

PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN INDIKATOR PEMBANGUNAN PENDIDIKAN MENGGUNAKAN METODE *SUBTRACTIVE FUZZY C-MEANS*

DINA MAULIDYA, YUDIANTRI ASDI*, HAZMIRA YOZZA

*Departemen Matematika dan Sains Data,
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,
Kampus UNAND Limau Manis Padang, Indonesia,
email : dinamaulidya33@gmail.com, yudiantriasdi@sci.unand.ac.id,
hazmirayozza@sci.unand.ac.id*

Diterima 29 Januari 2022 Direvisi 9 Juni 2022 Dipublikasikan 31 Juli 2022

Abstrak. Pendidikan menjadi salah satu tujuan utama dalam rencana pembangunan di Indonesia. Pembangunan pendidikan diukur dengan indikator-indikator terkait pendidikan. Ketercapaian pembangunan pendidikan di setiap provinsi di Indonesia berbeda, sehingga dengan cara mengelompokkan provinsi-provinsi tersebut berdasarkan kemiripan indikator yang tercapai dapat memudahkan pemerintah memberikan program peningkatan pembangunan pendidikan. Indikator pembangunan pendidikan yang digunakan dalam pengelompokan adalah sarana dan prasarana pendidikan, sanitasi sekolah, partisipasi sekolah, dan angka putus sekolah. Metode yang digunakan pada penelitian ini adalah metode *Subtractive Fuzzy C-Means*. Pengolahan data mengambil jari-jari yang beragam yaitu 1.00, 1.10, 1.20, 1.30, dan 1.50. Hasil indeks validitas kluster menunjukkan jari-jari 1.50 yang membentuk dua kluster merupakan jumlah kluster terbaik. Jumlah keanggotaan kluster pertama sebanyak 20 provinsi. Provinsi-provinsi yang menjadi anggotanya tersebar di seluruh Indonesia bagian Barat, kecuali Aceh, sebagian provinsi di Indonesia bagian Tengah, yaitu Pulau Kalimantan kecuali Kalimantan Tengah, Gorontalo dan Bali. Sedangkan jumlah provinsi yang masuk keanggotaan kluster kedua sebanyak 14 provinsi. Keanggotaannya tersebar pada semua provinsi di luar kluster pertama. Berdasarkan karakteristik kluster, kluster kedua merupakan kluster terbaik dibanding kluster pertama.

Kata Kunci: Pembangunan Pendidikan, Pengelompokan, *Subtractive Fuzzy C-Means*

1. Pendahuluan

Pendidikan merupakan salah satu upaya dalam pengembangan sumber daya manusia (untuk selanjutnya disingkat SDM), yang bertujuan untuk membangun masyarakat pekerja keras yang dinamis, produktif, terampil menguasai ilmu pengetahuan dan teknologi. Peningkatan akses layanan pendidikan selalu diupayakan pemerintah dalam rangka optimalisasi layanan pendidikan yang bermutu dan ber-

*Penulis Korespondensi

daya saing. Karena ketercapaian pembangunan pendidikan untuk setiap provinsi berbeda-beda, maka pengelompokan semua provinsi berdasarkan kemiripan indikator pendidikan yang tercapai dapat memudahkan pemerintah memberikan program/kebijakan peningkatan pembangunan pendidikan.

Salah satu metode analisis kluster adalah *fuzzy clustering*. *Fuzzy clustering* adalah metode pengklasteran data yang mana keberadaan setiap titik data dalam suatu kluster ditentukan oleh derajat keanggotaannya. Terdapat beberapa metode dalam *fuzzy clustering*, dua di antaranya yang sering digunakan adalah *fuzzy c-means* (FCM) dan *fuzzy subtractive clustering* (FSC). Metode FCM baik untuk mengelompokkan data yang memiliki lebih dari satu peubah. Namun, banyaknya peubah mengakibatkan kelompok yang terbentuk sangat banyak dan waktu komputasi yang lama. Di sisi lain, Metode FSC memberikan hasil pengelompokan yang lebih konsisten dan komputasi yang cepat. Namun metode ini memiliki akurasi yang lebih rendah. Agar hasil pengelompokan optimal, dilakukan penelitian untuk menggabungkan metode *Subtractive Clustering* dan *Fuzzy C-Means* membentuk metode baru yaitu *Subtractive Fuzzy C-Means* (SFCM) [5].

Untuk dapat menentukan seberapa bagus pengelompokan objek dengan menggunakan metode *fuzzy*, digunakan suatu indeks validitas. Terdapat dua kategori untuk menghitung indeks validitas, yaitu berdasarkan derajat keanggotaan dan bobot data itu sendiri. Salah satu indeks yang digunakan pada penelitian ini adalah *partition coefficient indeks* (PCI). PCI adalah indeks validitas yang mengevaluasi derajat keanggotaan antara setiap objek di dalam kelompok [6].

2. Landasan Teori

2.1. Analisis Kluster

Analisis kluster adalah salah satu metode peubah ganda yang bertujuan mengelompokkan objek-objek berdasarkan karakteristiknya ke dalam beberapa kelompok. Dengan analisis kluster, objek-objek tersebut diklasifikasikan ke dalam satu atau lebih kelompok (kluster) hingga setiap anggota dalam satu kelompok memiliki kemiripan karakteristik [2].

Salah satu pendekatan untuk mengukur ketaksamaan karakteristik dalam kluster adalah *Euclidian Distance* (jarak Euclidean). Persamaan untuk menghitung jarak euclidean adalah

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}. \quad (2.1)$$

2.2. Himpunan Fuzzy

Himpunan tegas adalah suatu himpunan yang terdefinisi secara tegas. Suatu objek dalam himpunan tersebut dapat ditentukan secara tegas masuk ke dalam anggota himpunan atau tidak. Karakteristik fungsi dari himpunan tegas ditentukan bernilai 1 jika merupakan anggota himpunan, atau 0 jika bukan merupakan anggota himpunan. Himpunan *fuzzy* didasari gagasan untuk memperluas jangkauan fungsi

keanggotaan pada himpunan tegas sedemikian sehingga fungsi tersebut berada di antara interval $[0, 1]$ [3].

2.3. Algoritma Fuzzy C-Means

Algoritma *Fuzzy C-Means* terdiri dari langkah-langkah berikut [5]:

- (1) Menginput data yang akan dikelompokkan berupa matriks X_{ij} , yaitu data sampel ke- i , $i = 1, 2, \dots, n$, peubah ke- j , $j = 1, 2, \dots, m$.
- (2) Menentukan nilai parameter awal yang berupa jumlah kluster, *fuzzifier* ($m = 2$), Fungsi objektif awal ($P_0 = 0$).
- (3) Menentukan maksimum iterasi (*MaxIter*) dan *error* terkecil yang diharapkan.
- (4) Hitung pusat kluster ke- l (C_{lj}) dengan rumus berikut.

$$C_{lj} = \frac{\sum_{i=1}^n (\mu_{li})^m (x_{ij})}{\sum_{i=1}^n (\mu_{li})^m}. \quad (2.2)$$

- (5) Menghitung fungsi objektif pada iterasi ke- t .

$$P_t = \sum_{i=1}^n \sum_{l=1}^c \left(\left[\sum_{j=1}^p (x_{ij} - C_{lj})^2 \right] (\mu_{li})^m \right). \quad (2.3)$$

- (6) Menghitung perubahan matrik partisi.

$$\mu_{li} = \frac{\left[\sum_{j=1}^p (x_{ij} - C_{lj})^2 \right]^{\frac{1}{m-1}}}{\sum_{l=1}^c \left[\sum_{j=1}^p (x_{ij} - C_{lj})^2 \right]^{\frac{1}{m-1}}}. \quad (2.4)$$

- (7) Memeriksa kondisi berhenti, jika $|P_t - P_{t-1}| < \epsilon$ atau $t > \text{MaxIter}$ maka iterasi berhenti, jika tidak maka tambah iterasi dan ulangi langkah ke-5.

2.4. Algoritma Subtractive Clustering

Langkah-langkah pengelompokan dengan *subtractive clustering* sebagai berikut [1]:

- (1) Input data yang akan dikelompokkan berupa matrik.
- (2) Tetapkan nilai parameter untuk semua peubah, yaitu jari-jari, *squash factor*, *accept ratio*, *reject ratio*.
- (3) Tentukan nilai minimum ($x_{\min}$) dan nilai maksimum ($x_{\max}$) untuk setiap peubah.
- (4) Normalisasi data menjadi data dengan bobot nilai 0 sampai 1 dengan menggunakan rumus berikut:

$$X_{ij}^* = \frac{x_{ij} - x_{\min j}}{x_{\max j} - x_{\min j}} \quad (2.5)$$

- (5) Menentukan potensi awal setiap titik data, D_k ($k = 1, 2, \dots, n$) dengan menggunakan rumus

$$D_k = \sum_{i=1}^n e^{\left(-\frac{4\|\mathbf{X}_k - \mathbf{X}_i\|^2}{r_a^2}\right)}. \quad (2.6)$$

- (6) Menentukan data dengan potensi tertinggi, ($\mathbf{C}_1$). Vektor $\mathbf{C}_1$ merupakan data yang memenuhi $D_{c1} = \max_k D_k$.

Selanjutnya tentukan $\mathbf{C}_1$ sebagai pusat kluster pertama.

- (7) Mengurangi potensi titik-titik data yang lain dengan rumus.

$$D_k' = D_k - D_{c1} \times e^{-\frac{4\|\mathbf{X}_k - \mathbf{X}_{c1}\|^2}{r_b^2}} \quad (2.7)$$

dengan : $r_b = q \times r_a$.

- (8) Tetapkan data dengan potensi tertinggi ($D_{k_{\max}}'$) sebagai calon pusat kluster, sebut saja $\mathbf{V}$.

- (9) Hitung rasio $\frac{D_{k_{\max}}'}{D_{c1}}$, didapatkan tiga kemungkinan kondisi :

- a. Jika rasio $> accept\ ratio$ maka $\mathbf{V}$ diterima sebagai pusat kluster.
- b. Jika nilai rasio $< accept\ ratio$ dan nilai rasio $> reject\ ratio$ maka $\mathbf{V}$ akan diterima jika berada cukup jauh dari $\mathbf{C}_1$. Untuk menetapkan $\mathbf{V}$ sebagai pusat kluster, lakukan langkah-langkah berikut.
 - i. Menghitung jarak $\mathbf{V}$ dengan pusat kluster yang sudah terbentuk $\mathbf{C}_l$ menggunakan rumus:

$$S_{dl} = \sum_{j=1}^p \left(\frac{V_j - C_{lj}}{r_a} \right)^2 \quad (2.8)$$

- ii. Menghitung jarak terdekat $\mathbf{V}$ dengan pusat kluster lain yang disimbolkan dengan Mds . Rumus menghitung $Mds = \sqrt{Md}$.

A. Jika $(rasio + Mds) \geq 1$ maka $\mathbf{V}$ dapat diterima sebagai pusat kluster baru.

B. Jika $(rasio + Mds) < 1$ maka $\mathbf{V}$ tidak dapat diterima dan tidak akan dipertimbangkan lagi sebagai pusat kluster.

Potensi data tersebut diatur menjadi 0.

- c. Jika nilai rasio $< accept\ ratio$ dan nilai rasio $< reject\ ratio$ maka $\mathbf{V}$ tidak diterima sebagai pusat kluster dan tidak ada lagi calon pusat kluster baru, Iterasi dihentikan.

- (10) Mengulangi langkah ke-3 sampai ke-5 hingga proses iterasi berhenti.

- (11) Mengembalikan pusat kluster dari bentuk ternormalisasi ke bentuk semula (denormalisasi).

$$C_{ljdenorm} = C_{lj} \times (x_{\max j} - x_{\min j}) + x_{\min j} \quad (2.9)$$

2.5. Subtractive Fuzzy C-Means

Metode *Subtractive Fuzzy C-Means* merupakan penggabungan dari metode *Subtractive Clustering* dan *Fuzzy C-Means*. Metode ini diawali dengan algoritma SC untuk menentukan pusat kluster [4]. Langkah-langkah yang digunakan sama dengan yang dijelaskan pada subbab 2.4. Setelah didapatkan matriks pusat kluster, ditentukan derajat keanggotaan menggunakan algoritma FCM. Langkah-langkah algoritma FCM meliputi membentuk matriks derajat keanggotaan awal dan menghitung fungsi objektif. Derajat keanggotaan terakhir ditentukan setelah mendapatkan nilai fungsi objektif yang stabil.

2.6. Validitas Kluster

Kriteria untuk menentukan jumlah kluster yang optimal dalam pengelompokan dengan metode *fuzzy* menggunakan indeks validitas kluster. Salah satu indeks validitas adalah *Partition Coefficient Indeks*. PCI berguna untuk mengevaluasi derajat keanggotaan antara setiap data di dalam kluster. Nilainya berada dalam rentang $[0, 1]$. Rumus untuk menghitung PC adalah [6]:

$$PC = \frac{1}{N} \sum_{l=1}^c \sum_{i=1}^N \mu_{li}^2. \quad (2.10)$$

Nilai Indeks PC yang semakin besar (mendekati nilai 1) menunjukkan bahwa perbedaan karakteristik antar kluster semakin besar sehingga kualitas kluster yang terbentuk semakin bagus [6].

3. Metode Penelitian

Objek pengamatan pada penelitian ini adalah provinsi-provinsi di Indonesia. Data ini merupakan data sekunder tentang indikator-indikator pembangunan pendidikan tahun 2020. Data berasal dari hasil Survei Sosial Ekonomi (SUSENAS) tahun 2020 dipublikasikan oleh Badan Pusat Statistik.

Peubah yang digunakan pada penelitian ini adalah Rasio jumlah sekolah terhadap penduduk usia terkait setiap tingkatan pendidikan ($X_1 - X_4$), rasio guru layak mengajar terhadap peserta didik setiap tingkatan pendidikan ($X_5 - X_8$), rasio sekolah yang memiliki ketersediaan toilet terpisah setiap tingkatan pendidikan ($X_9 - X_{12}$), persentase anak usia 3-6 tahun yang mengikuti pendidikan prasekolah (X_{13}), Persentase angka partisipasi sekolah (X_{14}), angka putus sekolah setiap tingkatan pendidikan ($X_{15} - X_{17}$).

Tahap-tahap dalam pengelompokan dengan menggunakan metode *Subtractive Fuzzy C-Means* adalah :

1. Melakukan pengelompokan provinsi di Indonesia berdasarkan indikator pembangunan pendidikan menggunakan *subtractive fuzzy c-means* (SFCM) dengan cara:
 - a. Menentukan pusat kluster dengan algoritma *subtractive clustering* (SC).
 - b. Menentukan derajat keanggotaan dengan *fuzzy c-means* (FCM) dengan pusat kluster awal yang diperoleh dari langkah ke-a.

2. Memberikan interpretasi karakteristik hasil pengelompokan dengan metode *subtractive fuzzy c-means* (SFCM).
3. Mengukur kinerja kluster dengan Indeks *partition coefficient*.

4. Pembahasan

4.1. Hasil Pengelompokan

Proses pengelompokan dengan menggunakan metode SFCM mengambil jari-jari yang beragam. Jari-jari yang digunakan adalah $r = 1.00, 1.10, 1.20, 1.30, \text{ dan } 1.50$. Parameter-parameter dalam pengelompokan yang digunakan menggunakan nilai yang disarankan pada penelitian, yaitu *squash factor* = 1.25; *accept ratio* = 0.5; *reject ratio* = 0.15; *fuzzifier* (m) = 2 dan *Maximum Iteration* (MaxIter) = 1000. Dengan jari-jari tersebut banyaknya jumlah kluster yang terbentuk adalah sebagai berikut.

Tabel 1. Banyak Kluster yang terbentuk berdasarkan nilai jari-jari

Jari-jari (r)	Banyaknya Kluster
1.00	11
1.10	9
1.20	6
1.30	3
1.50	2

Dari Tabel 1 dapat dilihat bahwa banyaknya kluster yang diperoleh dengan $r = 1.00$ adalah 11 kluster, $r = 1.10$ adalah 9 kluster, $r = 1.20$ sebanyak 6 kluster, $r = 1.30$ sebanyak 3 kluster dan $r = 1.50$ sebanyak 2 kluster.

4.2. Jumlah Kluster Terbaik

Penentuan jumlah kluster terbaik dapat dilihat berdasarkan indeks validitas kluster. Indeks validitas kluster yang digunakan yaitu indeks PC. Perhitungan PCI dilakukan pada seluruh jari-jari dengan hasilnya yang disajikan dalam Tabel 2 berikut.

Tabel 2. Nilai *Partition Coefficient Indeks*

Jari-jari	Banyaknya Kluster	PCI
1.00	11	0.1227
1.10	9	0.3006
1.20	6	0.2233
1.30	3	0.4366
1.50	2	0.6651

Berdasarkan Tabel 2 nilai PCI tertinggi berada pada jari-jari 1.50 yang menghasilkan dua klaster. Oleh karena itu, pengelompokan provinsi menjadi dua klaster adalah pilihan terbaik. Hasil pengelompokannya disajikan dalam Tabel 3.

Tabel 3. Hasil pengelompokan dengan jari-jari sebesar 1.50

Klaster	Provinsi
Klaster 1	Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Banten, Jawa Barat, Jawa Tengah, Jawa Timur, Bali, DI. Yogyakarta, Gorontalo, Kalimantan Timur, Kalimantan Selatan, Kalimantan Barat, Kalimantan Utara
Klaster 2	Sulawesi Utara, Sulawesi Tengah, Sulawesi Barat, Aceh, Bengkulu, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Tengah, Sulawesi Selatan, Sulawesi Tenggara, Maluku, Maluku Utara, Papua Barat, Papua

4.3. Karakteristik Klaster Terbaik

Kesimpulan dari karakteristik setiap klaster yang terbentuk, diambil dari membandingkan rata-rata umum setiap peubah indikator pembangunan pendidikan dengan rata-rata peubah pada setiap klaster. Misalkan rata-rata umum setiap indikator pembangunan pendidikan dilambangkan dengan $\bar{X}_j$, dan rata-rata indikator pendidikan klaster dinotasikan dengan $\bar{X}_j^l$. Apabila $\bar{X}_j^l > \bar{X}_j$, maka peubah X_j diberi tanda positif (+). Sebaliknya, jika $\bar{X}_j^l < \bar{X}_j$ maka peubah diberi tanda negatif (-).

Apabila suatu indikator bernilai positif (+), artinya pembangunan pendidikan telah mencapai/melebihi rata-rata umum indikator pembangunan pendidikan nasional. Sebaliknya, apabila indikator ini bernilai negatif (-) artinya pembangunan pendidikan di provinsi tersebut kurang dari rata-rata nasionalnya. Pada penelitian ini nilai positif (+) untuk peubah-peubah rasio sekolah tingkat SD (X_1), SMP (X_2), SMA (X_3) dan SMK (X_4) dengan penduduk usia terkait; rasio guru layak mengajar terhadap peserta didik tingkat SD (X_5), SMP (X_6), SMA (X_7) dan SMK (X_8); rasio sekolah yang memiliki ketersediaan toilet terpisah tingkat SD (X_9), SMP (X_{10}), SMA (X_{11}) dan SMK (X_{12}); anak yang mengikuti pendidikan pra sekolah (X_{13}); APS usia 19-24 tahun (X_{14}) berarti bahwa pembangunan pendidikan di klaster tersebut berjalan baik. Namun, untuk peubah angka putus sekolah tingkat SD (X_{15}), SMP (X_{16}), dan SM (X_{17}) suatu klaster dikatakan pembangunan pendidikan berjalan baik ketika tanda untuk peubah-peubah tersebut negatif (-). Adapun karakteristik setiap peubah disajikan dalam Tabel 4.

Karakteristik peubah X_1 sampai X_4 menunjukkan bahwa rasio jumlah SD sampai SMK berdasarkan usia terkait lebih baik pada klaster kedua. Kemudian karakteristik rasio guru layak mengajar di sekolah umum terhadap peserta didik lebih baik juga pada klaster kedua. Pada rasio sekolah yang memiliki ketersediaan toilet

Tabel 4. Karakteristik hasil pengelompokan provinsi dengan metode SFCM

Peubah	Klaster 1	Klaster 2	Peubah	Klaster 1	Klaster 2
X_1	—	+	X_{10}	+	—
X_2	—	+	X_{11}	+	—
X_3	—	+	X_{12}	+	—
X_4	—	+	X_{13}	+	—
X_5	—	+	X_{14}	—	+
X_6	—	+	X_{15}	—	+
X_7	—	+	X_{16}	—	+
X_8	—	+	X_{17}	+	—
X_9	+	—			

yang terpisah, karakteristik klaster pertama lebih baik dari klaster kedua. Persentase anak usia 3-6 tahun yang mengikuti pendidikan pra sekolah lebih baik pada klaster pertama dibandingkan dengan klaster kedua. Namun, APS usia 19-24 tahun lebih baik pada klaster kedua. Terakhir, peubah yang menunjukkan angka putus sekolah pada klaster pertama memiliki karakteristik lebih rendah dari klaster kedua.

5. Kesimpulan

Berdasarkan hasil penelitian pengklasteran provinsi-provinsi di Indonesia menggunakan metode *Subtractive Fuzzy C-Means* (SFCM) dengan jumlah klaster terbaik adalah dua klaster dengan jari-jari (r) = 1.50. Berdasarkan karakteristik klaster, klaster kedua dapat dikatakan lebih baik dari klaster pertama. Klaster pertama dikatakan sebagai klaster terendah karena hampir semua indikator pembangunan pendidikannya rendah. Oleh karena itu, provinsi-provinsi di klaster pertama harus mendapat program/kebijakan lebih dalam meningkatkan pembangunan pendidikannya.

6. Ucapan Terima kasih

Penulis mengucapkan terima kasih kepada Ibu Lyra Yulianti, Ibu Izzati Rahmi H.G., dan Ibu Arrival Rince Putri yang telah memberikan masukan dan saran sehingga artikel ini dapat diselesaikan dengan baik.

Daftar Pustaka

- [1] Azizah, N., D. Yuniarti, R. Goejantoro, 2018, Penerapan metode *Fuzzy Subtractive Clustering* (Studi kasus: pengelompokan kecamatan di Provinsi Kalimantan Timur berdasarkan luas daerah dan jumlah penduduk tahun 2015, *Jurnal Ekponensial*, Vol **9**(2): 197 – 206
- [2] Johnson, R.A., Wichern, D.E., 2007, *Applied Multivariate Statistical Analysis*, Prentice Hall, New Jersey

- [3] Klir, G.J., Folger T.A., 1988, *Fuzzy Set, Uncertainty and Information*, Prentice Hall, New Jersey
- [4] Kurniawan, R., B.N., Haqiqi, 2015, Pengelompokan menggunakan metode *Subtractive Fuzzy C-Means* (SFCM), Studi kasus : demam berdarah di Jawa Timur, *Jurnal Statistika*, Vol. **3**(2): 22 – 30
- [5] Kusumadewi, S., Purnomo, H., 2004, *Logika Fuzzy Untuk pendukung keputusan*, Jilid 2, Graha Ilmu, Jakarta
- [6] Pal, N.R. dan Bezdek, J.C., 1995, On Cluster Validity for the fuzzy c-means model, *IEEE Transactions On Fuzzy Systems*, Vol. **3**(3): 370 – 379