

Pemetaan Kinerja Pendidikan Provinsi di Indonesia Menggunakan Algoritma K-Means Berdasarkan Indikator Pendidikan Tahun 2025

Sutan M Arif¹, Indah Pangesti², Fadjriah Hapsari³

Teknik Informatika, FTIK, Universitas Indraprasta PGRI, Jakarta, Indonesia¹

Pendidikan Bahasa Inggris, FBS, Universitas Indraprasta PGRI, Jakarta, Indonesia²

Pendidikan Ekonomi, FIPPS, Universitas Indraprasta PGRI, Jakarta, Indonesia³

*Email Korespondensi: cutans.muhs@gmail.com

Diterima: 26-07-2026 | Disetujui: 01-08-2026 | Diterbitkan: 03-08-2026

ABSTRACT

Education is one of the key indicators in measuring the quality of human resources and regional development. This study aims to map the educational performance of provinces in Indonesia using the K-Means clustering algorithm based on three educational indicators, namely Average Years of Schooling (AYS), Expected Years of Schooling (EYS), and School Participation Rate (SPR) for the 16–18 age group in 2025. The study employed a quantitative approach using secondary data obtained from Statistics Indonesia (BPS), covering 38 provinces. Data were processed through descriptive statistical analysis, Z-score normalization, determination of the optimal number of clusters using the Elbow Method, clustering with the K-Means algorithm, and cluster quality evaluation using the Silhouette Score. The results identified three clusters representing high, moderate, and low educational performance. Most provinces were classified into the moderate-performance cluster, while a small number of provinces belonged to the low-performance cluster. The Silhouette Score of 0.453 indicates that the clustering results have acceptable quality and adequately represent the characteristics of educational performance across provinces. The findings provide useful information for supporting educational planning and policy formulation to reduce regional disparities in Indonesia.

Keywords: K-Means; Educational Performance; Average Years of Schooling; School Participation Rate; Clustering.

ABSTRAK

Penelitian ini bertujuan memetakan kinerja pendidikan provinsi di Indonesia menggunakan algoritma K-Means berdasarkan tiga indikator pendidikan, yaitu Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), dan Angka Partisipasi Sekolah (APS) usia 16–18 tahun pada tahun 2025. Penelitian menggunakan pendekatan kuantitatif dengan memanfaatkan data sekunder yang bersumber dari Badan Pusat Statistik (BPS) pada 38 provinsi di Indonesia. Pengolahan data dilakukan melalui analisis statistik deskriptif, normalisasi Z-Score, penentuan jumlah kluster menggunakan *Elbow Method*, proses klusterisasi menggunakan algoritma K-Means, serta evaluasi kualitas kluster menggunakan Silhouette Score. Hasil penelitian menunjukkan terbentuknya tiga kelompok provinsi, yaitu kinerja pendidikan tinggi, sedang, dan rendah. Sebagian besar provinsi berada pada kelompok kinerja pendidikan sedang, sedangkan hanya sebagian kecil yang termasuk kelompok kinerja rendah. Nilai Silhouette Score sebesar 0,453 menunjukkan bahwa hasil pengelompokan memiliki kualitas yang cukup baik dalam merepresentasikan karakteristik pendidikan antarprovinsi. Hasil penelitian diharapkan dapat menjadi referensi bagi pemerintah dalam mendukung perumusan kebijakan pemerataan dan peningkatan kualitas pendidikan di Indonesia.

Katakunci: K-Means; Kinerja Pendidikan; Rata-rata Lama Sekolah; Angka Partisipasi Sekolah; Klasterisasi.

Bagaimana Cara Sitasi Artikel ini:

Arif, S. M., Pangesti, I. ., & Hapsari, F. . (2026). Pemetaan Kinerja Pendidikan Provinsi di Indonesia Menggunakan Algoritma K-Means Berdasarkan Indikator Pendidikan Tahun 2025. Educational Journal, 2(1), 18-28. <https://doi.org/10.63822/nxy4da75>

PENDAHULUAN

Pendidikan merupakan salah satu indikator utama dalam pembangunan manusia karena berperan penting dalam meningkatkan kualitas sumber daya manusia. Peningkatan kualitas pendidikan berkontribusi terhadap pertumbuhan ekonomi, kesejahteraan masyarakat, dan daya saing suatu bangsa (Supriyadi & Martini, 2023). Selain itu, pendidikan menjadi faktor yang mendukung peningkatan Indeks Pembangunan Manusia (IPM) melalui peningkatan kompetensi, produktivitas, dan kemampuan individu dalam menghadapi perubahan (Delima et al., 2025). Oleh karena itu, pemerataan kualitas pendidikan di setiap daerah menjadi aspek penting dalam mendukung pembangunan yang berkelanjutan (Juita et al., 2024).

Meskipun pembangunan pendidikan di Indonesia terus mengalami peningkatan, kesenjangan kualitas pendidikan antardaerah masih menjadi tantangan yang perlu mendapat perhatian. Perbedaan akses terhadap layanan pendidikan, ketersediaan sarana dan prasarana, kualitas tenaga pendidik, serta kondisi sosial ekonomi menyebabkan capaian indikator pendidikan setiap provinsi tidak merata. Sejumlah penelitian menunjukkan bahwa provinsi di wilayah Indonesia timur cenderung memiliki tingkat kinerja pendidikan yang lebih rendah dibandingkan dengan provinsi di wilayah barat, khususnya Pulau Jawa dan Bali (Kamalia & Nawangsih, 2025; Nugraha, 2023; Wardani et al., 2025). Kondisi tersebut menunjukkan perlunya pemetaan kinerja pendidikan secara komprehensif sebagai dasar dalam penyusunan kebijakan yang lebih tepat sasaran untuk mengurangi kesenjangan pendidikan antarprovinsi.

Penelitian ini menggunakan tiga indikator pendidikan, yaitu Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), dan Angka Partisipasi Sekolah (APS) sebagai dasar dalam mengukur kinerja pendidikan setiap provinsi. RLS menggambarkan rata-rata jumlah tahun pendidikan yang telah ditempuh penduduk, sedangkan HLS menunjukkan peluang lama pendidikan yang diharapkan dapat dicapai oleh penduduk usia sekolah pada masa mendatang. Sementara itu, APS mencerminkan tingkat partisipasi penduduk usia sekolah dalam mengikuti pendidikan formal sehingga dapat digunakan untuk menilai akses terhadap layanan pendidikan (Alim et al., 2024; Maryam, 2025). Ketiga indikator tersebut saling melengkapi dalam memberikan gambaran mengenai kondisi pendidikan suatu daerah dan menjadi dasar yang relevan untuk proses pengelompokan menggunakan algoritma K-Means (Fitria et al., 2025).

Analisis data pendidikan menjadi salah satu pendekatan penting dalam mendukung pengambilan kebijakan yang berbasis data. Melalui teknik *clustering*, pemerintah dapat mengidentifikasi karakteristik dan tingkat kesamaan kondisi pendidikan antarprovinsi sehingga kebijakan yang disusun menjadi lebih tepat sasaran. Berbagai penelitian menunjukkan bahwa metode *clustering*, seperti *Self Organizing Maps*, *Fuzzy C-Means*, dan K-Means, mampu mengelompokkan wilayah berdasarkan indikator pendidikan serta mengidentifikasi perbedaan tingkat kinerja pendidikan antarprovinsi (Faqih & Mahdy, 2024; Maulidya et al., 2022; Vidyananta et al., 2025). Berdasarkan uraian tersebut, penelitian ini menerapkan algoritma K-Means untuk mengelompokkan provinsi di Indonesia berdasarkan indikator pendidikan sehingga diperoleh karakteristik kinerja pendidikan yang dapat menjadi dasar dalam mendukung penyusunan kebijakan pemerataan pendidikan.

Berbagai penelitian telah menerapkan algoritma K-Means pada bidang pendidikan untuk mengelompokkan data berdasarkan karakteristik tertentu, seperti pemetaan kualitas pendidikan, pengelompokan sekolah, dan identifikasi siswa berprestasi. Hasil penelitian menunjukkan bahwa algoritma K-Means mampu menghasilkan pengelompokan data secara efektif dengan proses komputasi yang relatif sederhana dan efisien (Chong, 2021; Mubarak & Rita Wahyuni Arifin, 2025). Meskipun demikian, kualitas

hasil pengelompokan sangat dipengaruhi oleh penentuan jumlah klaster dan karakteristik data yang digunakan. Oleh sebab itu, algoritma K-Means masih relevan untuk diterapkan dalam analisis data pendidikan, khususnya dalam mengidentifikasi karakteristik kinerja pendidikan antarprovinsi berdasarkan indikator pendidikan.

Meskipun berbagai penelitian telah menerapkan algoritma K-Means untuk mengelompokkan data pendidikan, sebagian besar masih berfokus pada analisis wilayah tertentu atau menggunakan indikator pendidikan yang terbatas. Selain itu, penelitian yang memanfaatkan data Badan Pusat Statistik (BPS) Tahun 2025 dengan mengombinasikan indikator Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), dan Angka Partisipasi Sekolah (APS) untuk memetakan kinerja pendidikan seluruh provinsi di Indonesia masih relatif terbatas. Kondisi tersebut menunjukkan adanya peluang untuk menghasilkan pemetaan yang lebih komprehensif sebagai dasar dalam memahami karakteristik kinerja pendidikan antarprovinsi.

Berdasarkan permasalahan dan *research gap* tersebut, penelitian ini bertujuan mengelompokkan provinsi di Indonesia berdasarkan tingkat kinerja pendidikan menggunakan algoritma K-Means dengan memanfaatkan indikator Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), dan Angka Partisipasi Sekolah (APS) Tahun 2025. Hasil pengelompokan diharapkan dapat mengidentifikasi karakteristik provinsi ke dalam kategori kinerja pendidikan tinggi, sedang, dan rendah, sehingga memberikan gambaran yang lebih jelas mengenai kondisi pendidikan di Indonesia.

Penelitian ini diharapkan memberikan kontribusi secara akademis maupun praktis. Secara akademis, penelitian ini memperkaya kajian mengenai penerapan algoritma K-Means dalam pemetaan kinerja pendidikan berbasis data sekunder. Secara praktis, hasil pengelompokan provinsi dapat menjadi informasi pendukung bagi pemerintah pusat maupun pemerintah daerah dalam merumuskan kebijakan pemerataan pendidikan, menetapkan prioritas pembangunan sektor pendidikan, serta mengevaluasi capaian indikator pendidikan secara lebih terarah.

METODE PENELITIAN

Jenis Penelitian

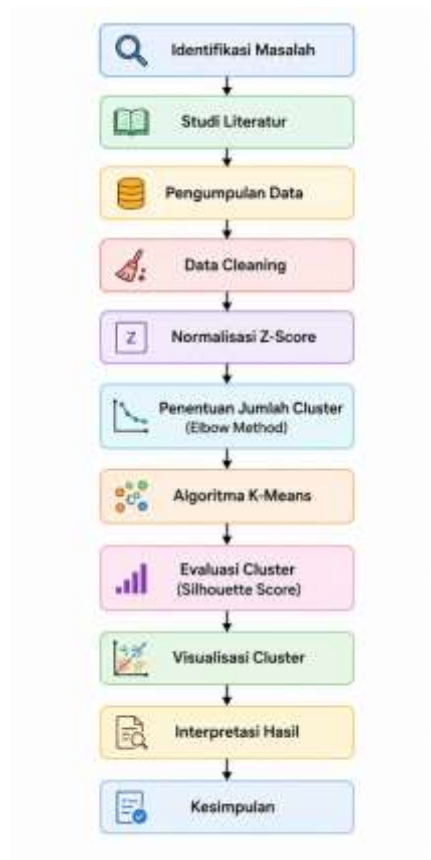
Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining untuk mengelompokkan provinsi di Indonesia berdasarkan karakteristik indikator pendidikan. Proses pengelompokan dilakukan menggunakan algoritma K-Means karena mampu mengelompokkan objek berdasarkan tingkat kemiripan karakteristik secara efektif dan efisien. Objek penelitian terdiri atas 38 provinsi di Indonesia, sedangkan variabel yang digunakan meliputi Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), dan Angka Partisipasi Sekolah (APS) usia 16–18 tahun yang diperoleh dari Badan Pusat Statistik (BPS) Tahun 2025. Hasil pengelompokan diharapkan dapat menggambarkan karakteristik kinerja pendidikan setiap provinsi sehingga menjadi informasi pendukung dalam penyusunan kebijakan pemerataan pendidikan.

Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari publikasi resmi Badan Pusat Statistik (BPS) Tahun 2025. Data tersebut dipilih karena bersifat valid, terstandar, dan merepresentasikan kondisi pendidikan pada seluruh provinsi di Indonesia. Objek penelitian mencakup 38 provinsi di Indonesia dengan

menggunakan tiga indikator pendidikan, yaitu Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), dan Angka Partisipasi Sekolah (APS) usia 16–18 tahun. Ketiga indikator tersebut dipilih karena mampu menggambarkan capaian pendidikan, harapan keberlanjutan pendidikan, serta tingkat partisipasi penduduk usia sekolah pada jenjang pendidikan menengah. Data yang telah dikumpulkan selanjutnya digunakan sebagai masukan dalam proses pengelompokan menggunakan algoritma K-Means.

Tahapan Penelitian



Gambar 1. Tahapan Penelitian

Algoritma K-Means

Algoritma **K-Means** merupakan metode *clustering* non-hierarki yang digunakan untuk mengelompokkan data berdasarkan tingkat kemiripan karakteristiknya. Pengelompokan dilakukan dengan menghitung jarak setiap data terhadap pusat klaster (*centroid*) menggunakan *Euclidean Distance*. Proses ini dilakukan secara berulang hingga posisi *centroid* stabil dan tidak terjadi perubahan keanggotaan klaster (Satyahadewi et al., 2013; Faizah et al., 2020).

Rumus *Euclidean Distance* yang digunakan adalah:

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

Tahapan algoritma K-Means dalam penelitian ini meliputi penentuan jumlah kluster menggunakan Elbow Method, penentuan *centroid* awal, perhitungan jarak data terhadap *centroid*, pembentukan kluster, pembaruan *centroid*, serta evaluasi hasil menggunakan Silhouette Score.

Evaluasi Hasil Kluster

Evaluasi hasil kluster dilakukan menggunakan Elbow Method dan Silhouette Score. Elbow Method digunakan untuk menentukan jumlah kluster (k) yang optimal berdasarkan nilai Within Cluster Sum of Squares (WCSS), di mana titik siku (elbow) menunjukkan jumlah kluster yang paling sesuai (Nurba et al., 2025). Selanjutnya, kualitas hasil kluster dievaluasi menggunakan Silhouette Score, yang mengukur tingkat kesamaan data dalam satu kluster dibandingkan dengan kluster lainnya. Nilai Silhouette Score berada pada rentang -1 hingga 1, di mana nilai yang semakin mendekati 1 menunjukkan kualitas kluster yang semakin baik (Fernaldy et al., 2025).

HASIL PENELITIAN

Statistik Deskriptif

Analisis statistik deskriptif dilakukan untuk memberikan gambaran awal mengenai karakteristik data penelitian yang terdiri atas Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), dan Angka Partisipasi Sekolah (APS) usia 16–18 tahun pada 38 provinsi di Indonesia tahun 2025. Hasil statistik deskriptif ditunjukkan pada Tabel 1

Tabel 1. Statistik Deskriptif Variabel Penelitian

Statistik	RLS	HLS	APS
Count	38	38	38
Mean	9,46	13,26	76,75
Std	1,19	1,03	8,95
Min	4,76	9,64	44,61
Max	11,58	15,78	93,03

Sumber: Hasil olahan Python (2025).

Berdasarkan Tabel 1, nilai rata-rata Rata-rata Lama Sekolah (RLS) sebesar 9,46 tahun, sedangkan Harapan Lama Sekolah (HLS) memiliki rata-rata 13,26 tahun. Sementara itu, Angka Partisipasi Sekolah (APS) usia 16–18 tahun memiliki rata-rata sebesar 76,75%. Perbedaan nilai minimum dan maksimum pada setiap indikator menunjukkan adanya variasi kondisi pendidikan antarprovinsi sehingga layak dilakukan proses klusterisasi menggunakan algoritma K-Means.

Normalisasi Data

Sebelum proses klasterisasi dilakukan, seluruh variabel dinormalisasi menggunakan metode *Z-Score* untuk menghilangkan perbedaan skala antarvariabel. Proses normalisasi bertujuan agar setiap indikator memiliki kontribusi yang seimbang dalam pembentukan klaster. Hasil normalisasi menunjukkan bahwa nilai rata-rata setiap variabel mendekati **0** dengan simpangan baku mendekati **1**, sehingga data telah memenuhi syarat untuk digunakan pada algoritma K-Means.

Tabel 2. Ringkasan Hasil Normalisasi Z-Score

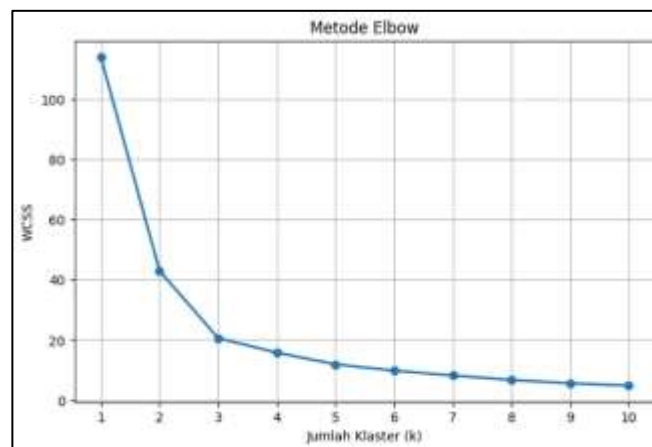
Statistik	RLS	HLS	APS
Mean	0,000	0,000	0,000
Std	1,013	1,013	1,013

Sumber: Hasil olahan Python (2025).

Nilai rata-rata yang mendekati nol dan simpangan baku mendekati satu menunjukkan bahwa proses normalisasi berhasil menghilangkan pengaruh perbedaan skala antarvariabel sehingga analisis klaster dapat dilakukan secara lebih objektif.

Penentuan Jumlah Klaster

Penentuan jumlah klaster dilakukan menggunakan metode *Elbow* dengan membandingkan nilai *Within Cluster Sum of Squares* (WCSS) pada jumlah klaster 1 sampai 10.



Gambar 2. Grafik Elbow Method

Berdasarkan Gambar 2, terlihat bahwa penurunan nilai WCSS sangat signifikan hingga $k = 3$, kemudian kurva mulai melandai pada jumlah klaster berikutnya. Oleh karena itu, jumlah klaster optimal yang digunakan pada penelitian ini adalah tiga klaster.

Hasil Klasterisasi Menggunakan Algoritma K-Means

Setelah jumlah klaster ditentukan, proses pengelompokan dilakukan menggunakan algoritma K-Means. Karakteristik masing-masing klaster ditunjukkan melalui nilai centroid sebagaimana disajikan pada

Pemetaan Kinerja Pendidikan Provinsi di Indonesia Menggunakan Algoritma K-Means Berdasarkan Indikator Pendidikan Tahun 2025

(Arif, et al.)

Tabel 3.

Tabel 3. Nilai Centroid Setiap Klaster

Cluster	RLS	HLS	APS	Kategori
Cluster 0	9,21	13,10	74,80	Sedang
Cluster 1	5,52	9,81	47,70	Rendah
Cluster 2	10,33	13,95	83,35	Tinggi

Sumber: Hasil olahan Python (2025).

Berdasarkan nilai centroid, Cluster 2 memiliki nilai rata-rata tertinggi pada seluruh indikator sehingga dikategorikan sebagai kelompok dengan kinerja pendidikan tinggi. Sebaliknya, Cluster 1 memiliki nilai rata-rata terendah sehingga termasuk kelompok kinerja pendidikan rendah, sedangkan Cluster 0 berada pada kategori kinerja pendidikan sedang.

Distribusi Hasil Klasterisasi

Distribusi jumlah provinsi pada masing-masing klaster disajikan pada Tabel 4.

Tabel 4. Distribusi Provinsi pada Setiap Klaster

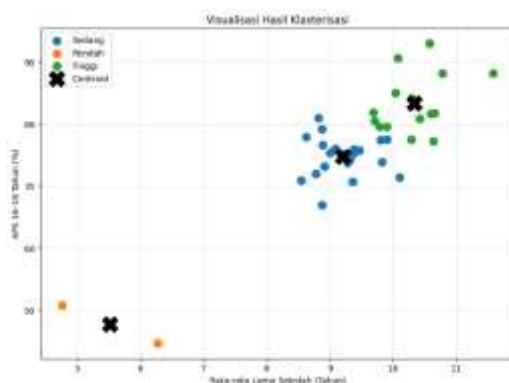
Cluster	Jumlah Provinsi	Persentase	Kategori
Cluster 0	21	55,26%	Sedang
Cluster 1	2	5,26%	Rendah
Cluster 2	15	39,47%	Tinggi

Sumber: Hasil olahan Python (2025).

Hasil klasterisasi menunjukkan bahwa sebagian besar provinsi berada pada Cluster 0 dengan jumlah 21 provinsi (55,26%). Selanjutnya, 15 provinsi (39,47%) termasuk dalam kelompok dengan kinerja pendidikan tinggi, sedangkan 2 provinsi (5,26%) berada pada kelompok kinerja pendidikan rendah.

Visualisasi Hasil Klasterisasi

Untuk memperjelas hasil pengelompokan, dilakukan visualisasi menggunakan grafik sebar (*scatter plot*) berdasarkan variabel RLS dan APS, serta posisi centroid pada setiap klaster.



Gambar 3. Hasil Klasterisasi Provinsi Berdasarkan Indikator Pendidikan Menggunakan Algoritma K-Means

Gambar 3 menunjukkan bahwa hasil klasterisasi membentuk tiga kelompok yang relatif terpisah. Cluster 2 berada pada wilayah dengan nilai RLS dan APS yang tinggi, sedangkan Cluster 1 berada pada wilayah dengan nilai terendah. Adapun Cluster 0 menempati posisi di antara kedua kelompok tersebut sehingga merepresentasikan provinsi dengan tingkat kinerja pendidikan sedang.

Evaluasi Hasil Klasterisasi

Evaluasi kualitas klaster dilakukan menggunakan *Silhouette Score*.

Tabel 5. Hasil Evaluasi Klaster Menggunakan Silhouette Score

Metode	Jumlah Klaster	Silhouette Score	Interpretasi
K-Means	3	0,453	Cukup Baik

Sumber: Hasil olahan Python (2025).

Nilai *Silhouette Score* sebesar **0,453** menunjukkan bahwa hasil klasterisasi memiliki kualitas yang **cukup baik**. Hal ini mengindikasikan bahwa sebagian besar provinsi berada pada kelompok yang sesuai dengan karakteristik indikator pendidikan yang dimilikinya, meskipun masih terdapat beberapa provinsi yang memiliki karakteristik yang relatif dekat dengan klaster lainnya.

pembahasan

Hasil penelitian menunjukkan bahwa algoritma K-Means mampu mengelompokkan 38 provinsi di Indonesia menjadi tiga kelompok berdasarkan indikator Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), dan Angka Partisipasi Sekolah (APS) usia 16–18 tahun. Pengelompokan ini menghasilkan tiga kategori, yaitu kinerja pendidikan tinggi, sedang, dan rendah, yang mencerminkan adanya perbedaan karakteristik pendidikan antarprovinsi.

Kelompok dengan kinerja pendidikan tinggi memiliki nilai rata-rata RLS sebesar 10,33 tahun, HLS sebesar 13,95 tahun, dan APS sebesar 83,35%. Kondisi ini menunjukkan bahwa provinsi dalam kelompok tersebut memiliki akses pendidikan yang lebih baik, harapan lama bersekolah yang lebih tinggi, serta

tingkat partisipasi sekolah yang relatif tinggi. Sebaliknya, kelompok kinerja pendidikan rendah memiliki nilai rata-rata RLS sebesar 5,52 tahun, HLS sebesar 9,81 tahun, dan APS sebesar 47,70%, yang mengindikasikan masih rendahnya akses dan partisipasi pendidikan. Oleh karena itu, wilayah dalam kelompok ini memerlukan perhatian yang lebih besar melalui pemerataan layanan pendidikan, peningkatan sarana dan prasarana, serta penguatan kualitas sumber daya pendidikan.

Sementara itu, sebagian besar provinsi berada pada kelompok kinerja pendidikan sedang dengan rata-rata RLS sebesar 9,21 tahun, HLS sebesar 13,10 tahun, dan APS sebesar 74,80%. Kelompok ini memiliki potensi untuk meningkat menjadi kelompok berkinerja tinggi apabila didukung oleh kebijakan yang mampu meningkatkan lama sekolah dan partisipasi pendidikan pada jenjang menengah. Selain itu, nilai Silhouette Score sebesar 0,453 menunjukkan bahwa hasil klusterisasi memiliki kualitas yang cukup baik sehingga mampu menggambarkan karakteristik pendidikan antarprovinsi secara representatif. Dengan demikian, hasil penelitian ini dapat menjadi dasar bagi pemerintah dalam menyusun kebijakan peningkatan mutu dan pemerataan pendidikan sesuai dengan karakteristik masing-masing kelompok provinsi.

SIMPULAN DAN SARAN

Penelitian ini berhasil memetakan kinerja pendidikan pada 38 provinsi di Indonesia berdasarkan indikator Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), dan Angka Partisipasi Sekolah (APS) usia 16–18 tahun. Hasil penelitian menunjukkan bahwa kondisi pendidikan antarprovinsi masih memiliki karakteristik yang beragam sehingga dapat dikelompokkan ke dalam tiga kategori, yaitu kinerja pendidikan tinggi, sedang, dan rendah. Kelompok dengan kinerja pendidikan tinggi memiliki capaian yang lebih baik pada seluruh indikator pendidikan, sedangkan kelompok dengan kinerja pendidikan rendah masih memerlukan perhatian lebih dalam peningkatan akses dan partisipasi pendidikan. Sementara itu, sebagian besar provinsi berada pada kategori kinerja pendidikan sedang yang menunjukkan adanya peluang untuk meningkatkan kualitas pendidikan melalui kebijakan yang tepat. Temuan penelitian ini memberikan gambaran mengenai karakteristik pendidikan antarprovinsi yang dapat dimanfaatkan sebagai dasar dalam mendukung perencanaan, pemerataan, dan peningkatan kualitas pendidikan di Indonesia. Berdasarkan hasil penelitian, disarankan agar penelitian selanjutnya menambahkan indikator pendidikan lain, seperti Angka Melek Huruf (AMH), Angka Partisipasi Kasar (APK), Angka Partisipasi Murni (APM), atau indikator mutu pendidikan lainnya sehingga hasil pemetaan menjadi lebih komprehensif. Selain itu, penggunaan algoritma klusterisasi lain, seperti Hierarchical Clustering, Fuzzy C-Means, atau DBSCAN, dapat dilakukan sebagai pembandingan untuk memperoleh hasil pengelompokan yang lebih optimal. Dari sisi praktis, hasil penelitian ini diharapkan dapat menjadi salah satu referensi bagi pemerintah pusat maupun pemerintah daerah dalam merumuskan kebijakan pemerataan dan peningkatan kualitas pendidikan sesuai dengan karakteristik masing-masing kelompok provinsi.

DAFTAR PUSTAKA

Alim, H. S., Rohmah, N., & Milawati, M. (2024). Study of education leverage factors to improve sampang human development index. *Cendikia: Media Jurnal Ilmiah Pendidikan*. <https://doi.org/10.35335/cendikia.v14i4.4624>

- Chong, B. (2021). K-means clustering algorithm: a brief review. *Academic Journal of Computing & Information Science*, 4(5). <https://doi.org/10.25236/AJCIS.2021.040506>
- Delima, A. S., Hasibuan, E., & Indonesia, D. (2025). *Kontribusi Pendidikan terhadap Kualitas Human Capital dan Dampaknya pada Indeks Pembangunan Manusia*. <https://doi.org/10.61132/jupenkei.v2i4.887>
- Faqih, M. F., & Mahdy, I. F. (2024). Penerapan Self Organizing Maps dalam Pengelompokan Provinsi di Indonesia Berdasarkan Aspek Pendidikan. *Jurnal Riset Statistika*, 93–102. <https://doi.org/10.29313/jrs.v4i2.4708>
- Fernaldy, F. A., Arifiyanti, A. A., & Kartika, D. S. Y. (2025). KLASTERISASI TRACER STUDY ALUMNI UNIVERSITAS XYZ MENGGUNAKAN ALGORITMA K-MEANS. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(1). <https://doi.org/10.23960/jitet.v13i1.5581>
- Fitria, D. A., Akbar, G. G., Ulumudin, A., Kania, I., & Yudiardi, D. (2025). Determinants of Average Years of Schooling in Garut Regency: An Exploratory Factor Analysis Approach. *Jurnal of Education Sciences*, 2(4), 344–365. <https://doi.org/10.62885/edusci.v2i4.655>
- Juita, D. P., Priya, Azwardi, M., & Amra, A. (2024). Pentingnya Pengembangan Sumber Daya Manusia pada Lembaga Pendidikan. *Indo-MathEdu Intellectuals Journal*, 5(3), 3068–3077. <https://doi.org/10.54373/imeij.v5i3.1243>
- Kamalia, A. Z., & Nawangsih, I. (2025). Identifikasi Pola Tingkat Kesenjangan Ketuntasan Pendidikan Di Indonesia Dengan Menggunakan Metode K-Medoids Clustering. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 12(2), 321–330. <https://doi.org/10.25126/jtiik.2025129219>
- Maryam, St. (2025). *Analysis of the Influence of Average School Length, Per Capita Expenditure, and Life Expectancy on the Human Development Index in West Nusa Tenggara Province in 2019-2023*. <https://doi.org/10.5281/zenodo.17438536>
- Maulidya, D., Asdi, Y., & Yozza, H. (2022). Pengelompokan provinsi di indonesia \ berdasarkan indikator pembangunan pendidikan menggunakan metode subtractive fuzzy c-means. *Jurnal Matematika Unand*, 11(3), 181. <https://doi.org/10.25077/jmua.11.3.181-189.2022>
- Mubarok, I., & Rita Wahyuni Arifin. (2025). Application of the K-Means Clustering Algorithm for Classifying High-Achieving Students at MI Irsyadul Athfal Depok. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, 5(1), 1773–1777. <https://doi.org/10.59934/jaiea.v5i1.1714>
- Nugraha, G. S. (2023). Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan Menggunakan Metode K-Means Clustering. *J-Cosine (Journal of Computer Science and Informatics Engineering)*, 7(1). <https://doi.org/10.29303/jcosine.v7i1.509>
- Nurba, A. S. W. A., Fathahillah, D., Pratama, M. S., Bachtiar, M. R., Putra, M. C. A., & Sitio, S. L. M. (2025). KLASTERISASI DATA : ANALISIS KINERJA K-MEANS PADA SEKTOR PAJAK, EKSPOR, PERIKANAN, MODAL, DAN SUMBER DAYA. *JRIS: JURNAL REKAYASA INFORMASI SWADHARMA*, 5(2), 108–120. <https://doi.org/10.56486/jris.vol5no2.883>
- Supriyadi, D., & Martini, N. (2023). Manajemen Sumber Daya Manusia dalam Lembaga Pendidikan. *Jurnal Manajemen Dan Sains*, 8(1), 450. <https://doi.org/10.33087/jmas.v8i1.776>
- Vidyananta, I. P. A., Dermawan, K. T., & Vidyananta, I. P. A. (2025). Perbandingan algoritma k-means dan k-medoids dalam pengelompokan provinsi di indonesia berdasarkan indikator keadaan sekolah dasar. *JITET: Jurnal Informatika Dan Teknik Elektro Terapan*, 13(3S1). <https://doi.org/10.23960/jitet.v13i3s1.8145>
- Wardani, M. T., Ramadhani, V. O., Anggreani, N., Sumanto, S., & Kuswanto, A. D. (2025). Pengelompokan Provinsi Berdasarkan Kelayakan Ruang Kelas dan Tenaga Kependidikan Sekolah Dasar Menggunakan Algoritma K-Means: Analisa Data Periode 2023-2024. *Riau Jurnal Teknik Informatika.*, 4(2), 163–172. <https://doi.org/10.30606/rjti.v4i2.3389>