

Grouping of Provinces in Indonesia Based on Quality Education Indicators in 2025 Using the Self-Organizing Maps Method

Dinda Putri Adilla dan Tessy Octavia Mukhti*

Departemen Statistika, Universitas Negeri Padang, Kota Padang, Negara Indonesia

*Corresponding author: tessyoctaviam@fmipa.unp.ac.id

Submitted : 15 Februari 2026

Revised : 03 Maret 2026

Accepted : 16 Maret 2026

ABSTRACT

Access to quality education plays an essential role in improving people's welfare and supporting sustainable development. As a fundamental component of social progress, quality education is not limited to academic attainment but also involves the development of skills, values, and character needed for meaningful participation in society. This study seeks to identify patterns and disparities in education quality among provinces by grouping regions based on multiple educational indicators. The indicators analyzed include Average Years of Schooling, Literacy Rate, Access to Information and Communication Technology, Gross Enrollment Rate, Net Enrollment Rate, and Teacher Qualifications. The data were examined using descriptive statistics, data visualization, and normalization, followed by clustering through the Self-Organizing Map (SOM) method as an unsupervised learning approach in data mining. Two clusters were formed to represent provinces with relatively higher and lower levels of educational quality. Cluster validity was assessed using internal validation measures, namely the Connectivity Index, Silhouette Index, and Dunn Index. The findings reveal that most basic education and literacy indicators show relatively favorable conditions; however, disparities remain evident in average years of schooling, ICT access, and participation in secondary and higher education. The clustering results indicate that 35 provinces fall into the group with relatively higher education quality, while 3 provinces are classified in the lower category. These results suggest that although the overall condition of education is relatively good, regional inequality in educational outcomes persists and requires targeted policy interventions to promote more balanced and inclusive development.

Kata kunci: Clustering, Kohonen, Quality education, Self Organizing-Maps (SOM)



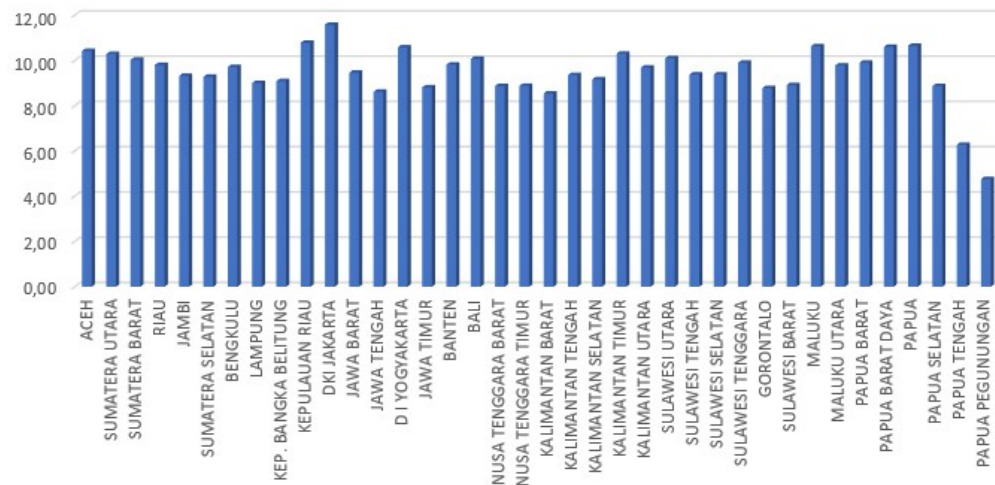
This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Pendidikan adalah upaya yang dilakukan dengan sengaja dan dirancang secara sistematis dengan maksud untuk menciptakan suasana dan alur pembelajaran yang kondusif agar tercipta kondisi yang mendukung peserta didik dalam mengembangkan potensi dirinya termasuk memperdalam nilai spiritual dan keagamaan, membangun pengendalian diri, meningkatkan kecerdasan, membentuk ahlak yang mulia serta mengasah keterampilan yang bermanfaat bagi diri sendiri dan masyarakat (Artawan dkk, 2023). Pendidikan sangat berperan penting untuk mencerdaskan kehidupan masyarakat dan memperluas pengetahuan bagi keberlangsungan hidup manusia. Pendidikan mendapatkan perhatian lebih oleh pemerintah karena salah satu indikator kemajuan bangsa adalah kualitas pendidikan yang baik (Patandung & Panggua, 2022). Pendidikan yang berkualitas diartikan sebagai pendidikan yang tidak hanya menitikberatkan pada pencapaian akademik, tetapi juga pada pengembangan keterampilan, penanaman nilai, serta pembentukan karakter yang diperlukan untuk berpartisipasi aktif dalam masyarakat (Situmeang dkk, 2021). Pendidikan berkualitas masuk kedalam target agenda pembangunan global yang disebut SDGs (*Sustainable Development Goal's*).

SDGs merupakan program pembangunan global yang terdiri dari 17 tujuan utama yang ditargetkan tercapai pada tahun 2030 dan ditetapkan oleh Perserikatan Bangsa-Bangsa (PBB) pada tahun 2015 sebagai kerangka kerja dalam mewujudkan pembangunan berkelanjutan yang meliputi tiga aspek utama, yaitu ekonomi, sosial, dan lingkungan (ECLAC, 2018). Pendidikan berkualitas merupakan tujuan SDGs poin ke-4. Meperoleh pendidikan berkualitas merupakan fondasi yang kuat bagi peningkatan taraf kehidupan masyarakat dan pembangunan berkelanjutan. Dengan demikian, poin ke-4 SDGs tidak hanya menyoroti akses terhadap pendidikan, tetapi juga mutu, relevansi, dan pemerataan pendidikan di seluruh lapisan masyarakat (Situmeang dkk, 2021).

Kutipan dari Tri Prasetyo dkk. (2023) ketimpangan kualitas pendidikan antar provinsi di Indonesia dapat diidentifikasi melalui berbagai data indikator pendidikan, seperti rata-rata lama sekolah, angka melek huruf, harapan lama sekolah, dan angka partisipasi murni. Data tersebut menunjukkan bahwa terdapat perbedaan yang cukup signifikan antar provinsi dalam hal capaian. Indikator Rata-rata Lama Sekolah (RLS) dipilih karena perpaduan yang dapat secara langsung mencerminkan tingkat pencapaian pendidikan masyarakat. Melalui perbandingan RLS antar provinsi, dapat diidentifikasi wilayah yang masih tertinggal dalam hal pemerataan pendidikan, sehingga indikator ini relevan untuk menunjukkan ketimpangan kualitas pendidikan di Indonesia. Ketimpangan tersebut dapat dilihat pada grafik RLS setiap provinsi di Indonesia pada tahun 2025.



Gambar 1. Rata-rata Lama Sekolah setiap provinsi di Indonesia tahun 2025

Pada Gambar 1 terlihat dengan jelas bahwa terdapat kesenjangan antara Provinsi DKI Jakarta dan Provinsi Papua Pegunungan. RLS penduduk di Provinsi DKI Jakarta mencapai 11,59 tahun sedangkan di Provinsi Papua Pegunungan hanya 4,3 tahun yang menunjukkan bahwa rata-rata penduduk Papua Pegunungan belum menyelesaikan pendidikan hingga tingkat SD (Sekolah Dasar).

Dalam upaya memahami ketimpangan serta pola pemerataan pendidikan antar wilayah secara lebih komprehensif, diperlukan metode analisis yang mampu mengelompokkan data berdasarkan kesamaan karakteristik secara tujuan. Salah satu metode yang sesuai untuk tujuan tersebut adalah *Self-Organizing Maps* (SOM), karena memiliki keunggulan dalam mengolah data *multivariat* serta mampu mengenali pola *non-linear* pada data (Imani dkk, 2023). Selain itu, SOM dapat menghasilkan visualisasi peta dua dimensi yang informatif dan mudah dipahami, sehingga efektif digunakan untuk memetakan kondisi pendidikan antar wilayah serta menunjukkan kedekatan karakteristik antar kelompok yang terbentuk. Dengan kemampuan tersebut, SOM menjadi metode yang tepat untuk mendukung analisis pemerataan pendidikan dan membantu mengidentifikasi kelompok wilayah dengan karakteristik pendidikan yang serupa secara lebih jelas (Imani dkk, 2023).

Penelitian ini bertujuan untuk melakukan pengelompokan provinsi di Indonesia berdasarkan indikator pendidikan berkualitas yang sesuai dengan tujuan *Sustainable Development Goals* (SDGs) ke-4 menggunakan metode *Self-Organizing Maps* (SOM). Hasil penelitian ini dapat dijadikan sebagai bahan pertimbangan dalam menyusun kebijakan pembangunan pendidikan di Indonesia, khususnya untuk menentukan prioritas wilayah yang perlu mendapatkan perhatian lebih berdasarkan hasil pengelompokan provinsi

II. METODE PENELITIAN

Penelitian ini merupakan penelitian kuantitatif deskriptif yang bertujuan untuk menggambarkan karakteristik setiap kelompok provinsi yang terbentuk sehingga dapat memberikan informasi mengenai kondisi ketimpangan kualitas pendidikan di Indonesia. Data yang digunakan berupa data sekunder yang diperoleh dari publikasi Badan Pusat Statistik (BPS) berjudul “Statistik Pendidikan 2025” Objek penelitian mencakup seluruh provinsi di Indonesia dengan variabel yang dianalisis berdasarkan indikator SDG’s poin ke-4 meliputi Rata-rata Lama Sekolah (X_1), Angka Melek Huruf (X_2), Akses Teknologi Informasi dan Komunikasi (X_3), Angka Partisipasi Murni SD (X_4), Angka Partisipasi Murni SMP

(X_5), Angka Partisipasi Murni SMA/SMK (X_6), Angka Partisipasi Murni PT (X_7), Angka Partisipasi Kasar Usia Dini (X_8), Angka Partisipasi Kasar SD (X_9), Angka Partisipasi Kasar SMP (X_{10}), Angka Partisipasi Kasar SMA/SMK (X_{11}), Angka Partisipasi Kasar PT (X_{12}) dan Kualifikasi Guru (X_{13}). Fokus utama penelitian adalah mengelompokkan provinsi-provinsi di Indonesia berdasarkan kualitas pendidikannya menggunakan metode *Self-Organizing Maps* (SOM).

Berikut merupakan tahapan-tahapan yang dilakukan dalam analisis *Self-Organizing Maps* (SOM) menggunakan software R:

1. Standarisasi Data

Standarisasi data merupakan proses untuk menyamakan jarak nilai antar variabel sehingga tidak terdapat perbedaan skala yang terlalu besar maupun terlalu kecil. Tujuan dari standarisasi adalah mengurangi perbedaan skala antar atribut dalam suatu dataset, sehingga setiap atribut dapat dibandingkan secara adil dan tidak mendominasi proses analisis statistik. Untuk melakukan standarisasi data dapat menggunakan *z-score* (Han, Kamber, & Pei. 2012) pada persamaan berikut:

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

Keterangan:

Z : Hasil standarisasi Z-score

x : Nilai data asli

\bar{x} : Nilai rata-rata seluruh data

σ : Simpangan baku

Dengan demikian, proses standarisasi menghasilkan matriks data terstandarisasi berukuran $N \times P$, di mana N adalah jumlah provinsi dan P adalah jumlah variabel yang digunakan dalam penelitian.

2. Menentukan jumlah kluster menggunakan validasi kluster

Menurut Brock dkk (2008) proses *clustering* perlu dilakukan validasi untuk menilai kualitas dan interpretasi kluster yang dihasilkan. Dalam penelitian ini, validasi internal dilakukan menggunakan tiga pendekatan, yaitu *Dunn Index*, *Silhouette Index*, dan *Connectivity Index*. Meskipun tidak terdapat ketentuan baku yang mewajibkan penggunaan lebih dari satu indeks validasi internal, penggunaan beberapa indeks secara simultan dianjurkan untuk meningkatkan reliabilitas hasil pengelompokan. Dengan mempertimbangkan ketiga aspek tersebut secara bersamaan, pemilihan jumlah kluster menjadi lebih komprehensif dan tidak bergantung pada satu ukuran evaluasi saja

(a) *Dunn index*

Nilai Indeks Dunn yang semakin besar menunjukkan kualitas kluster yang semakin baik. Indeks Dunn dirumuskan dalam persamaan 2 sebagai berikut:

$$DI = \min_{i=1, \dots, k} \left\{ \min_{j=i+1, \dots, k} \left(\frac{diss(c_i, c_j)}{\max_{m=1, \dots, m} (diam(c_i))} \right) \right\} \tag{2}$$

Keterangan:

DI : Indeks dunn

$diss(c_i, c_j)$: Jarak antara kluster i dan j

$diam(c_i)$: Diameter pasda kluster i

(b) *Silhouette index*

Nilai *silhouette* menunjukkan seberapa tepat suatu objek ditempatkan dalam klusternya. Objek yang terkelompok dengan bagus memiliki nilai *silhouette* mendekati 1, sedangkan objek yang kurang sesuai dengan klusternya memiliki nilai mendekati -1. Nilai *silhouette* untuk objek ke- i dihitung menggunakan persamaan 3 sebagai berikut:

$$S_{(i)} = \frac{b_i - a_i}{\max(b_i, a_i)} \tag{3}$$

$S_{(i)}$: *Silhouette coefficient* pada objek ke- i

a_i : Jarak rata-rata antara objek ke- i dengan seluruh objek lain dalam kluster yang sama

b_i : Jarak rata-rata antara objek ke- i dengan objek-objek pada kluster terdekat.

(c) *Connectivity index*

Suatu hasil pengelompokan dikatakan baik apabila nilai *connectivity* yang diperoleh semakin kecil, yang menandakan bahwa objek-objek dalam kluster memiliki keterkaitan yang semakin kuat. Koefisien *connectivity* dapat dinyatakan dengan persamaan 4 berikut:

$$Conn(C) = \sum_{i=1}^N \sum_{j=1}^L X_{i,nni(j)} \quad (4)$$

Keterangan:

$Conn(C)$: *Connectivity index*

$nni(j)$: Tetangga terdekat dari objek-j ke objek-i

L : Parameter jumlah tetangga terdekat

$X_{i,nni(j)}$: Nilai pada objek ke-i bernilai 0 jika objek i dan j dalam 1 kluster dan nilai 1/j ketika sebaliknya

3. Menentukan dimensi topologi jaringan (x, y) serta menetapkan nilai *learning rate* (α)

Dimensi topologi jaringan SOM ditentukan berdasarkan jumlah neuron yang disesuaikan dengan banyaknya data. Menurut Vesanto (2000), jumlah total neuron (m) dapat diperkirakan menggunakan pendekatan heuristik:

$$M \approx 5\sqrt{N} \quad (5)$$

Dimana N adalah jumlah data. Dengan jumlah provinsi sebanyak 38, diperoleh jumlah neuron mendekati 30, sehingga dipilih topologi jaringan berukuran 6x5 yang terdiri dari 30 neuron. *Learning rate* (α) merupakan parameter yang mengatur besar kesegaran bobot dalam setiap iterasi. nilai awal *learning rate* adalah 0,5 yang kemudian diturunkan secara bertahap mengikuti fungsi:

$$a(t) = a_0 \left(1 - \frac{t}{T}\right) \quad (6)$$

Keterangan:

$a(t)$: *Learning rate* pada iterasi ke-t

a_0 : *Learning rate* awal

t : Iterasi sekarang

T : Jumlah iterasi total

4. Memberi nilai awal bobot pada setiap neuron output secara acak dalam rentang $0 < w < 1$, dengan ukuran matriks bobot sama dengan jumlah variabel (P) dikalikan jumlah neuron (K)

5. Menentukan jarak minimum antara neuron input terhadap neuron output.

Neuron yang terpilih sebagai pemenang merupakan neuron dengan nilai jarak terkecil terhadap data input, di mana jarak tersebut dihitung menggunakan metode *Euclidean Distance*. Rumus yang digunakan untuk menentukan besarnya jarak *Euclidean* dinyatakan sebagai berikut:

$$D_{ij} = \sqrt{\sum_{k=1}^p (X_{ik} - X_{jk})^2} \quad (7)$$

Keterangan:

D_{ij} : Jarak *Euclidean* antara objek ke-i dan ke-j

X_{ik} : Nilai pengamatan pada objek ke-i untuk variabel ke-k

X_{jk} : Nilai pengamatan antara objek ke-j untuk variabel ke-k

p : Jumlah variabel yang digunakan

Neuron output yang memiliki jarak paqling kecil terhadap neuron input disebut neuron pemenang dan bobot neuron tersebut kemudian diperbarui.

6. Melakukan pembaruan nilai bobot pada neuron pemenang beserta neuron-neuron yang berada di sekitarnya (neuron tetangga).

Melakukan pembaruan bobot neuron pemenang agar mendekati data input menggunakan persamaan 8 berikut:

$$W_{ij.baru} = W_{ij.lama} + a(X_{ij} - W_{ij.lama}) \quad (8)$$

Keterangan:

- $W_{ij.baru}$: Bobot baru dari neuron baris ke-i dan kolom ke-j
- $W_{ij.lama}$: Nilai bobot sebelumnya dari neuron baris ke-i dan kolom ke-j
- a : *Learning rate* yang digunakan dalam proses pembaruan bobot
- X_{ij} : Nilai data input pada baris ke-i dan kolom ke-j

Untuk memperbarui bobot dari neuron tetangga menggunakan persamaan sebagai berikut:

$$W_{ij.baru} = W_{ij.lama} + \theta \times a(X_{ij} - W_{ij.lama}) \tag{9}$$

- θ : Nilai fungsi ketetanggaan

Variabel θ merupakan nilai yang diperoleh dari koefisien fungsi ketetanggaan. Fungsi ketetanggaan yang digunakan mengikuti distribusi normal sebagaimana ditunjukkan pada Persamaan 9, yang menyatakan jarak kuadrat antara dua neuron dimana

$$\theta = e^{-\left(\frac{\|r_1 - r_2\|^2}{2\delta_t}\right)} \tag{10}$$

Keterangan:

- e : Bilangan yang digunakan dalam fungsi *Gaussian* (distribusi normal)
- r_1 : Posisi neuron pemenang (BMU)
- r_2 : Posisi neuron tetangga
- $\|r_1 - r_2\|^2$: Jarak kuadrat antara kedua neuron
- δ_t : Radius neighborhood pada iterasi ke-t

Sementara itu, parameter δ_t merupakan nilai radius *neighborhood* yang berfungsi sebagai hiperparameter dalam algoritma SOM dan nilainya akan menurun seiring bertambahnya jumlah iterasi.

Dimana:

$$\delta_t = \delta \left(1 - \frac{t}{T}\right) \tag{11}$$

Keterangan:

- δ : Nilai awal radius neighborhood
- t : Indeks iterasi
- T : Jumlah iterasi total

7. Mengulai proses pada langkah 5 dan 6 untuk seluruh data input yang tersedia.
8. Mengelompokkan neuron-neuron yang telah diperbarui kedalam beberapa kalster, sehingga neuron dengan karakteristik yang paling mirip dapat berada dalam kelompok yang sama, dengan menggunakan persamaan (7).
9. Melakukan profilisasi hasil klaster, kemudian menyusun interpretasi serta membuat kesimpulan dari hasil analisis.

III. HASIL DAN PEMBAHASAN

Pada tahap awal eksplorasi data digunakan statistik deskriptif untuk melihat sebaran data dalam penelitian yang disajikan pada Tabel 1

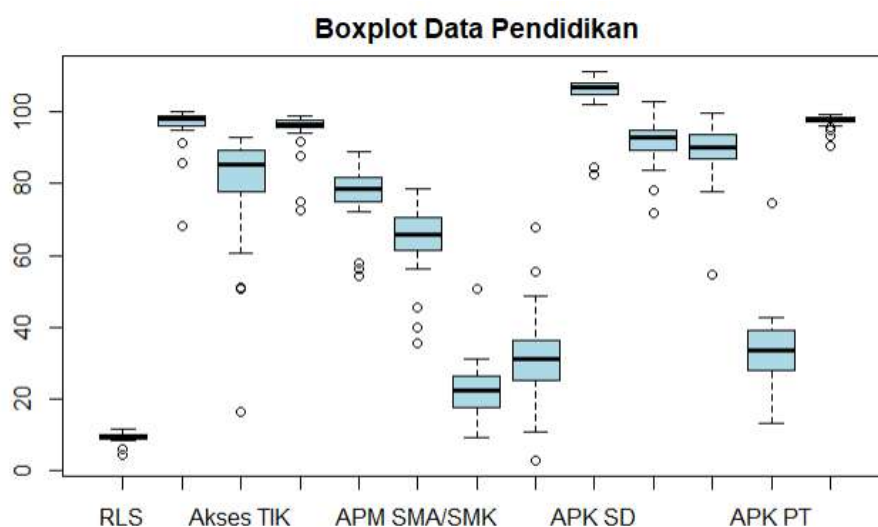
Tabel 1. Statistik Deskriptif

Variabel	Minimal	Median	Mean	Maksimal
RLS	4,76	9,57	9,45	11,58
AMH	68,32	98,17	96,58	99,8
Akses TIK	16,4	85,16	80	92,88
APM SD	72,55	96,39	95,09	98,77
APM SMP	54,33	78,47	77,34	88,67
APM SMA/SMK	35,59	65,61	64,42	78,42
APM PT	9,2	22,57	22,75	50,63
APK Usia Dini	3,1	31,36	31,63	67,63
APK SD	82,52	106,61	105,47	111,22

APK SMP	71,89	92,64	91,54	102,68
APKSMA/SMK	54,52	90,14	88,83	99,59
APK PT	13,34	33,55	33,56	74,7
Kualifikasi Guru	90,38	97,34	97,48	98,99

Hasil statistik deskriptif menunjukkan bahwa meskipun sebagian besar indikator pendidikan berada pada kondisi yang relatif baik, ketimpangan antarprovinsi masih terlihat cukup jelas. Perbedaan kemampuan terutama tampak pada rata-rata lama sekolah, akses teknologi informasi dan komunikasi, serta partisipasi pada jenjang pendidikan menengah. Selain itu, partisipasi pendidikan cenderung menurun seiring dengan meningkatnya jenjang pendidikan, yang mengindikasikan masih adanya hambatan dalam keinginan belajar ke tingkat yang lebih tinggi. Variasi pada pendidikan usia dini dan pendidikan tinggi juga menunjukkan bahwa pemerataan akses. Secara umum, kondisi ini menegaskan bahwa meskipun kualitas pendidikan nasional tergolong baik, disparitas antarwilayah masih menjadi isu penting yang perlu dianalisis lebih lanjut melalui pengelompokan klaster.

Untuk visualisasi data penelitian ditampilkan dalam bentuk boxplot pada Gambar 2



Gambar 2. Boxplot Data Pendidikan

Langkah awal sebelum masuk kedalam langkah analisis SOM perlu dilakukan standarisasi data karena beberapa variabel memiliki ukuran satuan yang berbeda, hal itu dapat dilihat dari Gambar 2. Hasil standarisasi data pendidikan disajikan dalam Tabel 2

Tabel 2. Hasil standarisasi data

Provinsi	X1	X2	X3	...	X13
Aceh	0,808	0,433	-0,51	...	0,588
Sumatera Utara	0,699	0,449	0,321	...	0,15
Sumatera Barat	0,481	0,462	0,674	...	0,844
Riau	0,287	0,448	0,41	...	0,118
⋮	⋮	⋮	⋮	⋮	⋮
Papua Pegunungan	-3,947	-5,238	-4,251	...	-4,143

Selanjutnya menentukan jumlah kluster, pada penelitian ini ingin dibentuk 2 kluster dimana pada data pendidikan di Indonesia, yang bertujuan untuk mengelompokkan wilayah berdasarkan kualitas pendidikan, yaitu kluster dengan kualitas cukup baik dan kluster dengan kualitas kurang baik.

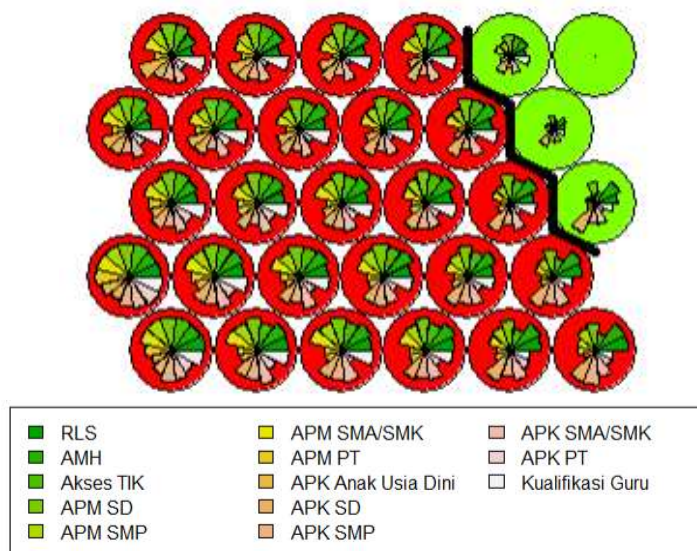
Setelah jumlah kluster ditentukan, langkah pertama dalam analisis SOM adalah uji validitas kluster untuk mengevaluasi jumlah kluster yang telah terbentuk. Uji validitas kluster yang digunakan *dunn index* dengan nilai kelayakan kluster yang terbentuk mendekati 1 menggunakan persamaan (2), nilai *silhouette index* yang dilihat adalah yang paling besar menggunakan persamaan (3) dan nilai *connectivity index* paling kecil menggunakan persamaan (4), dapat dilihat pada Tabel 3.

Tabel 3. Hasil Validasi Kluster

Index	Ukuran Kluster				
	2	3	4	5	6
Dunn	0,5212	0,5637	0,2463	0,2463	0,3244
Silhouette	0,6886	0,5631	0,3478	0,3043	0,2782
Connectivity	3,2623	7,4659	12,5238	14,7738	20,9135

Pada Tabel 3 terlihat bahwa nilai *dunn index* yang mendekati 1 terdapat pada kluster 3 yaitu 0,5637 dan kluster 2 yaitu 0,5212, lalu nilai *silhouette index* yang paling besar terdapat pada kluster 2 yaitu 0,6886 dan nilai *connectivity index* yang terkecil terdapat pada kluster 2 yaitu 3,2623. Sehingga dari hasil validasi kluster tersebut dapat disimpulkan bahwa kluster 2 adalah kluster terbaik yang dapat digunakan untuk mengelompokkan data pendidikan di Indonesia pada tahun 2025.

Selanjutnya adalah melakukan pengelompokan provinsi berdasarkan pendidikan di Indonesia menggunakan algoritma SOM. Ukuran topologi yang digunakan adalah 6x5 serta nilai *learning rate* sebesar 0,5, proses ini menghasilkan gambar diagram kipas (*fan*) yang terdapat pada Gambar 3.



Gambar 3. Diagram kipas (*fan*)

Diagram kipas (*fan diagram*) digunakan untuk menggambarkan ciri khas dari setiap kluster yang terbentuk, sehingga dapat memperlihatkan bagaimana masing-masing variabel tersebar dalam proses pemetaan. Semakin besar tampilan kipas pada diagram, semakin tinggi pula nilai variabel yang direpresentasikan. Pada diagram tersebut terdapat neuron berbentuk lingkaran yang memuat karakteristik setiap anggotanya. Neuron-neuron kemudian dikelompokkan dengan neuron lain yang memiliki tingkat kemiripan paling tinggi hingga terbentuk dua kelompok kluster. Masing-masing kluster dibedakan dengan warna yang berbeda, yaitu merah dan hijau, serta dipisahkan oleh vektor yang digambarkan dalam bentuk garis hitam pada diagram. Secara umum, kluster berwarna merah menunjukkan nilai variabel

yang relatif tinggi yang terlihat dari ukuran kipas yang lebih besar, sedangkan kluster berwarna hijau menunjukkan nilai variabel yang lebih rendah dengan ukuran kipas yang lebih kecil.

Penjelasan lebih rinci mengenai jumlah dan anggota yang termasuk dalam masing-masing kluster disajikan pada Tabel 4.

Tabel 4. Jumlah dan anggota kluster

Klaster	Jumlah Anggota	Anggota Kluster
1	35	Aceh, Bali, Banten, Bengkulu, DI Yogyakarta, DKI Jakarta, Gorontalo, Jambi, Jawa Barat, Jawa Tengah, Jawa Timur, Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Kepulauan Bangka Belitung, Kepulauan Riau, Lampung, Maluku, Maluku Utara, Nusa Tenggara Barat, Nusa Tenggara Timur, Papua, Papua Barat, Papua Barat Daya, Riau, Sulawesi Barat, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Utara, Sumatera Barat, Sumatera Selatan, dan Sumatera Utara.
2	3	Papua Pegunungan, Papua Selatan dan Papua Tengah

Berdasarkan Tabel 4, dapat diketahui bahwa kluster 1 mencakup 35 provinsi yang tergolong dalam kluster berwarna merah, sedangkan kluster 2 terdiri dari 3 provinsi yang termasuk dalam kluster berwarna hijau.

IV. KESIMPULAN

Penelitian ini berhasil mengelompokkan provinsi Indonesia berdasarkan pendidikan berkualitas tahun 2025 menggunakan metode *Self Organizing-Maps* (SOM) menjadi 2 kluster yang merupakan jumlah kluster paling optimal dalam penelitian ini. Hasil kluster yang didapatkan yaitu sebanyak 35 provinsi tergolong ke dalam kluster 1 yang menunjukkan kondisi pendidikan relatif lebih baik, ditinjau dari berbagai indikator yang digunakan dalam penelitian. Provinsi-provinsi pada kluster ini umumnya memiliki tingkat partisipasi pendidikan lebih tinggi dan akses pendidikan yang lebih baik, serta capaian indikator pendidikan lainnya yang lebih optimal. Sementara itu, sebanyak 3 provinsi tergolong ke dalam kluster 2, yang mencerminkan kualitas pendidikan yang relatif lebih rendah dibandingkan kluster 1. Provinsi pada kluster ini menunjukkan capaian indikator pendidikan yang masih tertinggal, sehingga memerlukan perhatian dan upaya peningkatan yang lebih intensif, baik dalam hal akses pendidikan, partisipasi sekolah, maupun pemerataan kualitas pendidikan.

Dengan demikian, hasil pengelompokan ini menunjukkan bahwa sebagian besar provinsi di Indonesia telah berada pada kategori kualitas pendidikan yang cukup baik, namun masih terdapat sejumlah wilayah yang memerlukan prioritas kebijakan dan intervensi khusus untuk meningkatkan kualitas pendidikan secara lebih merata.

Adapun untuk penelitian selanjutnya, disarankan untuk:

1. Menggunakan data panel beberapa tahun agar dapat menganalisis dinamika perubahan kluster dari waktu ke waktu.
2. Membandingkan metode SOM dengan metode klusterisasi lain, seperti *K-Means* atau *Hierarchical Clustering*, untuk konsistensi hasil pengelompokan.

DAFTAR PUSTAKA

- Artawan, P., Muhammadiyah, M., Hamsiah, A., Pongpalilu, F., Rachmandhani, S., Utari, T. I., Pratama, A., Mahmudah, K., Sumardi, M. S., & Wahyuningsih, N. S. (2023). *Pengantar Ilmu Pendidikan*. SONPEDIA.
- Badan Pusat Statistik. (2025). *Statistik Pendidikan* (Vol. 13). Badan Pusat Statistik.
- Brock, G., Pihur, V., Datta, Susmita, & Datta, Somnath. (2008). *clValid: An R Package for Cluster Validation*. In *JSS Journal of Statistical Software* (Vol. 25). <http://www.jstatsoft.org/>
- ECLAC. (2018). *The 2030 Agenda and the Sustainable Development Goals An opportunity for Latin America and the Caribbean*. www.issuu.com/publicacionescepal/stacks
- Han, Kamber & Pei. (2012) *Konsep dan Teknik Penambangan Data*. Edisi ke-3, Morgan Kaufmann Publishers, Waltham.

- Imani, N., Alfassa, A. I., & Yolanda, A. M. (2023). Analisis *Cluster* Terhadap indikator Data Sosial di Provinsi Nusa Tenggara Timur Menggunakan *Self Organizing-Maps* (SOM). *Jurnal Gaussian*, 11(3), 458–467. <https://doi.org/10.14710/j.gauss.11.3.458-467>
- Iyohu, L. R., Djakaria, I., & Nashar, L. O. (2023). Perbandingan Metode K-Means Clustering dengan Self-Organizing Maps (SOM) untuk Pengelompokan Provinsi di Indonesia Berdasarkan Data Potensi Desa. *Jurnal Statistika Dan Aplikasinya*, 7(2). www.bps.go.id
- Patandung, Y., & Panggua, S. (2022). Analisis Masalah-Masalah Pendidikan dan Tantangan Pendidikan Nasional. *Jurnal Sinestesia*, 12(2), 794–805. <https://sinestesia.pustaka.my.id/journal/article/view/277>
- Situmeang, D. E., Hawa, M. M., & Ismail, K. (2021). Pembangunan Berkelanjutan SDGs 2030 *goals 4 Ensure Inclusive and Equitable Quality Education and Promotr Lifelong Learning Opportunities for All* “Memastikan kualitas pendidikan yang inklusif dan adil dan mempromosikan kesempatan belajar seumur hidup untuk semua.” *ResearchGate*. <https://doi.org/10.13140/RG.2.2.11219.96809>
- Tri Prasetyo, H., Zein Azizah, S., Salmah, H., Anna Kusumawardani, N., Agistira Helena, R., & Setiyadi, D. (2023). Analisis Ketimpangan Gender pada Capaian Bidang Pendidikan di Jawa Tengah Periode 2019-2021. *ResearchGate*.
- Vesanto, J. and Alhoniemi, E. (2000) Clustering of the Self-Organizing Map. *IEEE Transactions on Neural Networks*, 11, 586-600. <https://doi.org/10.1109/72.846731>
- Yati Beti, I., & Juliansa, H. (2024). Penerapan Normalisasi Data Metode Decimal Scaling Dan Metode K-Means Dalam Mengelompokkan Kasus Demam Berdarah. *Media Online*, 4(6), 2928–2936. <https://doi.org/10.30865/klik.v4i6.1925>