Variance
Journal of Statistics and Its Applications
Volume 2 Nomor 2 | Desember 2020
p-ISSN: 2685-8738 e-ISSN: 2685-872X
doi : 10.30598/variancevol2iss2page63-74

PENERAPAN METODE *SUBTRACTIVE FUZZY C-MEANS* PADA TINGKAT PARTISIPASI PENDIDIKAN JENJANG SEKOLAH MENENGAH ATAS/SEDERAJAT DI KABUPATEN/KOTA PULAU KALIMANTAN TAHUN 2018

(*Application of the Fuzzy C-Means Subtractive Method at the Level of Education Participation at the Senior High School /Equivalent on District/City Kalimantan Island in 2018*)

Widya Suerni [1*], Memi Nor Hayati[2], Rito Goejantoro[3]
[1]Jurusan Matematika, Fakultas MIPA Universitas Mulawarman
[2] Lab. Statistika Terapan, Fakultas MIPA Universitas Mulawarman
[3] Lab. Statistika Komputasi, Fakultas MIPA Universitas Mulawarman
[1,2,3]Jln. Barong Tongkok No.4 Kampus Gunung Kelua, Samarinda – Kalimantan Timur 75123 Indonesia
e-mail:*widyasuerni15@gmail.com[1]*, meminorhayati@yahoo.com[2],ritogoejantoro@yahoo.com[3]*

Abstrak: Analisis *cluster* merupakan metode eksplorasi data yang digunakan untuk mendapatkan karakteristik yang tersembunyi dengan membentuk *cluster* data. Salah satu metode analisis *cluster* adalah *Subtractive Fuzzy C-Means* (SFCM). SFCM merupakan penggabungan dari metode *Subtractive Clustering* dan *Fuzzy C-Means*. Metode SFCM memiliki kelebihan yaitu tidak membutuhkan banyak iterasi, dan hasil yang didapatkan lebih stabil dan akurat dibandingkan dengan metode FCM dan SC. Penelitian ini bertujuan untuk mengetahui hasil *clustering* pada data angka partisipasi pendidikan jenjang Sekolah Menengah Atas (SMA)/sederajat. Data yang digunakan adalah data angka partisipasi pendidikan jenjang SMA/sederajat di kabupaten/kota di Pulau Kalimantan tahun 2018 menggunakan tiga variabel yaitu Angka Partisipasi Kasar (APK), Angka Partisipasi Sekolah (APS) dan Angka Partisipasi Murni (APM). Berdasarkan tiga indeks validitas yaitu *Partition Coefficient Index* (PCI), *Modified Partition Coefficient Index* (MPCI), dan Xie & Beni *Index* (XBI) pada metode SFCM diperoleh *cluster* yang optimal sebanyak dua *cluster*.

Kata Kunci: *Clustering,* pendidikan*, Subtractive Fuzzy C-Means*

Abstract: *Cluster analysis is a data exploration method uses to obtain hidden characteristics by forming data clusters. One of the cluster analysis methods is Subtractive Fuzzy C-Means (SFCM). SFCM is a combination of Subtractive Clustering and Fuzzy C-Means methods. The SFCM method has the advantages of not requiring many iterations and the results obtained are more stable and accurate than the FCM and SC methods. This study aims to determine the result of clustering on the enrollment rate data for Senior High School (SHS)/equivalent. The data used were the enrollment rate data for high school/equivalent level in the Regency/City of Kalimantan Island in 2018 using three variables, namely the Crude Participation Rate (CPR), the School Participation Rate (SPR) and the Net Enrollment Rate (NER). Based on the SFC three validity indices, namely Partition Coefficient Index (PCI) Validity Index, Modified Partition Coefficient Index (MPCI), and Xie & Beni Index (XBI) in the SFCM method, the optimal cluster were two clusters.*

Keywords**:** *clustering*, *education*, *Subtractive Fuzzy C-Means*

1. PENDAHULUAN

Analisis *cluster* merupakan metode eksplorasi data yang digunakan untuk mendapatkan karakteristik yang tersembunyi dengan membentuk kelompok/*cluster* data. Analisis *cluster* bertujuan mengelompokkan objek-objek dengan karakteristik yang mirip ke dalam satu *cluster* dan objek dengan karakteristik yang berbeda ke *cluster* yang lain dengan memaksimalkan kesamaan antar objek dalam satu *cluster* [1].

https://ojs3.unpatti.ac.id/index.php/variance/
Submitted:2020-11-04 *Accepted:2021-04-19*

Terdapat dua metode pengelompokan dalam analisis *cluster* yaitu metode berhirarki (*hierarchical methods*) dan metode tidak berhirarki (*nonhierarchical methods*). Pada proses pengelompokan (*clustering*) berhirarki atau tidak berhirarki, pembentukan kelompokan dilakukan sedemikian rupa sehingga setiap objek berada tepat pada satu kelompok. Akan tetapi pada suatu saat hal itu tidak dapat dilakukan, karena sebenarnya objek tersebut terletak di antara dua atau lebih kelompok yang lain. Oleh karena itu perlu dilakukan pengelompokan dengan menggunakan *Fuzzy Clustering* di mana dalam melalukan pengelompokan mempertimbangkan tingkat keanggotaan himpunan *Fuzzy* sebagai dasar pembobotan [2].

Ada beberapa algoritma *Fuzzy Clustering*, salah satu diantaranya adalah *Fuzzy C-Means* (FCM). FCM adalah suatu teknik pengelompokan data yang mana keberadaan tiap-tiap data dalam satu *cluster* ditentukan oleh fungsi keanggotaan [3]. Metode FCM mempunyai beberapa kelemahan, antara lain membutuhkan banyaknya kelompok dan matriks keanggotaan kelompok yang sudah ditetapkan sebelumnya [4]. Matriks keanggotaan kelompok awal diinisialisikan secara acak yang menyebabkan metode FCM memiliki masalah inskonsistensi.

Alternatif metode *cluster* yang lain yang dapat digunakan jika jumlah kelompok tidak diketahui sebelumnya adalah metode *Subtractive Cluster* (SC). SC memperoleh hasil yang lebih konsisten dibandingkan dengan FCM [5]. Selain itu SC memiliki kecepatan yang lebih baik dibandingkan FCM, namun SC memiliki akurasi yang lebih rendah dibandingkn FCM [6].

Untuk menghubungkan masalah kekurangan dan kelebihan kedua metode tersebut diusulkan sebuah metode baru yang merupakan penggabungan dari keduanya disebut *Subtractive Fuzzy C-Means* (SFCM). Metode ini digunakan Liu, Xiao, Wang, Shi, & Fang, S [7] dalam penelitiannya dan menyimpulkan bahwa SFCM memberikan solusi yang lebih baik dibandingkan dengan FCM serta memberikan kecepatan yang lebih tinggi dalam konvergensi fungsi objektif. [6] Kelebihan lainnya pada metode SFCM yaitu berdasarkan indeks validitas *Partition Coefficient* (PC) menyimpulkan bahwa SFCM dapat meningkatkan kecepatan, mengurangi jumlah iterasi, menghasilkan partisi data yang lebih stabil dan lebih akurat.

Pendidikan adalah salah satu aspek penting untuk mewujudkan perkembangan sumber daya manusia yang berkualitas. Dengan pendidikan, seseorang bisa mengembangkan kemampuan, bakat, serta kreativitas yang dimilikinya. Kesadaran pentingnya pendidikan dapat dilihat melalui partisipasi masyarakat dalam menyekolahkan anaknya. Akan tetapi tidak semua masyarakat sadar dan ikut mendukung program pemerintah tersebut, terutama masyarakat di daerah pedesaan yang kurang memperhatikan pentingnya pendidikan. Menurut BPS [8] berdasarkan tipe daerah, pada tahun 2018 masih terdapat sekitar 8,57% penduduk Indonesia usia 5 tahun ke atas di perdesaan yang belum pernah sekolah, dibandingkan di perkotaan yang hanya sekitar 5,16% penduduk yang belum sekolah. Melihat hal tersebut, tentunya perhatian pemerintah harus lebih ekstra dalam mensosialisasikan pentingnya pendidikan di daerah-daerah yang tingkat partisipasi pendidikannya masih rendah.

Tingkat pastisipasi sekolah untuk jenjang SMA/sederajat di Indonesia tahun 2018 hanya 71,99% paling rendah dibandingkan tingkat partisipasi sekolah SD yaitu 99,22% dan SMP yaitu 95,36%. Sedangkan Tingkat partisipasi sekolah untuk jenjang SMA/sederajat di Kalimantan Timur 81,55%, di Kalimantan Utara 75,62%, di Kalimantan Barat 68,35%, di Kalimantan Selatan 68,66%, dan di Kalimantan Tengah 66,95%. Dilihat dari tingkat partisipasi sekolah jenjang SMA/sederajat di Provinsi Kalimantan masih terdapat provinsi yang dibawah tingkat partisipasi sekolah di Indonesia yaitu Kalimantan Barat, Kalimantan Selatan dan Kalimantan Tengah. Oleh karena itu perlu dilakukannya pengelompokan di daerah-daerah berdasarkan tingkat partisipasi pendidikan sehingga dapat dilihat pengelompokan daerah-daerah dengan tingkat partisipasi pendidikannya yang masih rendah dan daerah-daerah yang tingkat partisipasi pendidikan untuk jenjang SMA/sederajat yang sudah tinggi. Tujuan Penelitian ini adalah untuk mengkaji hasil metode SFCM pada kabupaten/kota di Pulau Kalimantan berdasarkan tingkat partisipasi pendidikan jenjang SMA/sederajat tahun 2018. Dari hasil pengelompokan yang diperoleh dapat membantu pemerintah dalam mengelompokkan kabupaten/kota di Pulau Kalimantan berdasarkan partisipasi pendidikan jenjang SMA/sederajat.

2. METODOLOGI

2.1. Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian merupakan data sekunder, yaitu data yang diperoleh dari Buku Publikasi BPS untuk data APK & APS jenjang SMA/sederajat tahun 2018 dan Buku Publikasi Kemendikbud untuk data APM jenjang SMA/sederajat tahun 2018. Adapun variabel dalam penelitian ini adalah sebagai berikut [8]:

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Satuan
Angka Partisipasi Sekolah	Proporsi penduduk pada kelompok umur jenjang pendidikan tertentu yang masih bersekolah terhadap penduduk pada kelompok umur tersebut.	%
Angka Partisipasi Kasar	Proporsi jumlah penduduk yang sedang bersekolah pada suatu jenjang pendidikan terhadap jumlah penduduk umur sekolah yang sesuai dengan jenjang pendidikan tersebut.	%
Angka Partisipasi Murni	Menunjukkan seberapa banyak penduduk usia sekolah yang sudah dapat memanfaatkan fasilitas pendidikan sesuai pada jenjang pendidikannya.	%

2.2. Metode Analisis

***Fuzzy C-Means* (FCM)**

FCM merupakan teknik *clustering* data di mana keberadaan tiap titik-titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaan. Konsep dasar FCM pertama kali yaitu menentukan pusat *cluster*, yang akan menandai lokasi rata-rata untuk tiap-tiap *cluster*. Dengan memperbaiki pusat *cluster* derajat keanggotaan setiap titik data secara berulang, maka pusat *cluster* akan menuju ke tempat yang tepat. Perulangan tersebut didasarkan pada minimalisasi fungsi objektif yang menggambarkan jarak antara titik data dengan pusat *cluster* yang terboboti oleh derajat keanggotaan [3]. Algoritma FCM adalah sebagai berikut:

1. Penentuan jumlah *cluster (p)*, pangkat (*w*), maksimum iterasi (*MaxIter*), *error* terkecil yang diharapkan (ε), fungsi objektif awal ($J_0 = 0$).

2. Membangkitkan bilangan random u_{il} dengan i = 1,2,..*n* (banyaknya data) dan *l*=1,2,..*p* sebagai banyak *cluster* sebagai elemen-elemen awal matriks keanggotaan awal U.

$$\mu_{ij} = \frac{\mu_{ij}}{\sum_{l=1}^{p} \mu_{lj}} \tag{1}$$

3. Menghitung pusat *cluster* ke-*l* dengan persamaan

$$C_{lj} = \frac{\sum_{i=1}^{n}\left((\mu_{il})^{w} x_{ij}\right)}{\sum_{i=1}^{n}(\mu_{il})^{w}} \tag{2}$$

4. Menghitung fungsi objektif pada iterasi ke-*t* dengan persamaan

$$J_t = \sum_{i=1}^{n}\sum_{l=1}^{p}\left(\left[\sum_{j=1}^{m}(x_{ij} - C_{lj})^2\right](\mu_{ij})^{w}\right) \tag{3}$$

5. Menghitung perubahan matriks keanggotaan dengan persamaan:

$$\mu_{ij} = \frac{\left[\sum_{j=1}^{m}(x_{ij} - C_{lj})^2\right]^{-\frac{1}{w-1}}}{\sum_{l=1}^{p}\left[\sum_{j=1}^{m}(x_{ij} - C_{lj})^2\right]^{-\frac{1}{w-1}}} \quad (4)$$

6. Memeriksa kondisi berhenti
   a. Jika $|J_t - J_{t-1}| < \varepsilon$ atau $t > MaxIter$ maka berhenti
   b. Jika tidak : $t = t+1$, ulangi langkah ke-3

Subtractive Clustering

Konsep dasar SC adalah menentukan daerah-daerah dalam suatu variabel yang memiliki potensi tinggi terhadap titik-titik di sekitarnya. Titik dengan jumlah tetangga terbanyak akan dipilih sebagai pusat *cluster*. Titik yang sudah terpilih sebagai pusat *cluster* ini kemudian dikurang dengan titik potensinya. Kemudian algoritma akan memilih titik lain yang memiliki tetangga terbanyak untuk dijadikan pusat *cluster* yang lain. Hal ini akan dilakukan berulang-ulang sampai semua titik diuji. Pada metode ini jumlah *cluster* belum diketahui [3]. Algoritma SC untuk mencari pusat *cluster* adalah sebagai berikut:

1. Penginputan data (x_{ij}) dengan sampel ke-i $(i = 1, 2, \dots, n)$ dan atribut ke-j $(j = 1, 2, \dots, m)$.
2. Penetapan nilai jari-jari (r), *Squash factor* (q), *accept ratio* (ar), *reject ratio*(rr), nilai minimum dan maksimum data dari masing-masing atribut.
3. Penormalisasian setiap data dengan persamaan:

$$x_{ij_{norm}} = \frac{x_{ij} - x_{j_{min}}}{x_{j_{max}} - x_{j_{min}}} \quad (5)$$

4. Penentuan potensi awal tiap-tiap titik data
   a. $i = 1$
   b. Pengerjaan hingga $i = n$
      - $T_j = x_{kj_{norm}}$; $j = 1, 2, \dots, m$
      - Perhitungan:

$$Dist_{ij} = \frac{T_j - x_{ij_{norm}}}{r};\ i, k = 1, 2, \dots, n;\ j = 1, 2, \dots, m \quad (6)$$

      - Penentuan potensi awal tiap data dengan persamaan:

$$\text{Jika } m = 1, \text{ maka } D_k = \sum_{i=1}^{n} e^{-4\left(Dist_{ij}^2\right)} \quad (7)$$

$$\text{Jika } m > 1, \text{ maka } D_k = \sum_{i=1}^{n} e^{-4\left(\sum_{j=1}^{m} Dist_{ij}^2\right)} \quad (8)$$

5. Pencarian titik dengan potensi tertinggi

$$M = \max[D_k | k = 1, 2, \dots, n]; \text{ untuk iterasi ke-1} \quad (9)$$

$$Z = \max[D_k | k = 1, 2, \dots, n]; \text{ untuk iterasi ke-2, 3, dan seterusnya} \quad (10)$$

6. Penentuan pusat *cluster* dan kurangi potensinya terhadap titik-titik di sekitarnya dengan cara menetapkan vektor V sebagai calon pusat *cluster* dan menghitung nilai *ratio*.

$$R = \frac{Z}{M} \quad (11)$$

khusus pada iterasi pertama nilai Z=M.

Setelah nilai *ratio* diperoleh, ada 3 keadaan yang dapat terjadi:

1) Jika nilai *ratio* > *accept ratio*, calon pusat *cluster* dapat diterima sebagai pusat *cluster* baru dan ditulis C_l. Prosedur kerja yang harus dilakukan selanjutnya setelah mendapat *cluster* baru adalah mengurangi potensi titik-titik data yang lain dengan cara:

a. $$S_{ij} = \frac{C_{lj_{norm}} - x_{ij_{norm}}}{r * q};\ i = 1, 2, ..., n;\ j = 1, 2, ..., m \quad (12)$$

b. $$D_{C_{li_{norm}}} = M \times e^{-4\left[\sum_{j=1}^{m} (S_{ij})^2\right]};\ i = 1, 2, ..., n \quad (13)$$

c. $$D_i^t = D_i^{t-1} - D_{C_{li_{norm}}} \quad (14)$$

2) Jika nilai *ratio* < *accept ratio* dan nilai *ratio* > *reject ratio*, calon baru akan diterima jika keberadaannya cukup jauh dari pusat *cluster* yang telah ada. Prosedur dalam keadaan ini:

a) Misalkan $Md = -1$;

b) Mengerjakan untuk $l = 1$ sampai $l = p$; p = banyaknya *cluster*

$$Sd_l = \sum_{j=1}^{m} \left(\frac{V_j - C_{lj_{norm}}}{r} \right)^2;\ j = 1, 2, ..., m \quad (15)$$

Jika ($Md < 0$) atau ($Sd<Md$), maka $Md = Sd_l$

Jika (Sd_l>Md), maka *Md* tidak berubah

$$Mds = \sqrt{Md} \quad (16)$$

dengan *Mds* adalah jarak terdekat data calon pusat *cluster* dengan pusat *cluster*.

Jika (*ratio* + *Mds*) ≥ 1, calon pusat *cluster* diterima sebagai pusat *cluster* baru.

Jika (*ratio* + *Mds*) < 1, calon pusat *cluster* tidak diterima dan tidak akan dipertimbangkan kembali sebagai pusat *cluster* dan potensi data tersebut diset menjadi 0.

3) Jika nilai *ratio* < *accept ratio* dan nilai *ratio* < *reject ratio*, sudah tidak ada lagi calon pusat *cluster* baru dan iterasi dihentikan.

7. Pengembalian pusat *cluster* dari bentuk ternormalisasi ke bentuk semula (denormalisasi)

$$C_{lj} = C_{lj_{norm}} * (x_{\max_j} - x_{\min_j}) + x_{\min_j} \quad (17)$$

Subtractive Fuzzy C-Means

Metode *Subtractive Fuzzy C-Means* (SFCM) merupakan penggabungan antara metode *clustering Subtractive Clustering* (SC) dan *Fuzzy C-Means* (FCM). SC digunakan untuk menentukan jumlah *cluster* dan pusat awal *cluster* FCM, sehingga tidak perlu melakukan inisialisasi secara acak. Kemudian FCM digunakan untuk mendapatkan pusat *cluster* dan nilai keanggotaan data pada suatu *cluster*. Nilai keanggotaan untuk menentukan penempatan data ke dalam *cluster* [7]. Langkah metode SFCM adalah sebagai berikut:

1. Penormalisasian setiap data dengan Persamaan (5).
2. Penentuan potensi awal tiap-tiap data yang telah dinormalisasi menggunakan menggunakan Persamaan (7) dan (8)
3. Pencarian titik dengan potensi tertinggi dengan menggunakan persamaan (9) dan (10).

4. Penentuan pusat *cluster* dan kurangi potensinya terhadap titik-titik di sekitarnya dengan cara menetapkan vektor V sebagai calon pusat *cluster* dan menghitung nilai *ratio* menggunakan Persamaan (11). kemudian *ratio* dibandingkan dengan *accept ratio* dan *reject ratio.* Jika nilai *ratio* >*accept ratio*, calon pusat *cluster* dapat diterima sebagai pusat *cluster* baru. Langkah selanjutnya setelah mendapat *cluster* baru adalah mengurangi potensi titik-titik data yang lain dengan Persamaan (14). Jika nilai *ratio* < *accept ratio* dan nilai *ratio* > *reject ratio*, calon baru akan diterima jika keberadaannya cukup jauh dari pusat *cluster* yang telah ada. Jika nilai *ratio* < *accept ratio* dan nilai *ratio* < *reject ratio*, sudah tidak ada lagi calon pusat *cluster* baru dan iterasi dihentikan.
5. Pengembalian pusat *cluster* dari bentuk ternormalisasi ke bentuk semula (denormalisasi) dengan Persamaan (17).
6. Perhitungan matriks keanggotaan awal FCM dengan Persamaan (4) berdasarkan pusat *cluster* yang didapatkan melalu metode SC.
7. Perhitungan fungsi objektif pada iterasi ke-*t* dengan Persamaan (3).
8. Pembaharuan pusat *cluster* dengan Persamaan (2).
9. Perhitungan perubahan matriks keanggotaan dengan Persamaan (4).
10. Memeriksa kondisi berhenti
    a. Jika $\left|J_t - J_{t-1}\right| < \varepsilon$ atau t >*Maxlter* maka berhenti
    b. Jika tidak : $t=t+1$, ulangi langkah ke 10.

Indeks Validitas

Metode *clustering* yang menggunakan konsep *fuzzy*, sebuah data dapat menjadi anggota pada semua *cluster* dengan nilai derajat keanggotaan yang dimilikinya. Hal itu berarti data tidak mutlak menjadi anggota satu *cluster* saja, tetapi juga mempunyai nilai untuk menjadi anggota *cluster* yang lain. Oleh karena itu digunakan indeks validatas untuk mengatasi masalah tersebut. Indeks validitas yang sering digunakan dalam penelitian-penelitian adalah [9]

1. *Partition Coefficient Index* (PCI)

   PCI hanya mengevaluasi nilai derajat keanggotaan tanpa memandang nilai data. Nilai PCI yang semakin besar menandakan bahwa kualitas *cluster* yang didapatkan semakin baik. Berikut adalah persamaan untuk menghitungnya.

$$PCI = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p}\left(\mu_{ij}\right)^2 \tag{18}$$

2. *Modified Partition Coefficient Index* (MPCI)

   Nilai PCI memiliki kecendrungan monotonik terhadap *p*. Modifikasi nilai PCI (MPCI) dilakukan oleh Dave (1996) terhadap kecenderungan monotonik tersebut. Nilai MPCI yang didapatkan berada pada rentang [0,1], nilai MPCI yang semakin besar (mendekati 1) mempunyai arti bahwa kualitas *cluster* yang didapatkan semakin baik. Persamaan yang digunakan yang adalah:

$$MPCI = 1 - \frac{p}{p-1}(1 - PCI) \tag{19}$$

3. Xie & Beni *Index* (XBI)

   Xie & Beni [10] juga mengusulkan validitas untuk mengevaluasi *cluster* yang didapat dengan modifikasi oleh Pal dan Bezdek [11]. Secara umum, nilai yang terbaik untuk Xie & Beni *Index* (XBI) adalah nilai indeks yang semakin kecil. Nilai XBI yang semakin kecil mempunyai arti kualitas hasil pengelompokan yang semakin baik. Persamaan yang digunakan seperti berikut:

$$\text{XBI} = \sum_{j=1}^{p}\sum_{i=1}^{n}\frac{u_{ij}^{w}\times\left[\sum_{j=1}^{m}(x_{ij}-C_{lj})^2\right]}{n\times\min\left(\left[\sum_{j=1}^{m}(\text{C}_{fj}-C_{lj})^2\right]\right)} \quad l, f=1,2,..,p\ ;\ j=1,2,\ldots m \tag{20}$$

3. HASIL DAN PEMBAHASAN

3.1. Statistika Deskriptif

Analisis statistika deskriptif ini dilakukan untuk mengetahui gambaran umum data tingkat partisipasi pendidikan jenjang SMA/sederajat di kabupaten/kota Pulau kalimantan tahun 2018.

Tabel 2. Deskriptif APK, APS, APM

Ukuran Data	APK	APS	APM
Minimum	49,18	44,37	33,59
Maksimum	106,67	85,72	80,17
Rata-Rata	86,32	70,80	64,74
Simpangan Baku	12,09	8,73	9,60

3.2. Penentuan Nilai-nilai Parameter

Pada penelitian ini nilai-nilai parameter yang digunakan yaitu jari-jari (*r)* = 0,3; 0,4; 0,5 dan 0,6, *squash factor (q)* =1,25, *accept ratio (ar)* = 0,5, *reject ratio (rr)* = 0,15, pangkat pembobot(*w*) = 2, maksimum iterasi (*MaxIter*) = 1000 dan nilai *error* yang diharapkan (ε)= 0,001.

3.3. Penormalisasian Data

Langkah awal dalam proses pengelompokan dengan metode SFCM adalah normalisasi data. Normalisasi data dapat dihitung menggunakan Persamaan (5). Perhitungan normalisasi dilakukan pada data untuk setiap variabel dapat dilihat pada tabel berikut.

Tabel 3. Hasil Normalisasi Data

No	Kabupaten/Kota	APK	APS	APM
1	Sambas	0,5992	0,5826	0,7261
2	Mempawah	0,8841	0,6936	0,8700
⋮	⋮	⋮	⋮	⋮
56	Tarakan	0,5748	0,7712	0,6685

3.4. Penentuan Potensi Awal

Setelah menormalisasi data, langkah selanjutnya adalah menghitung potensi awal. Karena m > 1, maka perhitungan nilai potensi data dapat dihitung menggunakan Persamaan (8). Contoh perhitungan menggunakan nilai r = 0,5. Perhitungan untuk mencari potensi awal dapat dilihat pada Tabel 4 berikut.

Tabel 4. Nilai Potensi Awal

No. D	Potensi Awal
1	15,5047
2	15,4960
⋮	⋮
45	18,1221
⋮	⋮
56	14,9804

Langkah selanjutnya adalah mencari potensi tertinggi. Pada Tabel 4 dapat dilihat bahwa nilai potensi tertinggi terletak pada D_{45} yaitu 18,1221. Sehingga nilai tersebut menjukkan nilai M = 18,1221.

3.5. Penentuan Pusat *Cluster*

Data ke-45 akan dipilih menjadi calon pusat *cluster* karena memiliki nilai potensi tertinggi. Pada iterasi pertama didapatkan M =18,1221. Karena pada iterasi pertama nilai Z=M maka untuk iterasi pertama Z = 18,1221. Kemudian dilakukan perhitungan nilai *ratio* menggunakan Persamaan (5). Pada iterasi pertama didapatkan nilai *ratio* =1. Karena nilai *ratio* =1 >*accept ratio* = 0,5 maka calon pusat *cluster* akan diterima sebagai pusat *cluster*. Langkah selanjutnya adalah mengurangi potensi titik data dengan menggunakan Persamaan (6). Potensi data yang terbaru adalah sebagai berikut.

Tabel 5. Nilai Potensi Terbaru

No. D	Potensi Awal
1	5,8447
2	-0,3762
⋮	⋮
17	10,2240
⋮	⋮
56	6,0939

Berdasarkan Tabel 5 terlihat bahwa potensi tertinggi sebesar 10,2240 pada data kabupaten/kota ke-17. Dengan demikian data ke-17 ditetapkan menjadi calon pusat *cluster* kedua. Untuk menentukan pusat *cluster* kedua dan seterusnya digunakan nilai *ratio* yaitu hasil bagi nilai potensi terbaru dengan nilai potensi awal. Pada iterasi ke 2 ini, untuk nilai Z = 10,2240 sedangkan nilai M = 18,1221. Kemudian mencari kembali nilai *ratio* menggunakan Persamaan (11), sehingga diperoleh nilai *ratio* sebesar 0,5642. Karena *ratio* = 0,5642 > *accept ratio* = 0,5*,* maka data ditetapkan menjadi pusat *cluster* kedua. Untuk menentukan pusat *cluster* berikutnya dilakukan dengan cara yang sama. Penentuan pusat *cluster* dihentikan ketika *ratio* < *accept ratio* dan *ratio* < *reject ratio*. Dengan bantuan *Software R*x64 3.4.4 didapatkan hasil sebanyak 3 pusat *cluster* pada r = 0,5 seperti ditunjukkan pada Tabel 6 berikut.

Tabel 6. Tiga Pusat *Cluster* dan Satu Calon Pusat *Cluster*

Data	Nilai Potensi	APK	APS	APM
45	18,1221	0,7873	0,7173	0,8152
17	10,2240	0,4324	0,6641	0,4341
3	6,0485	0,5965	0,2888	0,6769

Langkah selanjutnya adalah pengembalian pusat *cluster* dari bentuk ternormalisasi ke bentuk semula (dernormalisasi) dengan menggunakan Persamaan (17). Sehingga diperoleh hasil pusat *cluster* yang telah didernormalisasi seperti pada Tabel 7 sebagai berikut/

Tabel 7. Elemen Pusat *Cluster* Hasil Denormalisasi dengan *r* = 0,5

Data	APK	APS	APM
45	94,44	74,03	71,56
17	74,04	71,83	53,81
3	83,47	56,31	65,12

Selanjutnya dilakukan penentuan banyaknya *cluster* dengan cara yang sama untuk r = 0,3; r = 0,4 dan r = 0,6. Banyaknya *cluster* yang dihasilkan dengan r = 0,3; r = 0,4 dan r = 0,6 dapat dilihat pada Tabel 8 berikut.

Tabel 8. Banyaknya *cluster* dengan jari-jari yang berbeda

Jari-jari	Banyaknya *Cluster*
0,3	6
0,4	4
0,5	3
0,6	2

3.6. Penentuan Matriks Keanggotaan

Pusat *cluster* yang telah didapatkan menggunakan metode *Subtractive Clustering,* kemudian digunakan untuk menghitung nilai keanggotaan awal menggunakan metode *Fuzzy C-Means* dengan Persamaan (4). Nilai keanggotaan awal menggunakan $r = 0,5$ dapat dilihat pada Tabel 9 sebagai berikut.

Tabel 9. Nilai Keanggotaan Awal

No	Nilai Keanggotaan		
	***Cluster* 1**	***Cluster* 2**	***Cluster* 3**
1	0,3770	0,0008	0,4071
2	0,9124	0,0000	0,0553
3	0,0000	0,0000	1,0000
⋮	⋮	⋮	⋮
56	0,4033	0,0019	0,2028

Selanjutnya nilai keanggotaan pada Tabel 9 digunakan untuk menghitung nilai fungsi objektif untuk t =1 dengan Persamaan (10). Nilai fungsi objektif untuk t =1 didapatkan nilai sebesar 2.782,5820. Nilai fungsi objektif awal $J_0 = 0$ maka perubahan fungsi objektifnya adalah 2.782,5820. $|J_1 - J_0| = 2.782,5820 > \varepsilon = 0,001$. Karena perubahan fungsi objektif masih lebih besar dari nilai ε, maka proses dilanjutkan ke iterasi berikutnya.

3.7. Perbaruan Pusat *Cluster*

Langkah selanjutnya adalah memperbarui pusat *cluster* dengan Persamaan (2). Sehingga pusat *cluster* yang telah diperbarui dapat dilihat pada Tabel 10 sebagai berikut.

Tabel 10. Elemen Pusat *Cluster* yang Diperbarui dengan r = 0,5

Data	APK	APS	APM
45	95,5903	75,1974	72,0980
17	74,0301	71,8900	53,8079
3	85,5584	59,7300	64,0056

3.8. Perhitungan Perubahan Matriks Keanggotaan

Setelah menghitung pusat *cluster*, langkah selanjutnya adalah menghitung fungsi objektif dengan pusat *cluster* yang telah diperbarui dan menghitung perubahan matriks keanggotaan. Iterasi berhenti ketika $|J_t - J_{t-1}| < 0,001$ dan $t < 1000$. Pada penelitian ini langkah berhenti pada iterasi ke-31. Diperoleh hasil akhir nilai keanggotaan akhir yang ditunjukkan pada Tabel 11 sebagai berikut.

Tabel 11. Nilai Keanggotaan dalam Tiga *Cluster*

No	Nilai Keanggotaan			***Cluster* yang Diikuti**
	***Cluster* 1**	***Cluster* 2**	***Cluster* 3**	
1	0,0791	0,0496	0,8712	3
2	0,9532	0,0109	0,0359	1
3	0,1057	0,1117	0,7826	3
⋮	⋮	⋮	⋮	⋮
56	0,2382	0,2226	0,5392	3

Selain perhitungan pada jari-jari (r) = 0,5 dilakukan juga langkah perhitungan yang sama terhadap r = 0,3; r = 0,4; dan r = 0,6. Sehingga diperoleh nilai keanggotan dengan r = 0,3 dengan enam *cluster* berikut.

Tabel 12. Nilai Keanggotaan dalam Enam *Cluster*

No	Nilai Keanggotaan						***Cluster* yang Diikuti**
	***Cluster* 1**	***Cluster* 2**	***Cluster* 3**	***Cluster* 4**	***Cluster* 5**	***Cluster* 6**	
1	0,0626	0,6205	0,1306	0,0362	0,0601	0,0901	2
2	0,9566	0,0053	0,0047	0,0013	0,0301	0,0020	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
56	0,0263	0,8412	0,0263	0,0283	0,0389	0,0390	2

Selanjutnya untuk nilai keanggotan dengan r = 0,4 dengan empat *cluster* adalah sebagai berikut.

Tabel 13. Nilai Keanggotaan dalam Empat *Cluster*

No	Nilai Keanggotaan				*Cluster* yang Diikuti
	Cluster 1	*Cluster* 2	*Cluster* 3	*Cluster* 4	
1	0,0721	0,7102	0,1722	0,0455	2
2	0,9458	0,0256	0,0218	0,0067	1
⋮	⋮	⋮	⋮	⋮	⋮
56	0,0410	0,8824	0,0383	0,0384	2

Selanjutnya untuk nilai keanggotan dengan r = 0,6 dengan dua *cluster* adalah sebagai berikut.

Tabel 14. Nilai Keanggotaan dalam Dua *Cluster*

No	Nilai Keanggotaan		*Cluster* yang Diikuti
	Cluster 1	*Cluster* 2	
1	0,5133	0,4867	1
2	0,9669	0,0331	2
⋮	⋮	⋮	⋮
56	0,4326	0,5674	2

3.9. Menghitung Indeks Validitas

Pada penelitian ini dilakukan validitas dengan *Partition Coefficient Index* (PCI), *Modified Partition Coefficient Index* (MPCI), dan Xie&Beni *Index* (XBI). untuk perhitungannya menggunakan r = 0,5. Nilai PCI, MPCI, dan XBI untuk r = 0,3; r =0,4; r =0,5 dan r =0,6 dapat ditunjukkan pada Tabel 15. sebagai berikut:

Tabel 15. Nilai PCI, MPCI, dan XBI

Jari-jari	Banyaknya *Cluster*	Indeks Validitas		
		PCI	MPCI	XBI
0,3	6	0,5138	0,4166	0,4041
0,4	4	0,6077	0,4769	0,2133
0,5	3	0,6592	0,4887	0,1978
0,6	2	0,7514	0,5029	0,1867

Berdasarkan Tabel 15. terlihat bahwa nilai PCI yang terbesar yaitu 0,7514 sehingga banyaknya *cluster* yang mempunyai nilai PCI sebesar 0,7514 yaitu dua *cluster* merupakan *cluster* yang optimal. Nilai MPCI yang terbesar yaitu 0,5029 sehingga banyaknya *cluster* yang mempunyai nilai MPCI sebesar 0,5029 yaitu dua merupakan *cluster* optimal. Nilai XBI yang terkecil yaitu 0,1867 sehingga banyaknya *cluster* yang mempunyai nilai XBI sebesar 0,1867 yaitu dua merupakan *cluster* yang optimal. Dari ketiga indeks validitas yaitu PCI, MPCI dan XBI maka dapat disimpulkan bahwa banyaknya *cluster* yang optimal digunakan adalah dua *cluster*.

Selanjutnya berdasarkan hasil dua *cluster* diperoleh statistik deskriptif yang ditunjukkan pada Tabel 16.

Tabel 16. Statistik Deskriptif untuk Dua *Cluster*

Cluster	Jumlah	Rata-rata		
		APK	APS	APM
1	32	94,7181	72,2394	71,2384
2	24	75,1125	68,8846	56,075

Dari Tabel 16 dapat diketahui bahwa pada *cluster* pertama jumlah kabupaten/kotanya adalah 32 kabupaten/kota, rata-rata nilai APK sebesar 94,7181 %, rata-rata nilai APS sebesar 72,2394 %, dan rata-rata nilai APM sebesar 71,2384 %.

Pada *cluster* dua jumlah kabupaten/kotanya adalah 24 kabupaten/kota, rata-rata nilai APK sebesar 75,1125 %, rata-rata nilai APS sebesar 68,8846%, dan rata-rata nilai APM sebesar 56,075 %. Jika dilihat pada nilai rata-rata, nilai rata-rata untuk APK, APS, dan APM jenjang SMA/sederajat pada *cluster* satu lebih tinggi dibandingkan dengan rata-rata APK, APS, dan APM jenjang SMA/sederajat yang terdapat pada *cluster* dua.

4. KESIMPULAN

Berdasarkan hasil dan pembahasan yang diperoleh maka dapat disimpulkan bahwa metode SFCM yang digunakan pada data tingkat partisipasi pendidikan di Kabupaten/Kota Pulau Kalimantan diperoleh *cluster* optimal sebanyak dua *cluster* dengan menggunakan jari-jari (r) = 0,6. Terdapat 32 yang kabupaten/kota yang masuk dalam *cluster* satu dan terdapat 24 yang termasuk dalam *cluster* dua. Dilihat pada nilai rata-rata, nilai rata-rata untuk APK, APS, dan APM jenjang SMA/sederajat pada *cluster* satu lebih tinggi dibandingkan dengan rata-rata APK, APS, dan APM jenjang SMA/sederajat yang terdapat pada *cluster* dua.

DAFTAR PUSTAKA

[1] S. Pramana, B. Yuniarto, M. Siti, I. Santoso and R. Nooraeni, *Data Mining dengan R*, Bogor: In Media, 2018.

[2] J.-S. R. Jang, S. C-T and E. Mizutani, Neuro-Fuzzy and Soft Computing: *A Computational Approach to Learning and Machine Intelligence*, New Jersey: Prentice-Hall Inc., 1997.

[3] S. Kusumadewi and H. Purnomo, *Aplikasi Logika Fuzzy untuk Pendukung Keputusan, Edisi 2,* Yogyakarta: Graha Ilmu, 2010.

[4] T. Le and T. Altman, "*Initialization Method for the Fuzzy C-Means Algorithm using Fuzzy Subtractive Clustering*," in International Conference on Information and Knowledge Engineering, Las Vegas, 2011.

[5] K. M. Bataineh and M. Saqer, "*A Comparison Study Between Various Fuzzy Clustering Algorithms*," Jourdan Journal of Medical and Industrial Engineering, vol. V, no. 4, pp. 335-343, 2011.

[6] J. Hossen, A. Rahman, S. Sayeed, K. Samsuddin and F. Rokhani, "A *Modified Hybrid Fuzzy Clustering Algorithm for Data Patitions*," Australian Journal of Basic and Applied Sciences, vol. V, no. 8, pp. 674-681, 2011.

[7] W. Y. Liu, C. J. Xiao, B. W. Wang, Y. Shi and S. F. Fang, "*Study on Combining Subtractive Clustering with Fuzzy C-Means Clustering*," in International Conference on Mchine Learning & Cybernetic, Xi'an, 2003.

[8] Badan Pusat Statistik, "Potret Pendidikan Indonesia Statistik Pendidikan 2018 Katalog Nomor 4301002," Badan Pusat Statistik, Jakarta, 2018.

[9] E. Prasetyo, *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*, Yogyakarta: ANDI, 2014.

[10] X. L. Xie and G. Beni, "*A Validity Measure for Fuzzy Clustering*," in IEEE Transactions on Pattern Analysis and Machine Intelligence, United States, 1991.

[11] N. R. Pal and J. C. Bezdek, "*On Cluster Validity for The Fuzzy C-Mean*s," in IEEE Transactions on Information Theory, United States, 1991.