

# Penerapan *Self Organizing Maps* dalam Pengelompokan Provinsi di Indonesia Berdasarkan Aspek Pendidikan

Muhammad Fariz Faqih, Ilham Faishal Mahdy\*

*Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.*

## ARTICLE INFO

### Article history :

Received : 20/08/2024  
Revised : 29/12/2024  
Published : 31/12/2024



Creative Commons Attribution-ShareAlike 4.0 International License.

Volume : 4  
No. : 2  
Halaman : 93 – 102  
Terbitan : **Desember 2024**

Terakreditasi Sinta [Peringkat 4](#)  
berdasarkan Ristekdikti  
No. 177/E/KPT/2024

## ABSTRAK

Indonesia berpartisipasi dalam Programme for International Student Assessment (PISA) pada tahun 2018, sebuah studi global yang mengevaluasi sistem pendidikan di 70 negara. Namun, Survei Sosial Ekonomi 2020, sebagaimana dilaporkan dalam Statistik Pendidikan 2020, menunjukkan bahwa Indonesia belum mencapai keberhasilan penuh dalam program wajib belajar di tingkat SMA, dengan Angka Partisipasi Kasar yang masih di bawah 95% dan 1,13% dari populasi gagal menyelesaikan pendidikan di tingkat tersebut. Artikel ini berfokus pada pengelompokan pendidikan di Indonesia berdasarkan provinsi-provinsi di Indonesia dan aspek-aspek pendidikan. Artikel ini bertujuan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan aspek pendidikan menggunakan metode Self Organizing Maps (SOM), sebuah metode dalam Artificial Neural Network yang digunakan untuk clustering. Data yang digunakan dalam artikel ini berasal dari 34 provinsi pada tahun 2023. Evaluasi performa model clustering dilakukan menggunakan Davies-Bouldin Index. Artikel ini mengidentifikasi bahwa jumlah kluster optimum adalah dua, dengan kluster pertama terdiri dari 32 provinsi dan kluster kedua terdiri dari 2 provinsi. Artikel ini memberikan pemahaman mendalam mengenai distribusi dan karakteristik pendidikan di berbagai provinsi di Indonesia, yang dapat membantu dalam perencanaan dan pengambilan kebijakan pendidikan di masa mendatang.

**Kata Kunci :** Pengelompokan; Indeks Davies-Bouldin; Pemetaan Pengorganisasian Diri

## ABSTRACT

Indonesia participated in the Programme for International Student Assessment (PISA) in 2018, a global study that evaluates education systems in 70 countries. However, the 2020 Socio-Economic Survey, as reported in the 2020 Education Statistics, indicated that Indonesia has not yet achieved full success in its compulsory education program at the high school level, with a Gross Enrollment Ratio still below 95% and 1.13% of the population failing to complete education at that level. This study focuses on clustering education in Indonesia based on provinces and educational aspects. The study aims to cluster provinces in Indonesia using the Self Organizing Maps (SOM) method, a technique within Artificial Neural Networks used for clustering. The data used in this study comes from 34 provinces in 2023. The performance of the clustering model is evaluated using the Davies-Bouldin Index. The study identifies that the optimal number of clusters is two, with the first cluster consisting of 32 provinces and the second cluster consisting of 2 provinces. This research provides a deep understanding of the distribution and characteristics of education across various provinces in Indonesia, which can assist in future educational planning and policy-making.

**Keywords :** Clustering; Davies-Bouldin Index; Self Organizing Maps.

Copyright© 2024 The Author(s)..

## A. Pendahuluan

Analisis kluster merupakan suatu teknik pengelompokan data yang bertujuan untuk mengelompokkan objek-objek yang mempunyai sifat serupa ke dalam kelompok (*cluster*) yang berbeda. *Clustering* merupakan pengelompokan objek dalam beberapa *cluster* (kelompok), yang mempunyai tingkat kemiripan sifat-sifat terhadap objek lain dibentuk menjadi satu *cluster*, sedangkan yang tidak mempunyai tingkat kesamaan karakteristik, dikelompokkan pada *cluster* yang lain [1]. *Clustering* sangat berguna untuk analisis dan eksplorasi data karena dapat mengungkap pola dan struktur yang sebelumnya tidak diketahui dalam data, hal ini membantu memahami data dengan lebih baik. Selain itu juga *clustering* memungkinkan data serupa secara otomatis dikelompokkan ke dalam beberapa *cluster* atau grup, hal ini membantu mengungkap pola dan struktur yang sebelumnya tidak diketahui dalam data [2].

Pada pengelompokan dalam analisis kluster secara umum terdiri dari dua metode, yakni metode Hierarki dan *Partitional Clustering*. Metode hierarki terdiri atas lima metode, yaitu *Single Linkage*, *Complete Linkage*, *Centroid Linkage*, *Average Linkage* dan Metode *Ward* sedangkan yang termasuk dalam metode *Partitional Clustering* (non-hierarki) dan yang paling banyak digunakan adalah *K-Means Clustering* [3]. Selain *K-Means Clustering* pada metode *partitional clustering* terdapat juga metode *Self Organizing Maps*, yaitu salah satu metode dalam *Artificial Neural Network* (ANN) yang dapat menyelesaikan permasalahan *clustering* atau pengelompokan seperti halnya metode *K-Means* yang digunakan [4].

*Artificial Neural Network* (ANN) merupakan salah satu cabang dari kecerdasan buatan (*Artificial Intelligence/AI*) yang meniru cara kerja otak manusia dalam memproses informasi dan menyelesaikan masalah [5]. ANN menggambarkan sebagai sistem pemrosesan paralel yang luas dengan kemampuan menyimpan dan menggunakan pengetahuan berdasarkan pengalaman [6]. ANN juga dapat didefinisikan sebagai sistem pengolahan informasi yang memiliki karakteristik serupa dengan jaringan syaraf biologis [7]. Konsep ANN didasarkan pada generalisasi model matematika dari sistem syaraf biologis, dengan beberapa asumsi dasar. (1) Pemrosesan informasi terjadi pada banyak elemen sederhana yang disebut neuron, (2) Sinyal ditransmisikan antar neuron melalui koneksi-koneksi, (3) Setiap koneksi antar neuron memiliki bobot yang dapat memperkuat atau memperlemah sinyal, (4) Setiap neuron menggunakan fungsi aktivasi untuk menentukan output berdasarkan jumlah input yang diterima, (5) Output dari fungsi aktivasi dibandingkan dengan nilai ambang tertentu. Adapun tiga komponen utama yang menentukan karakteristik ANN adalah, (1) Arsitektur jaringan: pola hubungan antar neuron, (2) Metode *training/learning/algorithm*: cara menentukan bobot koneksi, (3) Fungsi aktivasi: fungsi yang digunakan untuk menghasilkan output [8].

*Self Organizing Maps* diperkenalkan oleh Kohonen pada tahun 1982, dalam sebuah studi teoritis tentang pengorganisasian diri dari ruang output berdimensi rendah yang disebabkan oleh ruang input berdimensi tinggi. *Self Organizing Maps* merupakan bagian dari pembelajaran tanpa pengawasan dimana atau *unsupervised learning* dimana tidak ada target output yang telah ditentukan, *Self Organizing Maps* berupaya mengatur data masukan ke dalam peta dua dimensi berdasarkan kesamaan pola. Kelebihan dari *Self Organizing Maps* adalah memiliki fleksibilitas tinggi untuk paralelisasi yang membuat peningkatan *throughput* (kecepatan *rate*) transfer data efektif, yang diukur dalam BPS (bit per second) aplikasi, dan konsumsi daya yang rendah [4]. *Self Organizing Maps* digunakan untuk membentuk kelompok berdasarkan karakteristik data [9].

Pendidikan memegang peranan penting dalam mencerdaskan kehidupan masyarakat, memajukan ilmu pengetahuan dan teknologi, serta mengembangkan tenaga kerja yang kompeten dan mampu bersaing di era globalisasi [10]. Menurut Soraya [11] pendidikan dipengaruhi oleh berbagai aspek seperti politik, ekonomi, sosial budaya, perkembangan ilmu pengetahuan dan teknologi, globalisasi serta pendidikan juga harus tanggap terhadap perubahan zaman. Kualitas suatu negara dapat dilihat dari kualitas pendidikan yang diberikannya, salah satu tujuan pembangunan berkelanjutan (SDGs) adalah untuk “menjamin kualitas pendidikan yang inklusif dan adil serta mendukung kesempatan belajar seumur hidup bagi semua”.

Pada tahun 2018 Indonesia mengikuti tes PISA (*Programme for International Student Assessment*), PISA merupakan studi untuk mengevaluasi sistem Pendidikan yang diikuti oleh 70 negara di seluruh dunia. PISA dibentuk karena negara-negara mulai menyadari bahwa pentingnya evaluasi pendidikan [12]. Evaluasi berlangsung tiga tahun sekali, yang dinilai adalah siswa-siswa berusia 15 tahun dari sekolah-sekolah yang dipilih secara acak. Serta tes ini bersifat diagnostik, dalam tes ini yang dinilai yaitu membaca, matematika dan sains. Hasil tes tersebut Indonesia menduduki peringkat 70 dari 78 negara yang mengikuti tes pada saat itu

[13]. Selain itu, berdasarkan survei Sosial Ekonomi pada Maret 2020 dalam Statistik Pendidikan 2020 menunjukkan bahwa Indonesia belum mencapai status Tuntas Paripurna dalam program wajib belajar tingkat SMA/MA sederajat. Hal ini tercermin dari Angka Partisipasi Kasar yang masih di bawah 95%. Selain itu, masih ada 1,13% penduduk yang tidak menyelesaikan pendidikan di tingkat tersebut. Melihat kondisi ini, perlu dilakukan pengelompokan provinsi di Indonesia berdasarkan indikator-indikator pendidikan menggunakan metode clustering. Langkah ini penting untuk memahami variasi kinerja pendidikan antar provinsi dan merancang kebijakan yang lebih sesuai dengan kebutuhan masing-masing daerah [14].

Dengan demikian berdasarkan uraian diatas, artikel ini akan menggunakan metode Self Organizing Maps untuk melakukan pengelompokan pendidikan di Indonesia berdasarkan provinsi-provinsi di Indonesia dan aspek-aspek pendidikan. Artikel ini diharapkan dapat memberikan informasi perencanaan yang diperlukan pemerintah yaitu pemerataan pendidikan selain itu juga pemerintah dapat memperbaiki dan merancang program pendidikan di Indonesia yang lebih efektif untuk meningkatkan kualitas pendidikan di Indonesia.

## B. Metode Penelitian

Artikel ini menggunakan data sekunder yaitu data yang digunakan dalam artikel ini adalah data pendidikan yang dibagi dalam 5 aspek yaitu angka melek huruf penduduk yang berumur 15 tahun keatas, angka partisipasi murni berdasarkan jenjang SMA atau sederajat, angka partisipasi sekolah penduduk yang berumur 16-18, rata-rata lama sekolah penduduk umur 15 tahun ke atas dan angka partisipasi Kasar berdasarkan jenjang SMA atau sederajat berdasarkan provinsi yang berjumlah 34 provinsi pada tahun 2023.

### Langkah dalam Analisis Data

Tahapan penelitian merupakan serangkaian tahapan yang dilakukan dalam penelitian. Tahapan-tahapan dalam artikel ini mencakup dari mulai pengambilan data, persiapan data dan hasil. Penelitian ini dilakukan dengan bantuan *software R* dan dilakukan melalui beberapa tahapan sebagai berikut :

Pertama, melakukan standarisasi data menggunakan persamaan 1.

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \tag{1}$$

Kedua, menentukan jumlah kluster yang akan dibentuk.

Ketiga, menentukan jenis topologi. Topologi yang digunakan adalah topologi heksagonal (*hexagonal topology*), Ukuran topologi disesuaikan berdasarkan jumlah kluster yang akan dibuat, yaitu lebih banyak atau sama dengan jumlah kluster yang akan dibuat. Sedangkan untuk nilai ukuran ketetanggaan (R) awal adalah 1, hal ini dapat berubah dengan ketentuan hasil kluster yang dihasilkan terdapat kluster yang tidak memiliki anggota, untuk nilai  $\alpha$  atau learning rate pada iterasi pertama yaitu 0,1 dan akan berubah pada iterasi selanjutnya yaitu 0,5.

Keempat, menginisialisasikan bobot *neuron* output secara acak. Bobot acak pada setiap *neuron* ditentukan dengan rentang nilai antara 0 dan 1.

Kelima, menghitung jarak minimum antara data input terhadap *neuron* dan menentukan BMU menggunakan jarak euclidean pada persamaan (2).

$$D_{(ij)} = \sqrt{\sum_{i=1}^n (W_{ik} - x_{ik})^2} \tag{2}$$

Keenam, memperbarui nilai bobot *neuron* yang menjadi *neuron* pemenang pada persamaan (3) dan *neuron* tetangga pada persamaan (4).

$$W_{ij,new} = W_{ij,current} + \alpha(x_{ij} - W_{ij,current}) \tag{3}$$

$$W_{ij,new} = W_{ij,current} + \alpha * h_{ci}(x_{ij} - W_{ij,current}) \tag{4}$$

Ketujuh, mengelompokkan neuron yang telah diperbarui menjadi beberapa kluster pada persamaan (5).

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \tag{5}$$

Kedelapan, melakukan proses iterasi, proses ini dilakukan hingga kriteria konvergensi tercapai, yaitu ketika keanggotaan *neuron* dalam kelompok-kelompok (klaster) tidak mengalami perubahan lagi.

Kesembilan, melakukan Evaluasi cluster menggunakan metode DBI pada persamaan (6) sampai (10). Menghitung nilai jarak dalam cluster ke-*j* dengan rumus *Sum of Square Within Cluster* (SSW) sebagai berikut :

$$SSW_l = \frac{1}{n_l} \sum_{i=1}^{n_l} d(x_i, c_l) \tag{6}$$

Menghitung nilai jarak antar cluster dengan rumus *Sum of Square Between Cluster* (SSB) sebagai berikut :

$$SSB_{l,m} = d(c_l, c_m) \tag{7}$$

Membandingkan nilai *cluster* ke-*l* dan *cluster* ke-*m* dengan rumus :

$$R_{l,m} = \frac{SSW_l + SSW_m}{SSB_{l,m}} \tag{8}$$

Mencari nilai *Davies-Bouldin Index* (DBI) dengan rumus :

$$DBI = \frac{1}{K} \sum_{i=1}^k \max_{l \neq m} (R_{l,m}) \tag{9}$$

### C. Hasil dan Pembahasan

#### Standarisasi Data

Standarisasi data dilakukan untuk mengatasi perbedaan skala pada variabel, yang dapat mempengaruhi hasil analisis dan interpretasi. Dalam artikel ini, standarisasi dilakukan dengan metode Z-score, yang mengubah setiap nilai dalam dataset sehingga memiliki mean nol dan standar deviasi satu. Proses ini memastikan bahwa semua variabel memiliki skala yang sama, memungkinkan perbandingan yang lebih adil dan akurat. Dengan standarisasi Z-score, pengaruh skala variabel yang berbeda dihilangkan, sehingga analisis menjadi lebih reliabel dan valid.

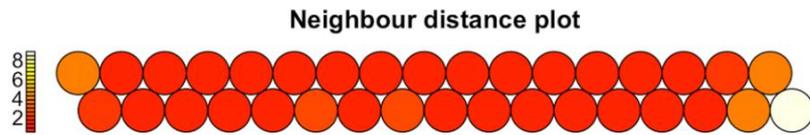
**Tabel 1.** Data Hasil Standarisasi

Provinsi	AMH	APS	APM	APS	RLS
Aceh	0,44	-0,80	1,32	0,65	0,72
Sumatera Utara	0,70	0,79	0,81	0,05	0,94
Sumatera Barat	0,74	0,59	0,89	1,65	0,35
Riau	0,70	0,29	0,16	0,23	0,36
⋮	⋮	⋮	⋮	⋮	⋮
Sulawesi Barat	-0,80	0,07	-0,46	-0,76	-1,01
Maluku	0,79	1,17	0,38	1,38	1,32
Maluku Utara	0,59	1,12	0,31	0,21	0,37
Papua Barat	0,29	1,31	0,34	0,72	1,02

#### Pengelompokkan Menggunakan Self Organizing Maps

Metode Self Organizing Maps (SOM) melibatkan empat tahap utama dalam prosesnya. Pertama, inialisasi bobot dilakukan secara acak. Selanjutnya, jarak antara data input dan bobot dihitung, dimana hasil perhitungan ini digunakan untuk menentukan jarak terdekat atau nilai minimum. Tahap ketiga melibatkan pembaruan bobot menggunakan nilai learning rate yang pada iterasi pertama telah ditetapkan. dan yang keempat untuk iterasi selanjutnya menggunakan learning rate yang telah diperbarui.

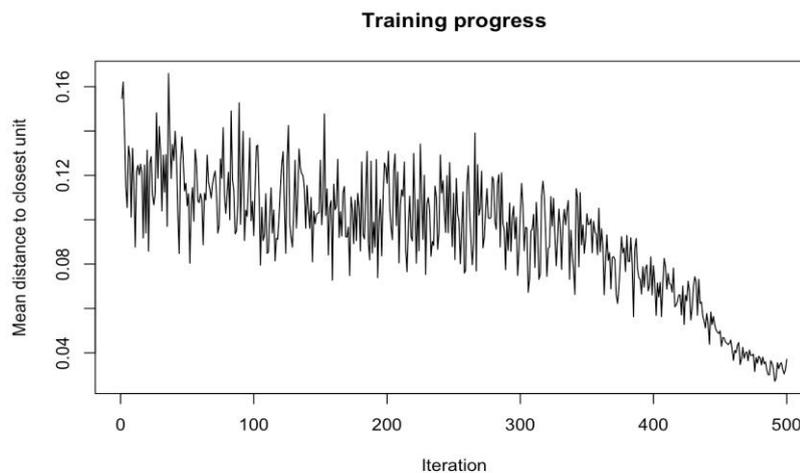
Proses dimulai dengan penentuan bobot awal secara acak. Struktur matriks bobot ditentukan oleh jumlah komponen dalam vektor input, dalam penelitian ini terdapat 5 variabel yang mewakili kolom matriks, sementara jumlah maksimum cluster yang akan dibentuk adalah 2 cluster, 3 cluster, 4 cluster dan 5 cluster. Jumlah cluster digunakan untuk menentukan jumlah baris matriks. Parameter awal ditetapkan dengan learning rate sebesar 0,1 sampai 0,5 dan jumlah maksimum iterasi sebanyak 500. Pendekatan ini memungkinkan SOM untuk mengorganisir data secara mandiri, mengidentifikasi pola, dan membentuk representasi topologi dari data input multidimensi dalam ruang output yang biasanya berdimensi lebih rendah.



**Gambar 1.** *Neighbour Distance Plot*

Gambar 1. menunjukkan jarak antara neuron tetangga pada peta Self-Organizing Maps (SOM), yang digunakan untuk mengidentifikasi kluster dalam data. Dalam plot ini, warna mencerminkan jarak antara neuron-neuron tersebut. Warna terang seperti kuning atau putih menunjukkan jarak yang lebih besar, menandakan perbedaan yang lebih signifikan antara data yang diwakili oleh neuron-neuron tersebut. Sebaliknya, warna gelap seperti merah menunjukkan jarak yang lebih kecil, menandakan bahwa data yang diwakili sangat mirip, seringkali membentuk kluster yang kompak.

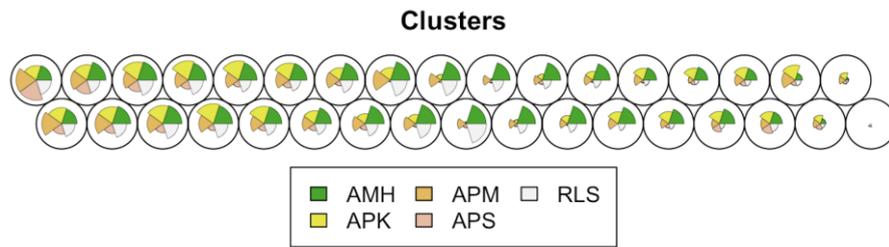
Area dengan warna merah seragam menunjukkan kluster dalam data, sementara transisi besar dalam warna (dari merah ke kuning atau putih) menunjukkan batas antara kluster-kluster tersebut. Dalam gambar ini, sebagian besar neuron memiliki jarak kecil satu sama lain (warna merah), namun ada beberapa area dengan jarak lebih besar (warna kuning dan putih), menunjukkan variasi dalam data yang diwakili.



**Gambar 2.** *Grafik Training Progress*

Gambar 2. menampilkan grafik kemajuan pelatihan dari Self-Organizing Map (SOM). Sumbu horizontal merepresentasikan jumlah iterasi selama proses pelatihan, sedangkan sumbu vertikal menunjukkan jarak rata-rata ke neuron terdekat, yang digunakan sebagai metrik untuk menilai seberapa representatif peta SOM terhadap data input. Grafik ini menunjukkan bahwa pelatihan SOM berjalan dengan baik, dengan penurunan jarak rata-rata yang menandakan bahwa model semakin baik dalam memetakan data ke neuron-neuronnya. Proses konvergensi mulai terlihat sekitar iterasi ke-300, di mana jarak rata-rata mulai berkurang secara konsisten. Stabilitas jarak rata-rata tercapai sekitar iterasi ke-400 hingga ke-500, menunjukkan bahwa pelatihan dapat dihentikan karena model sudah stabil dan representatif terhadap data input.

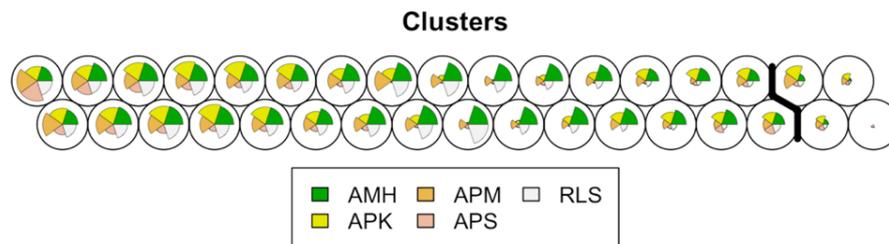
Visualisasi hasil SOM, yang divisualisasikan melalui software R, menunjukkan diagram lingkaran-lingkaran yang berdekatan secara topologis jika memiliki karakteristik serupa, menggambarkan bagaimana data terstruktur dalam peta SOM tersebut.



**Gambar 3.** Topologi Cluster

Gambar 3. menampilkan peta topologi kluster di mana setiap lingkaran merepresentasikan satu objek provinsi, dan di dalam lingkaran tersebut terdapat diagram pie chart yang menunjukkan distribusi kategori atau variabel dalam objek tersebut. Setiap irisan dalam pie chart mewakili proporsi satu variabel dalam unit tersebut, dengan total lima variabel yang diwakili oleh warna-warna berbeda. Dominasi suatu warna dalam pie chart menandakan bahwa variabel tersebut paling menonjol atau berpengaruh dalam lingkaran atau provinsi yang direpresentasikan. Misalnya, jika warna hijau (AMH) mendominasi sebuah pie chart, ini menunjukkan bahwa variabel AMH memiliki proporsi terbesar atau paling signifikan dalam provinsi tersebut. Variasi komposisi warna antar pie chart memungkinkan perbandingan karakteristik dan dominasi variabel di berbagai cluster atau provinsi, memberikan gambaran visual yang jelas tentang perbedaan distribusi variabel di seluruh dataset.

**Pengelompokan untuk 2 Cluster**



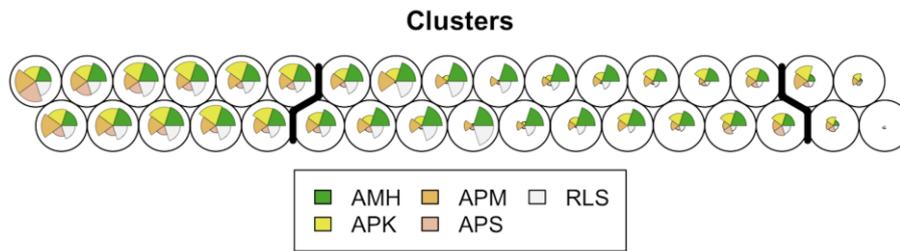
**Gambar 4.** Hasil Pengelompokan Dengan 2 Cluster

Gambar 4. menampilkan hasil pengelompokan 34 provinsi di Indonesia menggunakan metode *Self-Organizing Map* (SOM), yang menghasilkan dua kluster utama. Kluster pertama terdiri dari 32 provinsi, yang umumnya memiliki nilai tinggi dalam indikator pendidikan seperti Angka Melek Huruf (AMH) dan Angka Partisipasi Kasar (APK). Hal ini ditunjukkan oleh dominasi warna hijau (AMH) dan kuning (APK) dalam diagram pie di kluster tersebut, mengindikasikan tingkat melek huruf dan partisipasi pendidikan yang lebih tinggi. Sebaliknya, kluster kedua hanya terdiri dari 2 provinsi, yang menunjukkan karakteristik berbeda. Dalam kluster ini, indikator pendidikan mungkin lebih rendah, sebagaimana terlihat dari diagram pie yang memiliki warna selain hijau dan kuning yang lebih dominan. Hal ini menunjukkan bahwa kedua provinsi ini mungkin memerlukan perhatian khusus dalam kebijakan pendidikan untuk meningkatkan indikator pendidikan mereka.

**Tabel 2.** Jumlah dan Anggota 2 Cluster

Cluster	Jumlah Anggota	Anggota Cluster (Provinsi)
1	32	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat
2	2	Nusa Tenggara Barat, Papua

**Pengelompokkan untuk 3 Cluster**



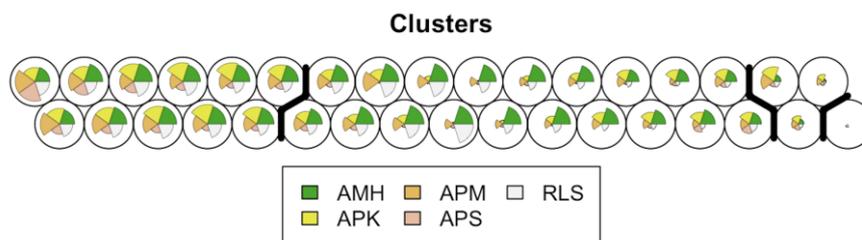
**Gambar 5.** Hasil Pengelompokkan Dengan 3 Cluster

Gambar 5. menunjukkan hasil pengelompokkan 34 provinsi di Indonesia menggunakan metode Self-Organizing Map (SOM), yang mengidentifikasi tiga kluster utama berdasarkan indikator pendidikan. Kluster pertama terdiri dari provinsi-provinsi dengan indikator pendidikan yang lebih baik, seperti Angka Melek Huruf (AMH) dan Angka Partisipasi Kasar (APK), yang terlihat dari dominasi warna hijau dan kuning pada diagram pie, menunjukkan kinerja yang baik dalam indikator tersebut. Kluster kedua mencakup provinsi-provinsi dengan kinerja pendidikan yang cukup baik namun masih memerlukan perbaikan pada beberapa indikator, yang tercermin dari variasi warna yang lebih seimbang pada diagram pie. Sementara itu, kluster ketiga terdiri dari provinsi-provinsi yang memiliki kinerja pendidikan yang membutuhkan perhatian khusus, dengan indikator-indikator yang lebih rendah dibandingkan kluster lainnya. Hal ini terlihat dari dominasi warna selain hijau dan kuning pada diagram pie, yang menunjukkan bahwa provinsi-provinsi dalam kluster ini memerlukan intervensi yang lebih fokus untuk meningkatkan kualitas pendidikan mereka.

**Tabel 3.** Jumlah dan Anggota 3 Cluster

Cluster	Jumlah Anggota	Anggota Cluster (Provinsi)
1	11	Aceh, Sumatera Utara, Sumatera Barat, Bengkulu, DI Yogyakarta, Bali, Kalimantan Timur, Kalimantan Utara, Maluku, Maluku Utara, Papua Barat
2	21	Riau, Jambi, Sumatera Selatan, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat
3	2	Nusa Tenggara Barat, Papua

**Pengelompokkan untuk 4 Cluster**



**Gambar 6.** Hasil Pengelompokkan Dengan 4 Cluster

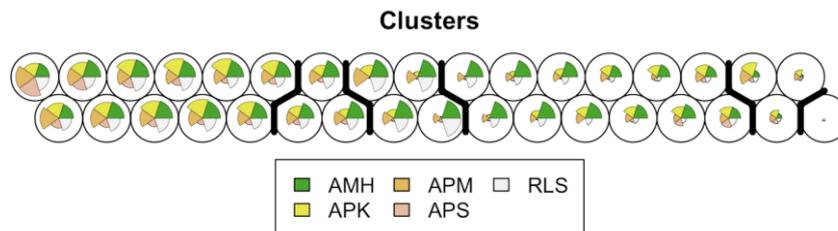
Gambar 6. menunjukkan hasil pengelompokkan 34 provinsi di Indonesia menggunakan metode Self-Organizing Map (SOM), yang menghasilkan empat kluster utama berdasarkan indikator pendidikan. Kluster pertama terdiri dari provinsi-provinsi dengan indikator pendidikan yang lebih baik, seperti Angka Melek Huruf (AMH) dan Angka Partisipasi Kasar (APK), yang terlihat dari dominasi warna hijau dan kuning pada diagram pie, menunjukkan kinerja pendidikan yang baik. Kluster kedua mencakup provinsi-provinsi dengan kinerja pendidikan yang cukup baik, namun masih memerlukan perbaikan pada beberapa indikator, tercermin dari

variasi warna yang lebih seimbang pada diagram pie. Klaster ketiga terdiri dari provinsi-provinsi dengan kinerja pendidikan yang lebih rendah dibandingkan klaster 1 dan 2, dan memerlukan intervensi yang lebih fokus untuk meningkatkan indikator-indikator pendidikan mereka. Klaster keempat menunjukkan provinsi-provinsi dengan kinerja pendidikan yang paling rendah di antara semua klaster, dengan nilai indikator pendidikan yang sangat membutuhkan perhatian khusus. Provinsi-provinsi dalam klaster ini memerlukan intervensi yang lebih intensif dan kebijakan yang spesifik untuk mengatasi tantangan pendidikan yang mereka hadapi.

**Tabel 4.** Jumlah dan Anggota 4 Cluster

Cluster	Jumlah Anggota	Anggota Cluster (Provinsi)
1	11	Aceh, Sumatera Utara, Sumatera Barat, Bengkulu, DI Yogyakarta, Bali, Kalimantan Timur, Kalimantan Utara, Maluku, Maluku Utara, Papua Barat
2	21	Riau, Jambi, Sumatera Selatan, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat
3	1	Nusa Tenggara Barat
4	1	Papua

**Pengelompokan untuk 5 Cluster**



**Gambar 7.** Hasil Pengelompokan Dengan 5 Cluster

Gambar 7. menunjukkan hasil pengelompokan 34 provinsi di Indonesia menggunakan metode *Self-Organizing Map* (SOM), yang menghasilkan lima klaster utama berdasarkan indikator pendidikan. Klaster pertama terdiri dari provinsi-provinsi dengan indikator pendidikan yang lebih baik, seperti Angka Melek Huruf (AMH) dan Angka Partisipasi Kasar (APK), yang terlihat dari dominasi warna hijau dan kuning pada diagram pie, menunjukkan kinerja pendidikan yang baik. Klaster kedua mencakup provinsi-provinsi dengan kinerja pendidikan yang cukup baik, namun masih memerlukan perbaikan pada beberapa indikator, yang tercermin dari variasi warna yang lebih seimbang pada diagram pie. Klaster ketiga terdiri dari provinsi-provinsi dengan karakteristik yang sedikit berbeda dan memerlukan strategi yang lebih terarah untuk memperbaiki indikator pendidikan tertentu yang membutuhkan perhatian tambahan. Klaster keempat mencakup provinsi-provinsi dengan kinerja pendidikan yang lebih rendah dibandingkan klaster sebelumnya, memerlukan intervensi yang lebih fokus untuk meningkatkan indikator pendidikan mereka. Klaster kelima menunjukkan provinsi-provinsi dengan kinerja pendidikan paling rendah di antara semua klaster, di mana nilai indikator pendidikan sangat memerlukan perhatian khusus. Intervensi yang lebih intensif dan kebijakan spesifik diperlukan untuk mengatasi tantangan pendidikan di provinsi-provinsi ini.

**Tabel 5.** Jumlah dan Anggota 4 Cluster

Cluster	Jumlah Anggota	Anggota Cluster (Provinsi)
1	10	Aceh, Sumatera Utara, Sumatera Barat, Bengkulu, DI Yogyakarta, Bali, Kalimantan Timur, Kalimantan Utara, Maluku, Maluku Utara, Papua Barat

Cluster	Jumlah Anggota	Anggota Cluster (Provinsi)
2	18	Riau, Jambi, Sumatera Selatan, Lampung, Kepulauan Bangka Belitung, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat
3	3	Kepulauan Riau, DKI Jakarta, Sulawesi Utara
4	1	Nusa Tenggara barat
5	1	Papua

**Evaluasi Cluster**

Adapun hasil perhitungan nilai DBI dengan langkah yang sama diperoleh nilai DBI untuk setiap K adalah sebagai berikut:

**Tabel 6.** Nilai Davies-Bouldin Index

Jumlah Klaster	Nilai DBI
2	0.6534
3	1.6906
4	3.3603
5	4.34

Berdasarkan Tabel 6. diperoleh bahwa nilai DBI untuk jumlah cluster optimal pada pengelompokan Provinsi di Indonesia berdasarkan aspek pendidikan adalah 2 cluster. Hal ini ditunjukkan oleh nilai DBI terendah, yaitu 0.6534. Clustering dengan 2 cluster memberikan hasil yang paling optimal.

**D. Kesimpulan**

Berdasarkan hasil pengelompokan 34 provinsi di Indonesia menggunakan metode Self-Organizing Map (SOM) berdasarkan aspek pendidikan, didapatkan hasil klaster optimum sejumlah 2 klaster. Klaster pertama terdiri dari 32 provinsi, sementara klaster kedua hanya terdiri dari 2 provinsi. Karakteristik utama provinsi dalam klaster pertama adalah memiliki nilai yang cukup tinggi dalam aspek pendidikan seperti Angka Melek Huruf (AMH) dan Angka Partisipasi Kasar (APK). Klaster kedua, yang terdiri dari 2 provinsi, menunjukkan karakteristik yang berbeda dibandingkan dengan klaster pertama. Provinsi-provinsi dalam klaster ini cenderung memiliki nilai yang lebih rendah untuk beberapa aspek pendidikan. Tingkat melek huruf (AMH) di provinsi-provinsi ini mungkin lebih rendah dibandingkan dengan yang ada di klaster pertama dan juga angka partisipasi kasar (APK) lebih rendah, menunjukkan bahwa partisipasi pendidikan di provinsi-provinsi ini perlu ditingkatkan.

**Daftar Pustaka**

- [1] L. R. Iyohu, Ismail Djakaria, and La Ode Nashar, “Perbandingan Metode K-Means Clustering dengan Self-Organizing Maps (SOM) untuk Pengelompokan Provinsi di Indonesia Berdasarkan Data Potensi Desa,” *Jurnal Statistika dan Aplikasinya*, vol. 7, no. 2, pp. 195–206, Dec. 2023, doi: 10.21009/JSA.07208.
- [2] P. R. Zain and T. S. Yanti, “Penerapan Multidimensional Scaling terhadap Pengelompokan Provinsi Berdasarkan Indikator Tujuan ke-6 SDGs,” *Jurnal Riset Statistika*, pp. 153–160, Dec. 2022, doi: 10.29313/jrs.v2i2.1465.
- [3] F. Zubedi, “Analisis Cluster Hasil Try Out Siswa MTS AlHuda Gorontalo dengan Chi-Sim Cosimilarity dan K-Means Clustering,” *JMPM: Jurnal Matematika dan Pendidikan Matematika*, vol. 5, no. 1, pp. 1–11, Mar. 2020, doi: 10.26594/jmpm.v4i2.1706.
- [4] D. M. Midyanti and S. Bahri, “Implementasi Self Organizing Maps untuk Pengelompokan Kabupaten/Kota Berdasarkan Indeks Pembangunan Manusia,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 6, pp. 1265–1272, Dec. 2023, doi: 10.25126/jtiik.1067647.

- [5] M. N. Zain, “Algoritma Artificial Neural Network dalam Klasifikasi Chest X-Rays Pasien COVID-19,” *Jurnal Riset Statistika*, pp. 137–144, Dec. 2022, doi: 10.29313/jrs.v2i2.1426.
- [6] W. Aryanti and N. A. K. Rifai, “Penerapan Artificial Neural Network dengan Algoritma Backpropagation untuk Memprediksi Harga Saham,” *Jurnal Riset Statistika*, vol. 3, no. 2, pp. 107–118, Dec. 2023, doi: 10.29313/jrs.v3i2.2953.
- [7] E. Harli, A. Fauzi, and T. H. Kusmanto, “Pengelompokan Kelas Menggunakan Self Organizing Map Neural Network pada SMK N 1 Depok,” *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 2, no. 2, Nov. 2016, doi: 10.26418/jp.v2i2.17574.
- [8] Widhi Aryanti and Nur Azizah Komara Rifai, “Penerapan Artificial Neural Network dengan Algoritma Backpropagation untuk Memprediksi Harga Saham,” *Jurnal Riset Statistika*, pp. 107–118, Dec. 2023, doi: 10.29313/jrs.v3i2.2953.
- [9] N. Imani, A. I. Alfassa, and A. M. Yolanda, “Self Organizing Map (SOM) Clustering Untuk Analisis Data Indikator Sosial di Provinsi Nusa Tenggara Timur,” *Jurnal Gaussian*, vol. 11, no. 3, pp. 458–467, Jan. 2023, doi: 10.14710/j.gauss.11.3.458-467.
- [10] E. Damanik and E. Simamora, “Estimasi Interval Kepercayaan Parameter Selisih Rata-rata IPK Kelas Pendidikan Reguler dan Ekstensi FMIPA Angkatan 2010 Unimed dengan Bootstrap Pensentil,” *Karismatika*, vol. 5, no. 3, pp. 1–9, 2019.
- [11] Z. Soraya, “Penguatan Pendidikan Karakter untuk Membangun Peradaban Bangsa,” *Southeast Asian Journal of Islamic Education Management*, vol. 1, no. 1, pp. 74–81, Jun. 2020, doi: 10.21154/sajiem.v1i1.10.
- [12] D. Camartya and A. I. Achmad, “Analisis Korespondensi pada Jumlah Pengangguran Terbuka Menurut Kabupaten/Kota Berdasarkan Pendidikan Tertinggi,” *Jurnal Riset Statistika*, pp. 119–128, Dec. 2022, doi: 10.29313/jrs.v2i2.1424.
- [13] P. A. Rizkianti, M. Asbari, N. P. Priambudi, and S. A. J. Asri, “Pendidikan Indonesia Masih Buruk?,” *Journal Of Information Systems And Management*, vol. 3, no. 2, 2024.
- [14] D. A. Alodia, A. P. Fialine, D. Endriani, and E. Widodo, “Implementasi Metode K-Medoids Clustering untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan,” *Sepren*, vol. 2, no. 2, pp. 1–13, Dec. 2021, doi: 10.36655/sepren.v2i2.606.