dari 91 desa, kelompok 2 terdiri dari 317 desa, kelompok 3 terdiri dari 498 desa, dan kelompok 4 terdiri dari 132 desa. Karakteristik kelompok dapat digambarkan melalui pusat kelompok dan rata-rata kelompoknya (Tabel 2).

**Tabel 2** Karakteristik Kelompok berdasarkan Pusat Kelompok

Variabel	Pusat Kelompok			
	kelompok 1	kelompok 2	kelompok 3	kelompok 4
<i>factor score</i> 1	0,4069	0,1067	0,1170	0,1457
<i>factor score</i> 2	0,7669	0,8133	0,5978	0,3392
Rata-rata	0.5869	0.4600	0.3574	0.2425

Berdasarkan karakteristik kelompoknya maka rata-rata *factor score* dapat diurutkan sesuai *ranking* kelompoknya dari kelompok tertinggi ke *ranking* terendah, yang dihitung menggunakan fungsi RANK(Number, Ref, Order), dan setelah diurutkan maka kelompok 1 adalah yang paling tinggi kelompoknya sedangkan kelompok 4 merupakan kelompok yang paling rendah *ranking*-nya. Susunan *ranking* kelompok berdasarkan pusat kelompok atau *centroid* sesuai persamaan (7) sebagai berikut:

$$centroid(v_i) = \begin{bmatrix} 0,4069 & 0,7669 \\ 0,1067 & 0,8133 \\ 0,1170 & 0,5978 \\ 0,1457 & 0,3392 \end{bmatrix} \rightarrow \begin{matrix} \text{kelompok 1} \rightarrow \text{ranking 1} \\ \text{kelompok 2} \rightarrow \text{ranking 2} \\ \text{kelompok 3} \rightarrow \text{ranking 3} \\ \text{kelompok 4} \rightarrow \text{ranking 4} \end{matrix} \dots\dots\dots(7)$$

Berdasarkan hasil *clustering ranking* kelompok tersebut, maka interpretasi dari tiap-tiap kelompok-kelompok yang ada, sebagai berikut:

- Kelompok 1 : Kelompok ini terdiri dari desa-desa yang paling maju ditinjau dari ke-4 dimensi Indeks Pembangunan Desa yang meliputi dimensi Pelayanan Dasar, Kondisi Infrastruktur, Aksesibilitas dan Pelayanan Umum di desa. Umumnya desa-desa yang ada dalam kelompok ini merupakan kategori daerah perkotaan. Kelompok ini memiliki nilai maksimum pada variabel pelayanan pendidikan, pelayanan kesehatan, infrastuktur ekonomi dan sarana serta aksesibilitas transportasi. Artinya, kelompok ini terdiri dari desa-desa yang paling maju dibanding 3 kelompok desa lainnya. Desa-desa dalam kelompok ini sebagian besarnya masuk dalam kategori Desa Mandiri (Badan Pusat Statistik, 2019).
- Kelompok 2 : Kelompok ini terdiri dari desa-desa yang cukup maju ditinjau dari dimensi Pelayanan Dasar dan Kondisi Infrastruktur desa, termasuk Aksesibilitas sosial ekonomi penduduknya. Kemajuan desa-desa yang ada dalam kelompok ini hampir menyamai kelompok 1, tetapi berdasarkan karakteristiknya masih dibawah kelompok tersebut.
- Kelompok 3 : adalah kelompok desa-desa yang kurang maju, ditinjau dari beberapa aspek pelayanan di desa tersebut, termasuk keadaan infrastuktur dan sarana serta aksesibilitas transportasi sosial ekonomi penduduknya, yang dicerminkan oleh cukup rendahnya keberadaan fasilitas pendidikan dan pelayanan dasar masyarakat. Kelompok ini dan kelompok 4, umumnya bisa digolongkan kepada desa-desa yang berkembang, merujuk pada Indeks Pembangunan Desa.
- Kelompok 4 : Kelompok ini terdiri dari desa-desa yang yang memiliki ciri-ciri daerah yang paling tertinggal dibanding kelompok lainnya. Desa-desa dalam kelompok ini memiliki nilai minimum pada hampir semua variabel. Artinya, kelompok ini terdiri dari desa-desa yang paling kurang maju (tertinggal) dibanding 3 kelompok desa lainnya.

**KESIMPULAN**

Dengan menggunakan kriteria hasil pengelompokan yaitu nilai fungsi objektif, waktu komputasi dan rasio simpangan baku minimum menggunakan algoritma FCM *clustering*, serta berdasarkan beberapa indeks validitas kelompok, maka dapat disimpulkan jumlah kelompok paling optimal untuk mengelompokkan desa-desa di Provinsi Kalimantan Timur berdasarkan algoritma tersebut yakni sebanyak 4 kelompok. Dengan kecenderungan desa untuk mengelompok adalah desa-desa maju/mandiri di kelompok 1, desa berkembang yang merupakan desa-desa yang cukup majunya

termasuk desa-desa kurang maju pada kelompok 2 dan 3, dan kelompok 4 untuk desa-desa yang tertinggal.

## DAFTAR PUSTAKA

- Abonyi, János, Feil, B., Basel, B., Boston, ·, & Berlin, ·. (2007). *Cluster Analysis for Data Mining and System Identification*.
- Abonyi, Janos, & Szeifert, F. (2003). Supervised fuzzy clustering for the identification of fuzzy classifiers. *Pattern Recognition Letters*, 24(14), 2195–2207. [https://doi.org/10.1016/S0167-8655\(03\)00047-3](https://doi.org/10.1016/S0167-8655(03)00047-3)
- Badan Pusat Statistik. (2019). *Indeks Pembangunan Desa 2018* (S. S. K. Wilayah (ed.)). Badan Pusat Statistik.
- Balazs Balasko, J. A. and B. F. (2013). Analysis tools: Clustering and Data Analysis Toolbox. *Applied Mathematical Sciences (Switzerland)*, 183, 49–120. [https://doi.org/10.1007/978-1-4614-5975-0\\_2](https://doi.org/10.1007/978-1-4614-5975-0_2)
- Bezdek, J. C. (1981). Pattern Recognition with Fuzzy Objective Function Algorithms. *Pattern Recognition with Fuzzy Objective Function Algorithms*. <https://doi.org/10.1007/978-1-4757-0450-1>
- Bunkers, M. J., Miller, J. R., & DeGaetano, A. T. (1996). Definition of climate regions in the northern plains using an objective cluster modification technique. *Journal of Climate*, 9(1), 130–146. [https://doi.org/10.1175/1520-0442\(1996\)009<0130:DOCRIT>2.0.CO;2](https://doi.org/10.1175/1520-0442(1996)009<0130:DOCRIT>2.0.CO;2)
- Dunn, J. C. (1975). Optimal Fuzzy Partitions: A Heuristic for Estimating the Parameters in a Mixture of Normal Distributions. *IEEE Transactions on Computers*, C-24(8), 835–838. <https://doi.org/10.1109/T-C.1975.224317>
- Gath, I., & Geva, A. B. (1989). Unsupervised Optimal Fuzzy Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7), 773–780. <https://doi.org/10.1109/34.192473>
- Gustafson, D. E., & Kessel, W. C. (1978). Fuzzy Clustering With a Fuzzy Covariance Matrix. *Proceedings of the IEEE Conference on Decision and Control, December*, 761–766. <https://doi.org/10.1109/cdc.1978.268028>
- Johnson, R.A. and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis 6th Edition* (p. 794). Prentice Hall International Inc.
- Kim, D. W., Lee, K. H., & Lee, D. (2003). Fuzzy cluster validation index based on inter-cluster proximity. *Pattern Recognition Letters*, 24(15), 2561–2574. [https://doi.org/10.1016/S0167-8655\(03\)00101-6](https://doi.org/10.1016/S0167-8655(03)00101-6)
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1(14), 281–297.
- Maxwell, B. A., & Pryor, F. (2014). CLUSTER ANALYSIS IN CROSS-CULTURAL. *International Journal of World Cultures, January 2002*, 22–38.
- Munaf, A. R. M. N. S. P. (2009). *PENGELOMPOKKAN KABUPATEN / KOTA DI JAWA TIMUR SEBAGAI DASAR PERENCANAAN PEMBIAYAAN KEGIATAN PENDATAAN LAPANGAN MENGGUNAKAN ALGORITMA FUZZY C - MEANS DENGAN MODIFIKASI FORMULA JARAK*.
- Santosa, S. (2014). *Statistik Multivariat (Edisi Revisi): Konsep dan Aplikasi dengan SPSS*. PT Elex Media Komputindo.
- Syoyer, R. dan Mashuri, M. (2014). Analisis Kelompok Dengan Algoritma Fuzzy C-Means dan Gath-Geva Clustering. *IndoMS Journal on Statistics*, 2, 11–26.
- Valente de Oliveira, J., & Pedrycz, W. (2007). Advances in Fuzzy Clustering and its Applications. *Advances in Fuzzy Clustering and Its Applications*, 1–434. <https://doi.org/10.1002/9780470061190>
- Wang, X. Y., Garibaldi, J. M., Bird, B., & George, M. W. (2005). Fuzzy clustering in biochemical analysis of cancer cells. *Proceedings - 4th Conference of the European Society for Fuzzy Logic and Technology and 11th French Days on Fuzzy Logic and Applications, EUSFLAT-LFA 2005 Joint Conference*, 1118–1123.
- Xie, X. L., & Beni, G. (1991). A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8), 841–847. <https://doi.org/10.1109/34.85677>
- Zadeh, L. A. (1965). FUZZY SETS. In *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems* (pp. 394–432). [https://doi.org/10.1142/9789814261302\\_0021](https://doi.org/10.1142/9789814261302_0021)

**Lampiran 1. Komponen Matriks Hasil Rotasi**

Variabel	Komponen							
	1	2	3	4	5	6	7	8
X1_jml_tk	.759	.286	.072	.069	.245	.038	.098	-.042
X2_jml_sd	.693	.256	.082	.076	.390	-.014	.224	-.020
X3_jml_smp	.557	.183	.148	.286	.395	-.012	.186	-.026
X4_jarak_smp	.036	.225	.747	.055	.066	.038	.134	.022
X4_jml_sma	.446	.090	.213	.224	.475	-.084	.032	-.011
X4_jarak_sma	.093	.350	.747	.058	.131	-.047	-.077	-.018
X4_LB	.143	.109	-.012	.574	.078	-.032	.017	-.020
X5_jml_rs	.326	.094	.016	.676	-.014	.023	-.069	.005
X5_jarak_rs	.226	.772	.134	.162	.113	-.016	-.008	-.010
X6_jml_rs_salin	.213	.028	.037	.674	.025	-.018	.131	.009
X6_jrk_rs_salin	.314	.647	.057	.119	.133	.080	.008	-.176
X7_puskesban	.109	.000	.000	.052	.248	.061	.750	.018
X7_jrk_puskes1	.122	.157	.757	-.043	-.066	.060	.002	.097
X7_jrk_puskes2	.147	.652	.147	.103	.036	.221	.057	-.145
X7_jrk_pusban	-.038	.151	.025	.035	-.244	-.004	.762	.045
X8_poliklinik	.637	.126	.008	-.031	-.153	.030	.012	.089
X8_jrk_poliklinik	.211	.802	.131	-.001	-.040	-.044	.011	.075
X9_praktek_dokter	.698	.094	.067	.207	-.132	-.079	-.065	-.009
X9_jrk_dokter	.166	.670	.402	-.009	-.081	-.037	.056	.160
X10_bidan	.545	.288	.105	-.039	.195	-.041	.123	-.164
X10_jrk_bidan	.148	.693	.373	-.022	-.070	-.037	.134	.032
X11_posdes	-.035	-.103	-.018	-.021	-.023	.836	.056	.049
X11_jrk_posdes	.007	.133	.086	-.030	.012	.842	-.007	-.019
X12_apotek_jamu	.830	.163	.051	.253	.043	-.044	-.046	.028
X12_jrk_apotek	.191	.791	.205	.018	.036	-.059	.073	.055
X12_jrk_jamu	.835	.208	.022	.126	.110	.017	.052	.025
X13_pertokoan	.320	.246	.148	-.229	.376	-.211	.166	-.112
X14_pasar	.746	.199	.007	.116	.036	.030	-.016	.062
X15_rmh_makan	.638	-.021	.095	.356	-.080	-.062	-.063	.156
X16_hote_lmotel	.784	.109	.064	.162	-.085	-.019	-.075	.070
X17_bank	.612	.195	.081	.034	.298	.074	.076	-.019
X24_BTS	.382	.566	.194	.068	.138	-.043	.049	.011
X24_operatorHP	.054	.046	.130	-.023	.107	.024	.080	.717
X30_jrk_camat	.090	.297	.428	.020	.048	.067	-.200	.362
X31_biaya_camat	.104	.622	.036	.058	.073	-.046	.044	.487
X32_jrk_bupati	.051	.624	-.067	.049	-.003	.021	-.058	.404
X33_biaya_bupati	.018	.022	.037	-.030	-.634	-.035	.058	-.152
X35_gizi_burukmin	.018	.022	.037	-.030	-.634	-.035	.058	-.152

**Lampiran 2. Ukuran Hasil Pengelompokan dengan Algoritma FCM Clustering**

Jumlah Kelompok	Jumlah Iterasi	Fungsi Objektif	Waktu Komputasi (detik)	Rasio Sw/S <sub>B</sub>
2	25	17,1769	1,2672	0,424
3	46	15,1185	1,3607	0,3319
4	51	12,852	1,3453	0,3961
5	133	11,5442	1,4254	0,3431
6	273	10,9343	1,4614	0,3121
7	124	9,9863	1,3920	0,2791
8	157	9,4581	1,4255	0,2652
9	307	9,2091	1,4914	0,2486
10	100	8,5873	1,4008	0,2230
11	139	8,2713	1,4557	0,1947
12	208	8,217	1,5351	0,1944

**Lampiran 3.** Nilai Indeks Validitas *Cluster* dengan metode FCM

<b>Jumlah Kelompok</b>	<b>CE</b>	<b>PI</b>	<b>SI</b>	<b>Indeks XB</b>	<b>Indeks Dunn</b>
2	0,3066	2,2954	0,0022	10,3470	0,0079
3	0,5263	2,0161	0,0032	17,7276	0,0045
4	0,5928	1,3527	0,0022	8,2507	0,0025
5	0,7004	1,1233	0,0017	30,1065	0,0026
6	0,7747	0,9736	0,0015	12,3037	0,0025
7	0,8202	0,8637	0,0014	7,8247	0,0027
8	0,8810	0,8135	0,0013	13,1576	0,0037
9	0,9329	0,7957	0,0013	6,4242	0,0027
10	0,9495	0,6698	0,0011	12,7393	0,0055
11	0,9679	0,5885	9,32E-04	10,1683	0,0024
12	1,0317	0,6113	9,29E-04	7,7408	0,0038