

Implementasi *Fuzzy C-Means* dalam Pengelompokan Provinsi di Indonesia untuk Pemerataan Kualitas Pendidikan

Galang Sumantri¹, Mardhani Dwi Novianto², Pusparani Puan Prihastuti³
Universitas Negeri Yogyakarta^{1,2,3}
galangsumantri.2019@student.uny.ac.id

Abstrak— Pendidikan berperan penting dalam membangun sumber daya manusia yang unggul dan berdaya saing. Namun, adanya pandemi COVID-19 telah menyebabkan hilangnya pembelajaran (*learning loss*) di kalangan peserta didik dan tenaga pendidik sehingga perlu dilakukan berbagai upaya pemulihan terkait hal tersebut. Selain itu, upaya pemulihan juga dilakukan untuk mendukung terlaksananya pendidikan berkualitas yang menjadi target dalam *Sustainable Development Goals* (SDGs) poin ke-empat. Namun, upaya tersebut juga perlu memperhatikan tingkat kualitas pendidikan di Indonesia yang berbeda-beda di setiap provinsinya. Oleh karena itu, diperlukan suatu pengelompokan atau klasterisasi provinsi di Indonesia berdasarkan indikator kualitas pendidikan untuk mempermudah pemerintah dalam upaya pemerataan kualitas pendidikan. Salah satu algoritma pengelompokan yang dapat digunakan adalah *Fuzzy C-Means*. *Fuzzy C-Means* merupakan suatu teknik pengelompokan data dimana keberadaan tiap titik data dalam suatu kelompok ditentukan oleh derajat keanggotaan. Data yang digunakan berasal dari publikasi Statistik Pendidikan 2022 oleh Badan Pusat Statistik. Adapun objek yang digunakan adalah 34 provinsi di Indonesia dengan 24 variabel yang merepresentasikan indikator kualitas pendidikan. Pada penelitian ini, dilakukan pengelompokan dengan skenario sebanyak 2 sampai dengan 5 *cluster* dan diperoleh *cluster* yang optimal yaitu 2 *cluster*. Validasi *cluster* optimal ini menggunakan indeks validitas yaitu *Partition Coefficient* (PC), *Partition Entropy* (PE), dan *Modified Partition Coefficient* (MPC). Berdasarkan hasil penelitian, *cluster* pertama terdapat 12 provinsi yang dapat digolongkan sebagai kelompok provinsi dengan kualitas pendidikan rendah, sementara *cluster* kedua terdapat 22 provinsi yang dapat digolongkan sebagai kelompok provinsi dengan kualitas pendidikan tinggi.

Kata kunci: Pandemi, Kualitas Pendidikan, Pengelompokan, *Fuzzy C-Means*

I. PENDAHULUAN

Pendidikan memiliki peran yang penting untuk membangun sumber daya manusia yang unggul dan berdaya saing. Pentingnya pendidikan untuk pembangunan suatu wilayah telah disebutkan dalam Undang-Undang No. 20 Tahun 2003, bahwa pendidikan merupakan usaha dasar dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dirinya, masyarakat, bangsa dan negara.

Namun, adanya pandemi COVID-19 telah menyebabkan hilangnya pembelajaran (*learning loss*) di kalangan peserta didik dan tenaga pendidik [1]. Hilangnya pembelajaran ini tentu sangat berdampak pada tingkat kualitas pendidikan. Kualitas pendidikan atau terlaksananya pendidikan berkualitas merupakan salah satu target dalam *Sustainable Development Goals* (SDGs) poin ke-empat. Hilangnya pembelajaran itu membuat tercapainya target tersebut menjadi lebih sulit. Oleh karena itu, perlu dilakukan berbagai upaya pemulihan setelah kurang lebih dua tahun tidak ada proses pembelajaran di sekolah. Upaya yang dilakukan tentunya juga harus dengan cara yang tepat agar target dalam SDGs poin ke-empat dapat segera tercapai sesuai dengan target awal.

Salah satu hal yang perlu diperhatikan dalam upaya pemulihan tersebut adalah pemerataan kualitas pendidikan. Pemerataan kualitas pendidikan di Indonesia harus diperhatikan mengingat Indonesia ini merupakan negara kepulauan yang terdiri dari berbagai provinsi. Hal itu tentu membuat keadaan kualitas pendidikan di Indonesia juga berbeda-beda di setiap provinsinya. Oleh karena itu, diperlukan suatu

pengelompokan atau klusterisasi provinsi di Indonesia berdasarkan indikator pendidikan untuk mempermudah pemerintah dalam upaya pemerataan kualitas pendidikan. Hal tersebut sesuai dengan yang dikemukakan oleh [2] bahwa indikator pendidikan di suatu daerah dapat menjadi salah satu indikator parameter keberhasilan pendidikan. Pengelompokan berdasarkan indikator pendidikan dapat menjadi suatu cara untuk memetakan keadaan kualitas pendidikan berdasarkan setiap kelompoknya. Hal ini dapat mendukung pemerintah Indonesia dalam membuat kebijakan dalam pemerataan kualitas pendidikan di Indonesia.

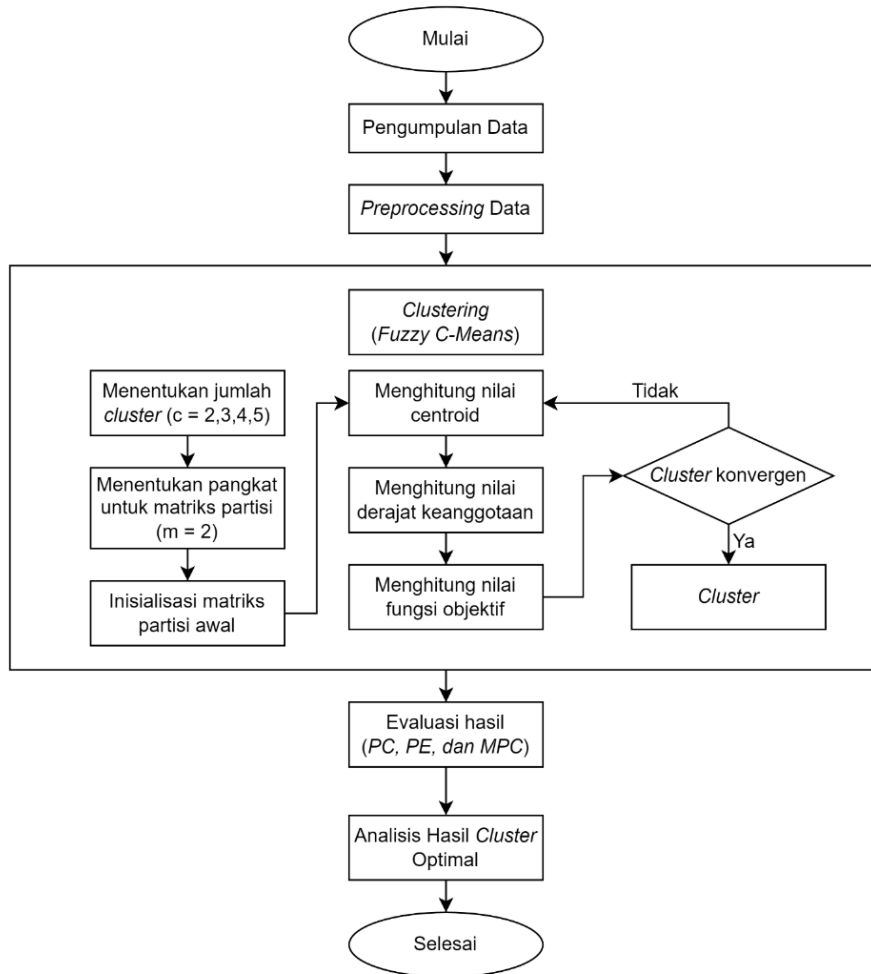
Algoritma pengelompokan yang dapat digunakan salah satunya adalah *Fuzzy C-Means*. *Fuzzy C-Means* suatu teknik pengelompokan data dimana keberadaan tiap titik data dalam suatu kelompok ditentukan oleh derajat keanggotaan [3]. Menurut [4], metode ini menggunakan model pengelompokan *fuzzy* dengan indeks kekaburan menggunakan *Euclidean Distance* sehingga data dapat menjadi anggota dari semua kelas atau *cluster* yang terbentuk dengan derajat keanggotaan yang berbeda antara 0 hingga 1. Metode *Fuzzy C-Means* dipilih karena menggunakan pemodelan *fuzzy* yang lebih fleksibel dengan melibatkan derajat keanggotaan dan memiliki keunggulan dalam menghasilkan pusat *cluster* yang lebih akurat dibandingkan dengan metode lainnya [5].

Penggunaan *Fuzzy C-Means* untuk pengelompokan suatu wilayah berdasarkan indikator pendidikan pernah dilakukan pada penelitian sebelumnya yang dilakukan oleh [2]. Metode *Fuzzy C-Means* pada penelitian tersebut digunakan untuk mengelompokkan kabupaten atau kota berdasarkan tingkat kemiripan atau kedekatan berdasarkan indikator pendidikan SMA/SMK/MA di Provinsi Jawa Timur. Pada penelitian tersebut menggunakan variabel indikator pendidikan Angka Partisipasi Murni, Angka Partisipasi Kasar, Angka Transisi, Angka Putus Sekolah, Angka Murid Mengulang, Angka Lulusan, Ratio Murid/Ruang Belajar, Ratio Kelas/Ruang Belajar, Ratio Murid/Guru, Ratio Murid/Sekolah. Penelitian lainnya dilakukan oleh [6] dengan mengimplementasikan *Fuzzy C-Means* dalam mengelompokkan provinsi di Indonesia juga dengan Indikator berupa Angka Partisipasi Sekolah (APS), Angka Partisipasi Murni (APM), dan Angka Partisipasi Kasar (APK).

Berdasarkan uraian tersebut, penelitian ini dilakukan untuk mengetahui jumlah kelompok yang optimal dan karakteristik setiap kelompok berdasarkan indikator kualitas pendidikan pada setiap provinsi di Indonesia dengan menggunakan metode *Fuzzy C-Means*. Dengan adanya penelitian ini, diharapkan dapat membantu pemerintah dalam menyusun kebijakan untuk meningkatkan kualitas pendidikan di Indonesia agar lebih efisien dan tepat sasaran terutama sebagai upaya dalam pemulihan pendidikan pasca pandemi COVID-19.

II. METODE PENELITIAN

Dalam penelitian ini instrumen yang digunakan yaitu data sekunder yang berasal dari publikasi Statistik Pendidikan 2022 oleh Badan Pusat Statistik [1]. Dalam hal ini, data yang digunakan adalah data yang sesuai dengan indikator kualitas pendidikan. Variabel yang digunakan yaitu: Persentase Sekolah Dasar (SD) menurut provinsi dan ketersediaan sanitasi ($X_1 - X_3$), Persentase Sekolah Menengah Pertama (SMP) menurut provinsi dan ketersediaan sanitasi ($X_4 - X_6$), Persentase Sekolah Menengah Atas (SMA) menurut provinsi dan ketersediaan sanitasi ($X_7 - X_9$), Persentase Sekolah Menengah Kejuruan (SMK) menurut provinsi dan ketersediaan sanitasi ($X_{10} - X_{12}$), Persentase guru yang memenuhi kualifikasi menurut provinsi dan jenjang pendidikan ($X_{13} - X_{16}$), Angka Partisipasi Murni (APM) PAUD anak usia 3-6 tahun menurut provinsi dan jenis kelamin (X_{17}), Angka Partisipasi Murni (APM) menurut provinsi dan jenjang pendidikan ($X_{18} - X_{21}$), dan Angka Melek Huruf (AMH) menurut provinsi dan kelompok umur ($X_{22} - X_{24}$).



GAMBAR 1. FLOWCHART PENELITIAN

Gambar 1 menunjukkan *flowchart* tahapan analisis dalam penelitian ini. Tahapan yang dilakukan, secara garis besar yaitu pengumpulan data, *preprocessing data*, *clustering*, evaluasi hasil, analisis hasil *cluster* optimal, dan kesimpulan.

A. *Preprocessing Data*

Preprocessing data merupakan salah satu langkah data *mining* yang berhubungan dengan persiapan dan transformasi data, juga sebagai langkah yang penting untuk meningkatkan efisiensi data. *Preprocessing data* bertujuan untuk mengurangi ukuran data, menemukan relasi antar data, normalisasi data, menghapus pencilan, dan mengekstrak fitur dari data. Langkah ini meliputi beberapa teknik seperti data *cleaning*, *integration*, *transformation*, dan *reduction* [7].

Preprocessing yang akan digunakan dalam penelitian ini adalah normalisasi dengan *Standard Scaler*. *Standard Scaler* adalah metode standarisasi fitur dengan menghapus rata-rata (*mean*) dan menskalakan unit varian (*standar deviasi*). *Preprocessing* dilakukan untuk mencegah adanya data yang memiliki nilai terlalu besar dibanding dengan nilai yang lain [8]. Rumus dari *Standard Scaler* ditunjukkan pada persamaan berikut.

$$Z = \frac{x_i - \bar{x}}{\sigma} \tag{1}$$

dimana \bar{x} adalah rata-rata nilai dan σ adalah standar deviasi.

B. *Fuzzy C-Means Clustering*

Logika *fuzzy* merupakan suatu logika yang memiliki nilai kekaburan (*fuzziness*) antara benar atau salah. Logika ini pertama kali diperkenalkan oleh Prof. Lotfi A. Zadeh pada tahun 1965. Suatu nilai bias dalam teori logika *fuzzy*, bernilai benar atau salah secara bersama. Akan tetapi, besarnya kebenaran dan kesalahannya tergantung pada derajat keanggotaan yang dimiliki. Adapun derajat keanggotaan dari logika *fuzzy*, berada dalam rentang 0 hingga 1. Pada teori himpunan *fuzzy*, peranan derajat keanggotaan sangat penting karena menjadi penentu keberadaan elemen dalam suatu himpunan. Logika *fuzzy* ini juga banyak digunakan untuk menerjemahkan suatu besaran yang diekspresikan menggunakan bahasa (*linguistic*). Contohnya, besaran kecepatan laju kendaraan yang diekspresikan menjadi pelan, agak cepat, cepat, dan sangat cepat [9].

Clustering adalah proses pengelompokan titik data menjadi dua atau lebih kelompok sehingga titik data dalam kelompok yang sama memiliki karakteristik yang lebih mirip satu sama lain daripada titik data pada kelompok yang berbeda. Metode *clustering* yang digunakan dalam penelitian ini menggunakan metode *Fuzzy C-Means*. *Fuzzy C-Means* menerapkan pengelompokan *fuzzy* dimana keberadaan tiap-tiap titik data dalam suatu kelompok ditentukan oleh derajat keanggotaan [10]. Metode ini meminimalkan fungsi objektif berhubungan dengan nilai anggota *fuzzy* u_{ij} dan *cluster centroid* V_i .

$$J_m = \sum_{i=1}^c \sum_{j=1}^n (\mu_{ij})^m d^2(X_j, V_i) \tag{2}$$

dimana

$$d^2(X_j, V_i) = (X_j - V_i)^T A (X_j - V_i). \tag{3}$$

Fungsi objektif ini memperhitungkan derajat keanggotaan setiap titik data pada setiap *cluster* dan menyesuaikan pusat *cluster* secara berkala untuk menghasilkan hasil pengelompokan akhir. Tujuannya adalah untuk menemukan solusi optimal dimana jumlah kuadrat jarak antara setiap titik data dan *centroid cluster* terkait diperkecil seoptimal mungkin. *Fuzzy C-Means* dalam melakukan *clustering* mempunyai langkah-langkah sebagai berikut [11]:

- 1) Menentukan nilai-nilai parameter
- 2) Inialisasi keanggotaan u_{ij} dari X_j pada *cluster* i sehingga

$$\sum_{i=1}^n \mu_{ij} = 1 \tag{4}$$

- 3) Menghitung *fuzzy centroid* V_i untuk $i = 1, 2, \dots, c$ menggunakan

$$V_i = \frac{\sum_{j=1}^n (\mu_{ij})^m X_j}{\sum_{j=1}^n (\mu_{ij})^m} \tag{5}$$

- 4) Perbaharui keanggotaan *fuzzy* μ_{ij} menggunakan

$$\mu_{ij} = \frac{\left(\frac{1}{d^2(X_j, V_i)}\right)^{\frac{1}{m-1}}}{\sum_{j=1}^n \left(\frac{1}{d^2(X_j, V_i)}\right)^{\frac{1}{m-1}}} \tag{6}$$

- 5) Ulangi langkah 3) dan 4) hingga nilai J_m tidak berkurang lagi.

C. *Evaluasi Hasil*

Cluster yang dihasilkan dari proses *clustering* ini perlu dilakukan validasi agar diperoleh hasil *cluster* yang optimal. Metode *clustering* yang menggunakan konsep *fuzzy* cukup berbeda dengan yang lain karena setiap data dapat menjadi anggota pada semua *cluster* dengan nilai derajat keanggotaannya. Berdasarkan hal itu, berarti setiap data tidak mutlak menjadi anggota satu *cluster* tertentu, melainkan juga mempunyai nilai untuk menjadi anggota *cluster* yang lain. Oleh karena itu, diperlukan suatu indeks validitas yang berkaitan dengan *Fuzzy C-Means* ini. Berikut beberapa indeks validitas yang digunakan yaitu *Partition Coefficient* (PC), *Partition Entropy* (PE), dan *Modified Partition Coefficient* (MPC) [12].

Partition Coefficient didefinisikan sebagai berikut.

$$PC(c) = \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^n (\mu_{ij})^2 \tag{7}$$

dimana $\frac{1}{c} \leq PC(c) \leq 1$. Nilai PC semakin dekat dengan 1 menunjukkan *cluster* yang lebih baik.

Sementara, *Partition Entropy* didefinisikan sebagai berikut.

$$PE(c) = -\frac{1}{n} \sum_{i=1}^c \sum_{j=1}^n \mu_{ij} \log_2 \mu_{ij} \tag{8}$$

dimana $0 \leq PE(c) \leq \log_2 c$. Nilai PE semakin dekat dengan 0 menunjukkan *cluster* yang lebih baik. Baik PC maupun PE keduanya memiliki kecenderungan perubahan yang monoton dengan c . Modifikasi dari indeks PC dapat mengurangi kecenderungan monoton dan didefinisikan sebagai berikut.

$$MPC(c) = 1 - \frac{c}{c-1} (1 - PC(c)) \tag{9}$$

dimana $0 \leq MPC(c) \leq 1$. Nilai MPC semakin besar menunjukkan *cluster* yang lebih baik.

III. HASIL DAN PEMBAHASAN

Jumlah data yang digunakan pada penelitian ini adalah 34 yang merupakan provinsi dengan 24 variabel yang merupakan indikator kualitas pendidikan. Data dapat dilihat pada Tabel 1. Data yang telah dikumpulkan tersebut perlu melalui tahap *preprocessing*.

TABEL 1. DATA

No	Provinsi	X ₁	X ₂	X ₃	X ₄	X ₅	...	X ₂₄
1.	Aceh	53	22	25	65	15	...	98,25
2.	Sumatera Utara	53	17	30	71	10	...	99,11
3.	Sumatera Barat	61	18	21	79	10	...	99,29
4.	Riau	54	21	24	68	12	...	99,18
5.	Jambi	65	18	17	79	9	...	98,10
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
34.	Papua	29	11	60	52	12	...	81,19

A. Preprocessing Data

Preprocessing data yang dilakukan dalam penelitian ini adalah normalisasi data. Hal ini dilakukan agar rentang nilai antar data tidak terlalu jauh, tidak ada yang terlalu besar maupun terlalu kecil. Normalisasi dilakukan menggunakan *standard scaler*. Hasil akhir dari normalisasi data yang dilakukan untuk setiap data ditunjukkan oleh Tabel 2.

TABEL 2. HASIL NORMALISASI DATA

No	Provinsi	X ₁	X ₂	X ₃	X ₄	X ₅	...	X ₂₄
1.	Aceh	-0,22973	0,44150	0,07463	-0,55586	0,88466	...	0,43173
2.	Sumatera Utara	-0,22973	-0,53324	0,48389	0,01112	-0,52087	...	0,66924
3.	Sumatera Barat	0,33321	-0,33829	-0,25278	0,76709	-0,52087	...	0,71895
4.	Riau	-0,15936	0,24655	-0,00722	-0,27237	0,04134	...	0,68857
5.	Jambi	0,61468	-0,33829	-0,58019	0,76709	-0,80198	...	0,39030
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
34.	Papua	-1,91853	-1,70294	2,93947	-1,78432	0,04134	...	-4,27984

B. Fuzzy C-Means Clustering

Setelah data melalui tahap *preprocessing*, data kemudian dikelompokkan dengan *Fuzzy C-Means* untuk mendapatkan hasil *cluster* dan pusat *cluster*. Langkah-langkah yang dilakukan adalah sebagai berikut.

1) Menentukan nilai-nilai parameter

Pada penelitian ini, nilai-nilai parameter yang digunakan yaitu jumlah *cluster* (c) = 2, 3, 4, dan 5; indeks *fuzzy* (m) = 2; nilai *error* yang diharapkan (ϵ) = 0.00005; dan maksimum iterasi (maxiter) = 100.

2) Inisialisasi keanggotaan u_{ij} dari X_j pada *cluster* i

Inisialisasi ini ditentukan secara acak dengan syarat jumlah nilai elemen dari kolom dalam setiap baris harus sama dengan 1. Inisialisasi keanggotaan u_{ij} dari X_j pada *cluster* i dalam 2 *cluster* ditunjukkan pada Tabel 3.

TABEL 3. HASIL INISIALISASI NILAI KEANGGOTAAN

No	Provinsi	Nilai Keanggotaan 1	Nilai Keanggotaan 2	No	Provinsi	Nilai Keanggotaan 1	Nilai Keanggotaan 2
1.	Aceh	0,96690	0,03310	18.	Nusa Tenggara Barat	0,65498	0,34502
2.	Sumatera Utara	0,53660	0,46340	19.	Nusa Tenggara Timur	0,44050	0,55950
3.	Sumatera Barat	0,49616	0,50384	20.	Kalimantan Barat	0,89502	0,10498
4.	Riau	0,46899	0,53101	21.	Kalimantan Tengah	0,82410	0,17590
5.	Jambi	0,30982	0,69018	22.	Kalimantan Selatan	0,83205	0,16795
6.	Sumatera Selatan	0,48647	0,51353	23.	Kalimantan Timur	0,41404	0,58596
7.	Bengkulu	0,54898	0,45102	24.	Kalimantan Utara	0,75499	0,24501
8.	Lampung	0,67111	0,32889	25.	Sulawesi Utara	0,20232	0,79768
9.	Kep. Bangka Belitung	0,58007	0,41993	26.	Sulawesi Tengah	0,72361	0,27639
10.	Kep. Riau	0,86426	0,13574	27.	Sulawesi Selatan	0,47417	0,52583
11.	DKI Jakarta	0,54284	0,45716	28.	Sulawesi Tenggara	0,89538	0,10462
12.	Jawa Barat	0,44092	0,55908	29.	Gorontalo	0,44293	0,55707
13.	Jawa Tengah	0,72973	0,27027	30.	Sulawesi Barat	0,75005	0,24995
14.	DI Yogyakarta	0,87774	0,12226	31.	Maluku	0,57370	0,42630
15.	Jawa Timur	0,18381	0,81619	32.	Maluku Utara	0,67739	0,32261
16.	Banten	0,19326	0,80674	33.	Papua Barat	0,35716	0,64284
17.	Bali	0,03424	0,96576	34.	Papua	0,85410	0,14590

Setelah diperoleh inisialisasi keanggotaan u_{ij} dari X_j pada *cluster* i , langkah yang berikutnya adalah sebagai berikut.

- 3) Menghitung *fuzzy centroid* V_i untuk $i = 1, 2, \dots, c$
- 4) Perbaharui keanggotaan *fuzzy* μ_{ij}
- 5) Ulangi langkah 3) dan 4) hingga nilai J_m tidak berkurang lagi.

Langkah tersebut menyesuaikan *centroid* (titik pusat) *cluster* berdasarkan perhitungan jarak antara data dan *centroid* dan memperbaharui keanggotaan *fuzzy* u_{ij} berdasarkan *centroid* baru. Langkah tersebut berulang hingga nilai *fuzzy membership* dan *centroid* tidak lagi berubah secara signifikan atau apabila iterasi sudah maksimal. *Centroid* akhir (final) dapat dilihat pada Tabel 4.

TABEL 4. CENTROID AKHIR

Cluster	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	-0,71426	0,29941	0,70608	-0,74864	0,55931	0,73726	-0,70904	0,49328
2	0,45785	-0,20361	-0,44786	0,47763	-0,38548	-0,45850	0,45359	-0,32326
Cluster	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}
1	0,68979	-0,79596	0,43952	0,83458	-0,58101	-0,32037	-0,14059	-0,22837
2	-0,43392	0,50269	-0,31116	-0,51385	0,36134	0,21808	0,09488	0,13352
Cluster	X_{17}	X_{18}	X_{19}	X_{20}	X_{21}	X_{22}	X_{23}	X_{24}
1	-0,00303	-0,34223	-0,41051	-0,14495	0,06381	-0,16022	-0,26648	-0,26406
2	-0,01845	0,25142	0,27037	0,12111	-0,03219	0,13321	0,20597	0,20564

Perubahan dikatakan tidak lagi berubah secara signifikan apabila tidak menyebabkan perubahan yang besar pada hasil akhir *clustering*. Hal ini diukur dengan membandingkan nilai *objective function* (fungsi objektif) sebelum dan setelah perubahan. Apabila perubahan fungsi objektif kurang dari *threshold* (batas) yang sudah ditentukan pada langkah 1, maka proses iterasi selesai dan hasil *clustering* dianggap konvergen atau stabil. Iterasi yang dilakukan dalam 2 *cluster* berhenti pada iterasi ke-25. Riwayat fungsi objektif dari iterasi pertama hingga iterasi terakhir dapat dilihat pada Tabel 5.

TABEL 5. RIWAYAT FUNGSI OBJEKTIF

Iterasi	Fungsi Objektif	Iterasi	Fungsi Objektif	Iterasi	Fungsi Objektif	Iterasi	Fungsi Objektif	Iterasi	Fungsi Objektif
1	521,5894646	6	397,5985776	11	394,9079034	16	394,8783097	21	394,8779758
2	406,5581631	7	396,0932351	12	394,8900620	17	394,8781105	22	394,8779734
3	404,9173463	8	395,3675204	13	394,8828933	18	394,8780287	23	394,8779725
4	402,6718057	9	395,0690359	14	394,8799836	19	394,8779952	24	394,8779721
5	399,9777072	10	394,9530112	15	394,8787959	20	394,8779814	25	394,8779719

Kemudian setelah iterasi berhenti, didapatkan nilai keanggotaan terakhir (final). Nilai ini menunjukkan seberapa kuat keanggotaan suatu titik data dengan setiap *cluster*. Nilai keanggotaan ini berupa angka antara 0 dan 1 dimana nilai mendekati 1 menunjukkan bahwa titik data termasuk dalam *cluster* tersebut sedangkan nilai mendekati 0 menunjukkan bahwa titik data tidak termasuk dalam *cluster* tersebut. Nilai keanggotaan final ini dapat dilihat pada Tabel 6 beserta dengan hasil *cluster*.

TABEL 6. HASIL NILAI KEANGGOTAAN PADA ITERASI TERAKHIR

No	Provinsi	Nilai Keanggotaan pada Iterasi Terakhir		Cluster	No	Provinsi	Nilai Keanggotaan pada Iterasi Terakhir		Cluster
		(μ_1)	(μ_2)				(μ_1)	(μ_2)	
1.	Aceh	0,45804	0,54196	2	18.	Nusa Tenggara Barat	0,68994	0,31006	1
2.	Sumatera Utara	0,38851	0,61149	2	19.	Nusa Tenggara Timur	0,75578	0,24422	1
3.	Sumatera Barat	0,15565	0,84435	2	20.	Kalimantan Barat	0,39687	0,60313	2
4.	Riau	0,24775	0,75225	2	21.	Kalimantan Tengah	0,56982	0,43018	1
5.	Jambi	0,09003	0,90997	2	22.	Kalimantan Selatan	0,26326	0,73674	2
6.	Sumatera Selatan	0,28549	0,71451	2	23.	Kalimantan Timur	0,11220	0,88780	2
7.	Bengkulu	0,17739	0,82261	2	24.	Kalimantan Utara	0,47465	0,52535	2
8.	Lampung	0,31468	0,68532	2	25.	Sulawesi Utara	0,59883	0,40117	1
9.	Kep. Bangka Belitung	0,21673	0,78327	2	26.	Sulawesi Tengah	0,63744	0,36256	1
10.	Kep. Riau	0,17568	0,82432	2	27.	Sulawesi Selatan	0,56024	0,43976	1
11.	DKI Jakarta	0,22813	0,77187	2	28.	Sulawesi Tenggara	0,77700	0,22300	1
12.	Jawa Barat	0,17776	0,82224	2	29.	Gorontalo	0,30506	0,69494	2
13.	Jawa Tengah	0,18385	0,81615	2	30.	Sulawesi Barat	0,75537	0,24463	1
14.	DI Yogyakarta	0,32893	0,67107	2	31.	Maluku	0,72335	0,27665	1
15.	Jawa Timur	0,39504	0,60496	2	32.	Maluku Utara	0,70834	0,29166	1
16.	Banten	0,20118	0,79882	2	33.	Papua Barat	0,79387	0,20613	1
17.	Bali	0,20946	0,79054	2	34.	Papua	0,58978	0,41022	1

C. Evaluasi Hasil

Selanjutnya, pemilihan jumlah *cluster* yang optimal dilakukan dengan melihat kriteria pada nilai *Partition Coefficient* (PC), *Modified Partition Coefficient* (MPC), dan *Partition Entropy* (PE). Adapun hasil validasi menggunakan nilai PC, MPC, dan PE untuk data penelitian dapat dilihat pada Tabel 7.

TABEL 7. HASIL VALIDASI

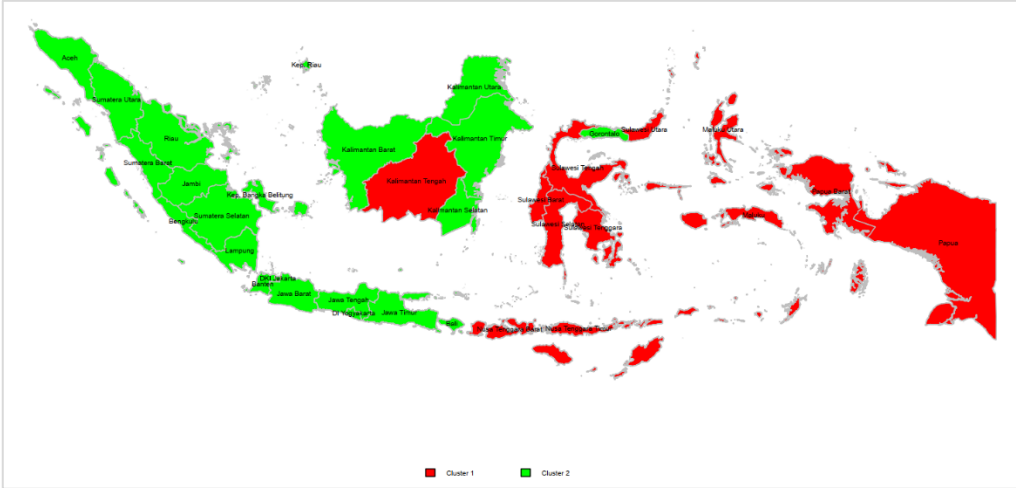
Jumlah Cluster	Partition Coefficient	Modified Partition Coefficient	Partition Entropy
2	0,61470	0,22941	0,56985
3	0,42329	0,13494	0,95207
4	0,34454	0,12605	1,18449
5	0,27978	0,09973	1,40791

Berdasarkan Tabel 7, dapat diketahui bahwa: (1) nilai *Partition Coefficient* (PC) pada jumlah *cluster* sebanyak dua lebih dekat dengan 1 daripada nilai PC pada jumlah *cluster* yang lainnya sehingga dengan indeks validitas PC, jumlah *cluster* terbaik adalah dua; (2) nilai *Modified Partition Coefficient* (MPC) pada

jumlah cluster sebanyak dua lebih tinggi daripada nilai MPC pada jumlah cluster yang lainnya sehingga dengan menggunakan nilai MPC jumlah cluster terbaik adalah dua; (3) nilai Partition Entropy (PE) pada jumlah cluster sebanyak dua lebih dekat dengan 0 daripada nilai PE pada jumlah cluster yang lainnya sehingga dengan menggunakan nilai PE jumlah cluster terbaik adalah dua. Terlihat dari hasil ketiga indeks validitas yang digunakan, semuanya menunjukkan bahwa jumlah cluster optimal adalah dua sehingga dapat disimpulkan bahwa dua merupakan jumlah cluster yang paling optimal.

D. Analisis Hasil Cluster Optimal

Setelah mendapatkan jumlah cluster optimal untuk pengelompokan provinsi di Indonesia berdasarkan kualitas pendidikan, selanjutnya hasil pengelompokan tersebut divisualisasikan pada Gambar 2.



GAMBAR 2. VISUALISASI CLUSTER DI SETIAP PROVINSI

Setelah divisualisasikan, untuk mengetahui karakteristik dari setiap cluster akan dihitung rata-rata data setiap variabel yang dimiliki oleh masing-masing cluster. Hasil rata-rata data setiap variabel tersebut disajikan pada Tabel 8.

TABEL 8. PROFILISASI CLUSTER

Variabel	Cluster 1	Cluster 2	Variabel	Cluster 1	Cluster 2	Variabel	Cluster 1	Cluster 2
Dasar (SD)	41,83	64,13	Tidak Ada (SMA)	19,91	11,31	APM (PAUD)	31,91	33,26
Terbatas (SD)	22,25	18,36	Dasar (SMK)	57	73,81	APM (SD)	95,44	98,25
Tidak Ada (SD)	36	17,59	Terbatas (SMK)	15,16	11,63	APM (SMP)	74,86	80,84
Dasar (SMP)	59,75	76,95	Tidak Ada (SMK)	28	14,36	APM (SM)	61,19	64,15
Terbatas (SMP)	15	10,13	Guru (SD)	89,16	95,46	APM (PT)	24,16	23,59
Tidak Ada (SMP)	25,41	12,86	Guru (SMP)	96,51	97,53	AMH (15-24)	99,12	99,90
Dasar (SMA)	66,16	78,72	Guru (SMA)	98,46	98,63	AMH (15-59)	96,68	99,18
Terbatas (SMA)	14	10	Guru (SMK)	95,98	96,62	AMH (>15)	95,04	97,58

Berdasarkan Tabel 8 di atas, dapat diketahui karakteristik cluster sebagai berikut.

1) Cluster 1

Cluster pertama terdiri dari 12 provinsi yaitu Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Tengah, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua. Cluster 1 ini memiliki karakteristik rata-rata persentase ketersediaan sanitasi yang masih kurang (ditandai dengan sanitasi dasar yang masih kurang, banyaknya sanitasi yang terbatas, dan banyak yang tidak memiliki sanitasi). Selain itu, rata-rata persentase guru yang memenuhi kualifikasi, rata-rata persentase APM, dan rata-rata persentase AMH lebih rendah daripada cluster 2. Berdasarkan hal tersebut dapat disimpulkan bahwa cluster 1 ini dapat digolongkan sebagai cluster dengan kualitas pendidikan rendah.

2) *Cluster 2*

Cluster kedua terdiri dari 22 provinsi yaitu Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Kalimantan Barat, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Gorontalo. *Cluster 2* ini memiliki karakteristik rata-rata persentase ketersediaan sanitasi yang cukup (ditandai dengan sanitasi dasar yang banyak, sedikitnya sanitasi yang terbatas, dan sedikit yang tidak memiliki sanitasi). Selain itu, rata-rata persentase guru yang memenuhi kualifikasi, rata-rata persentase APM, dan rata-rata persentase AMH lebih tinggi daripada *cluster 1*. Berdasarkan hal tersebut dapat disimpulkan bahwa *cluster 2* ini dapat digolongkan sebagai *cluster* dengan kualitas pendidikan tinggi.

Jumlah *cluster* optimal sebanyak dua yang dihasilkan hanya membagi data ke dalam dua kelompok saja sehingga tidak memungkinkan pemisahan data yang lebih detail. Ada beberapa faktor yang dapat menyebabkan hal ini terjadi dalam suatu analisis *cluster*, salah satunya berkaitan dengan pemilihan indikator. Indikator yang memiliki variabilitas atau variasi data yang rendah dapat mengakibatkan perbedaan antar *cluster* sulit dilihat.

IV. SIMPULAN DAN SARAN

Berdasarkan hasil dan pembahasan pada penelitian ini, dapat disimpulkan bahwa terdapat 2 *cluster* optimal untuk mengelompokkan provinsi di Indonesia berdasarkan indikator kualitas pendidikan. Hal ini dikarenakan jumlah 2 *cluster* memiliki nilai PC, MPC terbesar dan PE terkecil dibandingkan dengan jumlah *cluster* lainnya. *Cluster* pertama terdapat 12 provinsi yang dapat digolongkan sebagai kelompok provinsi dengan kualitas pendidikan rendah, sementara *cluster* kedua terdapat 22 provinsi yang dapat digolongkan sebagai kelompok provinsi dengan kualitas pendidikan tinggi.

Mempertimbangkan hasil tersebut, peneliti berharap agar pemerintah dapat memaksimalkan sistem pendidikan di Indonesia. Pemerintah sebaiknya segera mengambil tindakan terhadap *cluster* pertama dengan mengupayakan pemerataan program wajib belajar, meningkatkan infrastruktur pendidikan, dan meningkatkan kualitas tenaga pendidik. Akan tetapi, dengan tidak mengabaikan *cluster* kedua agar kualitas pendidikan dapat merata dan tidak terjadi lagi ketertinggalan antar provinsi. Untuk penelitian lebih lanjut, disarankan dapat menambah atau menggunakan indikator kualitas pendidikan yang lainnya yang belum digunakan dalam penelitian ini. Selain itu, dapat menggunakan metode yang lain pula selain *Fuzzy C-Means* untuk mengelompokkan provinsi di Indonesia berdasarkan kualitas pendidikannya.

DAFTAR PUSTAKA

- [1] B. P. Statistik, "Statistik Pendidikan 2022," *Badan Pus. Stat.*, 2022.
- [2] H. S. Karti, "Pengelompokan Kabupaten / Kota di Provinsi SMA / SMK / MA dengan Metode C-Means dan Fuzzy C-Means," vol. 2, no. 2, 2013.
- [3] A. K. Muchsin and M. Sudarma, "Penerapan Fuzzy C-Means Untuk Penentuan Besar Uang Kuliah Tunggal Mahasiswa Baru," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 6, no. 3, p. 175, 2015, doi: 10.24843/lkjiti.2015.v06.i03.p04.
- [4] E. T. Luthfi, "Fuzzy C-Means Untuk Clustering Data (Studi Kasus : Data Performance Mengajar Dosen)," vol. 2007, no. November, pp. 1–7, 2007.
- [5] N. Nidyashofa and D. Istiawan, "Penerapan Algoritma Fuzzy C-Means untuk Pengelompokan Kabupaten / Kota di Jawa Tengah Berdasarkan Status Kesejahteraan Tahun 2015," *6th Univ. Res. Colloq.*, no. September, pp. 23–30, 2017.
- [6] N. Yuliani, M. Fadli, and W. Nengsih, "Pengelompokan Angka Partisipasi Pendidikan Wajib Belajar Se-Indonesia Menggunakan Metode Fuzzy C-Means," *J. Aksara Komput. Terap.*, vol. 3, no. 2, 2014.
- [7] S. A. Alasadi and W. S. Bhaya, "Review of data preprocessing techniques in data mining," *J. Eng. Appl. Sci.*, vol. 12, no. 16, pp. 4102–4107, 2017, doi: 10.3923/jeasci.2017.4102.4107.
- [8] A. Ambarwari, Q. Jafar Adrian, and Y. Herdiyeni, "Analysis of the Effect of Data Scaling on the Performance of the Machine Learning Algorithm for Plant Identification," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 1, pp. 117–122, 2020, doi: 10.29207/resti.v4i1.1517.
- [9] D. L. Rahakbauw, V. Y. I. Ilwaru, and M. H. Hahury, "Implementasi Fuzzy C-Means Clustering Dalam Implementation Of Fuzzy C-Means Clustering In," *J. Ilmu Mat. dan Terap.*, vol. 11, pp. 1–12, 2017.
- [10] C. L. Simbolon, N. Kusumastuti, and B. Irawan, "Clustering lulusan mahasiswa matematika fmpa utan pontianak menggunakan algoritma fuzzy c - means," *Bul. Ilm. Mat. Stat. Dan Ter.*, vol. 02, no. 1, pp. 21–26, 2013.
- [11] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media, 2013.
- [12] K. L. Wu and M. S. Yang, "A cluster validity index for fuzzy clustering," *Pattern Recognit. Lett.*, vol. 26, no. 9, pp. 1275–1291, 2005, doi: 10.1016/j.patrec.2004.11.022.