

Pengelompokan Kabupaten/Kota Di Pulau Kalimantan Dengan Fuzzy C-Means Berdasarkan Indikator Kemiskinan

Clustering Regency/City In Borneo With Fuzzy C-Means Based On Poverty Indicators

Retno Ayu Ningtyas¹, Yuki Novia Nasution^{2,3}, dan Syaripuddin³

^{1,2,3}Laboratorium Matematika Komputasi, Program Studi Matematika, FMIPA, Universitas Mulawarman
Email: retnoayas@gmail.com

ABSTRACT

Cluster analysis is a branch of statistical science that is used to grouping data that have similar characteristics between each other. The grouping method used in this research is Fuzzy C-Means. Fuzzy C-Means method is one of the grouping methods developed from the C-Means method by applying the properties of fuzzy sets. With the existence of each data is determined by the degree of membership. This method is applied to data from 56 districts/cities on Borneo based on poverty indicators with variables namely the percentage of average length of schooling, life expectancy, percentage of the poor, percentage of open unemployment rate, percentage of households with proper sanitation, and percentage of households with proper drinking water. This study aims to obtain the results of grouping districts/cities on Borneo based on poverty indicators and to obtain optimal cluster results based on three validity indices, namely Connectivity, Dunn, and Silhouette values. Based on the results of the study, it was found that there were 2 optimal clusters, namely the first cluster consisted of 36 regencies/cities while the second cluster consisted of 20 regencies/cities.

Keywords: Cluster analysis, Fuzzy C-Means, Poverty Indicators, Validity Index

Pendahuluan

Analisis *cluster* merupakan sebuah cabang ilmu statistik analisis multivariat yang digunakan khusus untuk pengelompokan objek yang memiliki kesamaan karakter. Objek yang sama akan dikelompokkan menjadi sebuah kelas yang memiliki beberapa anggota (Prasetyo, 2012). Terdapat beberapa metode dalam analisis cluster, diantaranya adalah *non-fuzzy clustering* dan *fuzzy clustering*. Perbedaan antara *fuzzy clustering* dengan *non-fuzzy clustering* terletak pada derajat keanggotaannya. Pada *fuzzy clustering* harus ditentukan terlebih dahulu derajat keanggotaannya sedangkan pada *non-fuzzy clustering* tidak perlu ditentukan derajat keanggotaannya melainkan langsung dilakukan proses *clustering* (Kusumadewi, 2002).

Fuzzy clustering merupakan salah satu teknik untuk menentukan *cluster* optimal dalam suatu ruang vektor yang didasarkan pada jarak antar vektor. *Fuzzy clustering* sangat berguna bagi pemodelan *fuzzy* terutama dalam mengidentifikasi aturan-aturan *fuzzy* (Kusumadewi, 2006). Ada beberapa algoritma *clustering* data, diantaranya adalah *K-Means* dan *Fuzzy C-Means*. *Fuzzy C-Means* adalah suatu teknik *clustering* data dimana keberadaan tiap-tiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaan. Teknik ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981 (Kusumadewi dan Purnomo, 2010).

Kelebihan dari *Fuzzy C-Means* adalah dapat melakukan *clustering* lebih dari satu variabel sekaligus. *Fuzzy C-Means* menggunakan model

pengelompokan *fuzzy* sehingga data dapat menjadi anggota dari semua *cluster* dengan derajat atau tingkat keanggotaan yang berbeda antara 0 hingga 1. Kedekatan objek data ke pusat *cluster* menentukan tingkat keanggotaan objek data ke sebuah *cluster* (Kusumadewi, 2007). Metode *Fuzzy C-Means* juga dapat menempatkan pusat *cluster* yang lebih tepat dibandingkan dengan metode *clustering* lainnya. Selain memiliki kelebihan *fuzzy c-means* juga memiliki keterbatasan diantaranya mudah terjebak dalam optimum lokal dan sensitif terhadap pusat *cluster* awal. Untuk mengatasi keterbatasan tersebut, maka pada penelitian ini digunakan tiga indeks validitas yaitu Indeks *Dunn*, Indeks *Silhouette*, dan Indeks *Connectivity*.

Validitas *cluster* merupakan suatu tahap kegiatan yang digunakan untuk memberikan penilaian hasil dari analisis suatu *cluster* dengan berdasarkan jumlah atau kuantitatif dan sesuai dengan keadaan sebenarnya (Kapita & Abdullah, 2020). Selain itu, validitas *cluster* juga salah satu metode yang digunakan untuk melakukan evaluasi hasil dari sebuah algoritma *clustering* yang telah dilakukan dengan tujuan untuk memperoleh jumlah *cluster* terbaik (Khairati, Adlina, Hertono & Handari, 2019).

Pada penelitian yang telah dilakukan, dapat dilihat bahwa penerapan *Fuzzy C-Means* dapat diterapkan pada Indikator Kemiskinan. Kemiskinan merupakan salah satu masalah kesejahteraan yang harus dituntaskan atau mengalami penurunan tingkat kemiskinan,

sehingga upaya yang harus dilakukan secara komprehensif dalam mencakup berbagai aspek kehidupan masyarakat dan dilaksanakan secara terpadu. Badan Pusat Statistik mencatat penduduk Indonesia yang masih hidup dibawah garis kemiskinan hingga September 2020 sebesar 27,55 juta orang atau 10,15 persen dari total penduduk Indonesia. Pulau Kalimantan memiliki jumlah penduduk miskin terendah yaitu sebanyak 969,64 ribu orang.

Teori/Metodologi

Analisis *cluster* melakukan pengelompokan data ke dalam sejumlah kelompok (*cluster*) berdasarkan kesamaan karakteristik masing-masing data pada kelompok-kelompok yang ada.

Fuzzy C-Means

Metode *Fuzzy C-Means* menurut 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 (Kusrini dan Luthfi, 2009).

Konsep dasar FCM pertama kali adalah menentukan pusat kelompok, yang akan menandai lokasi rata-rata untuk tiap-tiap kelompok. Pada kondisi awal, pusat kelompok ini masih belum akurat. Tiap-tiap titik data memiliki derajat keanggotan untuk tiap-tiap kelompok. Dengan cara memperbaiki pada kelompok dan derajat keanggotaan tiap-tiap titik data secara berulang, maka akan dapat dilihat bahwa pusat kelompok akan bergerak menuju lokasi yang tepat. Perulangan ini didasarkan pada minimisasi fungsi objektif yang menggambarkan jarak dari titik data yang diberikan ke pusat kelompok yang terbobot oleh derajat keanggotaan titik data tersebut (Kusumadewi, 2006). Dalam metode FCM dipergunakan variabel derajat keanggotaan atau *membership function* (μ_{ik}) yang merujuk seberapa besar kemungkinan suatu data bisa menjadi anggota ke dalam suatu kelompok. *Membership function* untuk suatu data ke suatu kelompok tertentu mempunyai jangkauan nilai $0 \leq \mu_{ik} \leq 1$. Metode FCM memperkenalkan suatu variabel w yang merupakan *weighting exponent* dari *membership function*, variabel ini dapat mengubah besar pengaruh dari *membership function* dalam proses pengelompokan menggunakan metode FCM, variabel w mempunyai nilai lebih besar dari 1 ($w > 1$). Nilai

variabel w yang umumnya digunakan adalah 2 (Bezdek dan Dunn, 1975).

Algoritma Fuzzy C-Means

Tahapan dari metode FCM untuk pengelompokan data secara umum dirumuskan sebagai berikut:

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

$$\mu_{ik} \in [0,1], i = 1,2, \dots, n; k = 1,2, \dots, c \quad (1)$$

μ_{ik} adalah derajat keanggotaan yang merujuk pada seberapa besar kemungkinan suatu data bisa menjadi anggota ke dalam suatu kelompok. Misalkan Q_i adalah jumlah derajat keanggotaan μ_{ik} pada setiap kolom matriks U , yaitu:

$$Q_i = \sum_{k=1}^c \mu_{ik} \quad (2)$$

Nilai elemen matriks yang ternormalisasi dihitung dengan persamaan:

$$\mu_{iknorm} = \frac{\mu_{ik}}{Q_i} \quad (3)$$

Bentuk matriks partisi awal yang telah dinormalisasi sebagai berikut :

$$U^0 = \begin{bmatrix} \mu_{11norm} & \mu_{21norm} & \dots & \mu_{n1norm} \\ \mu_{12norm} & \mu_{22norm} & \dots & \mu_{n2norm} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{c1norm} & \mu_{c2norm} & \dots & \mu_{cnorm} \end{bmatrix} \quad (4)$$

Nilai elemen-elemen pada setiap kolom matriks partisi awal U^0 harus mendekati atau sama dengan satu.

3. Menghitung *centroid* v_{kj} dengan $k=1,2,\dots,c$; dan $j=1,2,\dots,m$; menggunakan persamaan berikut :

$$v_{kj} = \frac{\sum_{i=1}^n (\mu_{iknorm})^w x_{ij}}{\sum_{k=1}^c (\mu_{iknorm})^w} \quad (5)$$

Persamaan untuk proses iterasi pusat kelompok :

$$v_{kj} = \frac{\sum_{i=1}^n (\mu_{ikiter})^w x_{ij}}{\sum_{i=1}^n (\mu_{ikiter})^w} \quad (6)$$

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

$$v^t = \begin{bmatrix} v_{11} & v_{21} & \dots & v_{1m} \\ v_{21} & v_{22} & \dots & v_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ v_{c1} & v_{c2} & \dots & v_{cm} \end{bmatrix} \quad (7)$$

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

$$\mu_{ikiter}^t = \frac{[\sum_{j=1}^m (x_{ij} - v_{kj}^t)^2]^{-\frac{1}{w-1}}}{\sum_{k=1}^c [\sum_{j=1}^m (x_{ij} - v_{kj}^t)^2]^{-\frac{1}{w-1}}} \quad (8)$$

Dengan μ_{ikiter}^t merupakan derajat keanggotaan pada data kelompok ke- k untuk objek ke- i dalam proses iterasi ke- t . Derajat keanggotaan yang lebih diperbaiki ini merupakan elemen matriks U^t berukuran $c \times n$. Bentuk matriksnya adalah sebagai berikut:

$$v^t = \begin{bmatrix} v_{11} & v_{21} & \dots & v_{1m} \\ v_{21} & v_{22} & \dots & v_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ v_{c1} & v_{c2} & \dots & v_{cm} \end{bmatrix} \quad (9)$$

Dalam menghitung μ_{ikiter}^t pada iterasi ke- t (elemen matriks U^t), maka nilai v_{ij} yang digunakan adalah v_{ij} pada iterasi ke- t pula (pada 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 langkah akhir. Persamaan untuk nilai awal fungsi objektif untuk $t = 1$:

$$P^t = \sum_{i=1}^n \sum_{k=1}^k ([\sum_{j=1}^m (x_{ij} - v_{kj}^t)^2] (\mu_{iknorm})^2) \quad (10)$$

Persamaan untuk proses iterasi fungsi objektif Untuk $t = 1, 2, \dots$:

$$P^t = \sum_{i=1}^n \sum_{k=1}^k ([\sum_{j=1}^m (x_{ij} - v_{kj}^t)^2] (\mu_{ikiter}^t)^2) \quad (11)$$

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 μ_{ikiter}^t iterasi ke- $(t-1)$ yaitu elemen matriks U^{t-1} .

5. Mengecek kondisi penghentian iterasi Jika $|P_t - P_{t-1}| < \epsilon$ atau $t >$ maksimum iterasi maka iterasi dihentikan, jika tidak, maka iterasi dinaikkan $t=t+1$ dengan mengulangi langkah ke-4 untuk memulai iterasi baru.

Indeks Validitas

Indeks validitas merupakan suatu ukuran yang digunakan untuk menentukan jumlah kelompok yang optimal (Wang & Zhang, 2007). Menurut

Halkidi, Batistakis & Vazirgiannis (2001), terdapat tiga jenis kriteria indeks validitas, yaitu:

- 1) Kriteria Eksternal: Mengevaluasi hasil algoritma *clustering* berdasarkan struktur yang sudah ditentukan.
- 2) Kriteria Internal: Mengevaluasi hasil algoritma *clustering* secara kuantitatif yang mencakup vektor dari himpunan data tersebut.
- 3) Kriteria Relatif: Mengevaluasi hasil algoritma *clustering* dengan membandingkan algoritma tersebut dengan skema *clustering* lainnya.

Validasi hasil *clustering* dapat menggunakan beberapa indeks diantaranya yaitu indeks *Dunn*, *Silhouette*, dan *Connectivity*. Pada indeks *Connectivity* membentuk jumlah *cluster* terbaik apabila nilai yang dihasilkan semakin kecil dibandingkan dengan nilai dari *cluster* yang lain (Halim & Widodo, 2017). Sedangkan untuk indeks *Dunn* menghasilkan *cluster* terbaik jika nilai *Dunn* yang diperoleh semakin besar. Kemudian untuk indeks *Silhouette* menghasilkan *cluster* terbaik jika nilai yang diperoleh semakin dekat dengan angka 1 (Khairati, Adlina, Hertono & Handari, 2019).

Kemiskinan

Secara umum, kemiskinan diartikan sebagai kondisi ketidak mampuan pendapatan dalam mencukupi kebutuhan pokok sehingga kurang mampu untuk menjamin kelangsungan hidup. Definisi kemiskinan dalam arti luas adalah keterbatasan yang dimiliki oleh seseorang, keluarga, komunitas, bahkan negara yang menyatakan ketidaknyamanan dalam kehidupan, terancamnya penegakan hak dan keadilan, terancamnya posisi tawar (*bargaining*) dalam pergaulan dunia, hilangnya generasi serta suramnya masa depan bangsa (Haryadi, 2005). Secara sederhana, miskin berarti tidak mampu memenuhi kebutuhan pokok anggota keluarga baik berupa pangan maupun non pangan (Tjondronegoro, 1996 dalam Girsang, 2011). Badan Pusat Statistik menentukan garis kemiskinan berdasarkan pengeluaran rumah tangga yang nilainya sekitar Rp 211.726 (Maret 2010) atau Rp 233.740 per kapita per bulan (Maret 2011) (Girsang, 2011). Berbeda dengan BPS, Bank Dunia memberikan definisi kemiskinan yang lebih komprehensif, meliputi kebutuhan dasar, deprivasi kesejahteraan dan ketidakcukupan kapasitas untuk hidup lebih baik (Girsang, 2011).

Hasil Penelitian dan Pembahasan

1. Data Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder. Pada Teknik pengambilan data ini dilakukan dengan mencatat data sampel di

website Badan Pusat Statistik tentang indikator kemiskinan 56 kabupaten/kota di Pulau Kalimantan. Adapun variabel yang digunakan dalam penelitian ini yaitu enam variabel: rata-rata lama sekolah (X_1), angka harapan hidup (X_2), tingkat pengangguran terbuka (X_3), persentase penduduk miskin (X_4), persentase rumah tangga dengan akses sanitasi layak (X_5) dan persentase rumah tangga dengan akses air minum layak (X_6).

2. Fuzzy C-Means

Adapun tahapan-tahapan clustering metode *Fuzzy C-Means* dalam clustering data kabupaten/kota di Pulau Kalimantan berdasarkan indikator kemiskinan sebagai berikut:

1. Menghitung indeks validitas

Hasil uji validasi dilakukan dengan menggunakan *package cValid* pada *software R*. Kemudian didapatkan nilai indeks *Dunn*, indeks *Silhouette* dan indeks *Connectivity* dari *clustering* menggunakan *Fuzzy C-Means* untuk data kabupaten/kota di Pulau Kalimantan berdasarkan indikator kemiskinan pada masing-masing *cluster* yang terbentuk sesuai dengan Tabel 1 dan Tabel 2.

Tabel 1. Hasil Uji Validasi Internal

	<i>Connectivity</i>	<i>Dunn</i>	<i>Silhouette</i>
2 cluster	27,2587	0,1377	0,2365
3 cluster	35,7607	0,1668	0,2014
4 cluster	42,6833	0,1978	0,2091
5 cluster	52,7480	0,2055	0,2061

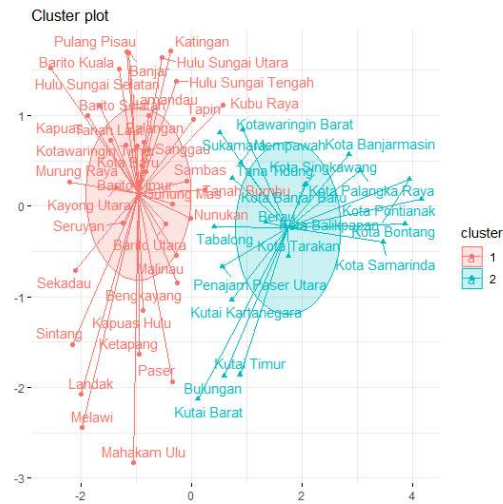
Tabel 2. Jumlah Cluster Optimal

Indeks Validitas	Nilai	Cluster Optimal
<i>Connectivity</i>	27,2587	2
<i>Dunn</i>	0,2055	5
<i>Silhouette</i>	0,2365	2

Tabel 2 menunjukkan nilai indeks validasi internal terbaik serta jumlah *cluster* optimal yang digunakan untuk metode *Fuzzy C-Means Clustering* yaitu dengan menggunakan pengelompokan 2 *cluster* yang mempunyai indeks *connectivity* lebih kecil yaitu sebesar 27,2587 dan indeks *silhouette* yaitu 0,2365 lebih besar dibandingkan jumlah *cluster* lainnya namun nilai indeks *dunn* lebih kecil dibandingkan jumlah *cluster* lainnya.

2. Pengelompokan dengan Fuzzy C-Means

Tahap selanjutnya dilakukan proses pengelompokan menggunakan metode *Fuzzy C-Means* pada aplikasi *R* dengan menggunakan banyak *cluster* $c = 2$. Diperoleh hasil *cluster* yang terdiri dari 56 Kabupaten/Kota di Pulau Kalimantan yang ditunjukkan pada Gambar 1 sebagai berikut:



Gambar 1. Hasil Cluster Kabupaten/Kota yang terbentuk

Berdasarkan hasil analisis diatas dapat diketahui bahwa anggota *cluster* dengan $c = 2$ pada metode *Fuzzy C-Means* adalah sebagai berikut:

1. *Cluster 1* beranggotakan kabupaten/kota di Pulau Kalimantan yaitu Sambas, Bengkayang, Landak, Sanggau, Ketapang, Sintang, Kapuas Hulu, Sekadau, Melawi, Kayong Utara, Kubu Raya, Tanah Laut, Kota Baru, Banjar, Barito Kuala, Tapin, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara, Tanah Bumbu, Balangan, Malinau, Nunukan, Kotawaringin Barat, Kotawaringin Timur, Kapuas, Barito Selatan, Barito Utara, Lamandau, Seruyan, Katingan, Pulang Pisau, Gunung Mas, Barito Timur, Murung Raya, Paser, dan Mahakam Ulu.
2. *Cluster 2* beranggotakan kabupaten/kota di Pulau Kalimantan yaitu Mempawah, Kota Pontianak, Kota Singkawang, Tabalong, Kota Banjarmasin, Kota Banjar Baru, Bulungan, Tana Tidung, Kota Tarakan, Kotawaringin Barat, Sukamara, Kota Palangkaraya, Kutai Barat, Kutai Kartanegara, Kutai Timur, Berau, Penajam Paser Utara, Kota Balikpapan, Kota Samarinda, dan Kota Bontang.

Kesimpulan

Berdasarkan hasil analisis data dan pembahasan diperoleh kesimpulan sebagai berikut:

1. Berdasarkan hasil indeks validitas *connectivity* sebesar 27,2587 dan nilai *silhouette* sebesar 0,2365 pada metode *Fuzzy C-Means* bahwa jumlah *cluster* yang optimal untuk mengelompokkan Kabupaten/Kota di Pulau Kalimantan berdasarkan indikator kemiskinan adalah sebanyak 2 *cluster*.

2. Anggota pada *cluster* 1 berjumlah 36 Kabupaten/Kota. Sedangkan anggota pada *cluster* 2 berjumlah 20 Kabupaten/Kota.

Saran

Saran yang dapat diberikan untuk penelitian selanjutnya yaitu melakukan perbandingan antara metode *Fuzzy clustering* yang lainnya seperti *Fuzzy subtractive clustering*. Serta disarankan menggunakan indeks validitas yang lain seperti *partition coefficient* dan *exponential*.

Daftar Pustaka

- Bezdek, J.C dan Dunn, J. (1975). Optimal Fuzzy Partitions: A Heuristic for Estimating the Parameters in a Mixture of Normal Distributions. *IEEE Transactions on Computers*, 835-838.
- Girsang, Wardis. (2011). *Kemiskinan Multidimensional Di Pulau-Pulau Kecil*. Ambon: BFP-UNPATTI.
- Halkidi, Maria., Batistakis, Yannis, Batistakis., dan Michalis, Vazirgiannis. (2001). On Clustering Validation Techniques. *Journal of Intelligent Information Systems* 17, 107-145.
- Haryadi, Agus. (2005). *Kemiskinan, Mentalitas Budaya*. Yogyakarta: Pustaka Rahima.
- Kapita, S.N., & Abdullah, S. Do. (2020). Pengelompokan Data Mutu Sekolah Menggunakan Metode Self-Organizing Map (SOM) dan K-Means. *Jurnal Teknik ITS*, A368-A373.
- Khairati, A.F., Adlina, A., Hertono, G., & Handari, B. (2019). Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA. In: *Prosiding Seminar Nasional Matematika*
- Klawonn, F, dkk. (2007). Evolving Fuzzy Rule-Based Classifiers. *IEEE Press* 220-225
- Kusrini dan Luthfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Offset
- Kusumadewi, Sri. (2002). *Analisis dan Desain Sistem Fuzzy Menggunakan Tool Box*. Yogyakarta: Graha Ilmu.
- Kusumadewi, S, dkk. (2006). *Fuzzy Multi-Attribute Decision Making (Fuzzy MADM)*. Yogyakarta: Graha Ilmu.
- Kusumadewi, Sri. (2007). Klasifikasi Kandungan Nutrisi Bahan Pangan Menggunakan Fuzzy C-Means. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, 53-58.
- Kusumadewi, S, dan Purnomo, H. (2010). *Aplikasi Logika Fuzzy Untuk Pendukung Keputusan*. Yogyakarta: Graha Ilmu.
- Wang, Wei., dan Zhang, Qi. (2007). A Fast Algorithm for Approximate Quantiles in High-Speed Data Streams. *IEEE*.

