



Research Paper

JHP: Jurnal Harmoni Pendidikan

Vol 2, No 1. (2026), 57- 69

e-ISSN: 3109-8991

Doi: 10.64845/jhp.v1i2

Journal homepage: <https://athallahpublishing.com/index.php/jhp/index>

Determinan Pendidikan di Indonesia Menggunakan Metode *K-Means Clustering*

Mochammad Luthfan Nur Rafif Falah^{1*}, Novia Amilatus Solekha²

¹Universitas Terbuka, Indonesia

²STMIK PPKIA Tarakan, Indonesia

*Corresponding author: luthfan.nur.rafif@gmail.com

ARTICLE INFO	ABSTRAK
<p>Kata Kunci Analisis Klaster Faktor Pendidikan K-Means Clustering</p> <p>Article history Received: 12 August 2025 Revised: 25 October 2025 Accepted: 25 December 2025 Available online: 06 Januari 2026</p>	<p>Pendidikan merupakan faktor utama dalam pembangunan sumber daya manusia yang berkualitas dan unggul. Penelitian ini bertujuan untuk mengelompokkan faktor-faktor yang mempengaruhi tingkat pendidikan di Indonesia pada tahun 2024 menggunakan metode <i>K-Means Clustering</i>. Data yang digunakan adalah data sekunder dari Badan Pusat Statistik (BPS) yang mencakup lima variabel utama: Angka Melek Aksara, Angka Partisipasi Sekolah (APS), Angka Buta Aksara, Angka Partisipasi Kasar (APK), dan Angka Partisipasi Murni (APM). Metode <i>K-Means Clustering</i> diterapkan untuk mengelompokkan 38 provinsi di Indonesia ke dalam tiga klaster berdasarkan karakteristik pendidikan masing-masing. Hasil penelitian menunjukkan bahwa klaster pertama terdiri dari provinsi dengan tingkat pendidikan rendah, klaster kedua dengan tingkat pendidikan sedang, dan klaster ketiga dengan tingkat pendidikan tinggi. Uji ANOVA menunjukkan bahwa variabel Angka Melek Aksara memiliki pengaruh paling signifikan dalam pembentukan klaster.</p>

Copyright © 2026 Authors

This is an open access article under [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/) license



Pendahuluan

Pendidikan berperan penting dalam pembangunan Sumber Daya Manusia (SDM) yang tangguh, produktif, dan berdaya saing, didukung oleh industri dan talenta global. Hal ini sejalan dengan Rencana Pembangunan Jangka Menengah Nasional 2020-2024 untuk meningkatkan kualitas SDM (BPS, 2024). Pendidikan merupakan indikator utama kemajuan bangsa yang mencerminkan kualitas sumber daya manusia dan daya saing nasional. Sistem Pendidikan Nasional berperan penting dalam menciptakan masyarakat maju, didukung oleh SDM berkualitas dan fasilitas yang memadai. Pendidikan, baik formal, nonformal, maupun informal, berfungsi membentuk individu bermartabat dan mengembangkan pola pikir konstruktif sepanjang hayat (Via Septiani, 2023). Pendidikan bertujuan mendorong dan memfasilitasi kegiatan belajar siswa agar mereka dapat mengembangkan potensi diri. Sesuai dengan UUD RI No. 20 Tahun 2003, pendidikan berperan dalam membentuk watak serta peradaban bangsa yang bermartabat. Belajar menjadi aspek penting dalam pendidikan, karena melalui proses ini, individu mengalami perubahan dan perkembangan (Auliya et al., 2024; Salsabila, 2020). Berdasarkan definisi pendidikan yaitu tidak hanya mentransfer pengetahuan dan keterampilan, tetapi juga membentuk nilai dan karakter siswa, dimana Pendidikan bukan hanya menekankan pada proses membaca namun juga proses memahami serta menganalisis materi pembelajaran (Kasa et al., 2022; Lestari, 2019; Safitri et al., 2024).

Clustering atau klasterisasi adalah teknik untuk mengelompokkan data dalam analisis statistik. Metode ini digunakan dalam berbagai bidang seperti ekonomi, psikologi, kesehatan, dan sosial. Analisis Klaster adalah teknik multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Analisis Klaster mengklasifikasi objek sehingga setiap objek yang hampir memiliki kesamaan dengan objek lain akan di posisikan dalam klaster yang sama. Klaster-klaster yang terbentuk memiliki homogenitas internal yang tinggi dan heterogenitas eksternal yang tinggi (Abid & Setiawan, 2023; Cantika et al., 2024; Nur Afidah, 2023). Salah satu metode clustering yang terkenal adalah K-Means, yaitu algoritma non-hirarki yang membagi objek ke dalam beberapa cluster berdasarkan karakteristiknya. K-Means bekerja melalui proses iteratif hingga mencapai kondisi akhir yang optimal. Metode K-means merupakan suatu metode yang dapat melakukan pengelompokan data dalam jumlah yang cukup besar dengan perhitungan waktu yang relatif cepat dan efisien. (Aulia, 2021; Fakhri et al., 2021; Nur Afidah, 2023; Tri Cahaya et al., 2024).

Berbagai penelitian tentang penggunaan metode clustering telah dilakukan yaitu penelitian yang dilakukan oleh Azzam Al Fauzie & Akhir Putra (2023) menunjukkan hasil pengelompokan data ke dalam delapan klaster dengan jumlah siswa yang bervariasi, menggunakan tiga kategori ukuran klaster, yaitu pintar, sedang, dan cukup. Berdasarkan analisis data, ditemukan bahwa Bahasa Inggris

perlu mendapatkan perhatian khusus dalam pembelajaran. Siswa dengan nilai seimbang antara mata pelajaran IPA dan IPS dapat memilih jurusan sesuai minat mereka. Siswa dalam klaster sedang cenderung memiliki nilai yang dominan pada jurusan IPA, sedangkan siswa dalam klaster cukup lebih dominan pada jurusan IPS. Penelitian selanjutnya oleh Asroni & Adrian, (2016) menunjukkan bahwa algoritma K-Means bisa digunakan untuk mengelompokkan mahasiswa berdasarkan IPK dan beberapa atribut mata kuliah. Penelitian selanjutnya oleh Arifah et al (2023) yaitu Pengujian aplikasi segmentasi citra jajanan tradisional menggunakan Elbow Visualizer untuk menentukan jumlah klaster optimal dan waktu pemrosesan. Dari 10 pengujian, algoritma K-Means memperoleh akurasi 76,47%, sementara Fuzzy C-Means mencapai 68.63%. Hasil ini menunjukkan bahwa K-Means lebih unggul dalam akurasi dan lebih efisien dalam waktu pemrosesan dibandingkan Fuzzy C-Means dalam segmentasi warna citra jajanan tradisional.

Pada Penelitian ini membahas penerapan metode clustering dengan algoritma K-Means dalam mengelompokkan faktor-faktor yang mempengaruhi tingkat pendidikan di Indonesia berdasarkan klasifikasi Provinsi ke dalam klaster yang telah ditetapkan. K-Means dipilih karena menggunakan data fisik yang jelas dan tidak abstrak, sesuai dengan karakteristik penduduk di Indonesia. Selain itu, metode ini fleksibel dalam menentukan jumlah klaster yang akan dibuat. Tujuan penelitian ini adalah untuk menganalisis penerapan dan hasil clustering dalam pengelompokan faktor-faktor yang mempengaruhi tingkat pendidikan di tiap Provinsi di Indonesia

Metode

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining, khususnya teknik clustering menggunakan algoritma K-Means. Pendekatan kuantitatif dipilih karena penelitian bertujuan untuk mengidentifikasi pola dan pengelompokan determinan pendidikan di Indonesia berdasarkan data numerik yang bersumber dari indikator-indikator pendidikan. Metode K-Means Clustering digunakan untuk mengelompokkan wilayah atau satuan analisis ke dalam beberapa klaster yang memiliki karakteristik determinan pendidikan yang relatif homogen. Data yang digunakan adalah data sekunder yaitu data yang diperoleh secara tidak langsung untuk mendapatkan informasi (keterangan) dari objek yang diteliti, yang biasanya telah dikumpulkan oleh lembaga pengumpul data. Data sekunder yang digunakan berasal dari suatu badan (instansi) yang dengan sengaja mengumpulkan data yaitu Badan Pusat Statistik (BPS), y terdiri atas 5 variabel pada tahun 2024 di Indonesia (Sumber: Publikasi Statistik Pendidikan)

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari lembaga resmi, seperti Badan Pusat Statistik (BPS), Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, serta sumber data nasional lainnya yang relevan. Data mencakup berbagai indikator determinan pendidikan di Indonesia, antara lain angka partisipasi sekolah, rata-rata lama sekolah, harapan lama sekolah, rasio guru terhadap murid, tingkat melek huruf, tingkat kemiskinan, serta ketersediaan sarana dan prasarana pendidikan. Satuan analisis dalam penelitian ini adalah wilayah administratif (provinsi/kabupaten/kota) sesuai dengan ketersediaan data.

Tabel 1. Variabel faktor yang Mempengaruhi Pendidikan di Indonesia

Variabel	Keterangan
X_1	Angka Melek Aksara Penduduk 15-24 tahun Menurut Provinsi
X_2	Angka Partisipasi Sekolah (APS) 16-18 tahun Menurut Provinsi
X_3	Angka Buta Aksara Penduduk 10 tahun ke Atas Menurut Provinsi
X_4	Angka Partisipasi Kasar (APK) 16-18 tahun Menurut Provinsi
X_5	Angka Partisipasi Murni (APM) 16-18 tahun Menurut Provinsi

Sumber: Badan Pusat Statistika, 2024

Metode yang digunakan dalam pengolahan adalah K-Means Clustering dengan bantuan perangkat lunak SPSS. Tahapan dalam metode ini adalah sebagai berikut:

1. Menentukan jumlah kluster (k-cluster). Pada penelitian ini, jumlah kluster yang digunakan adalah tiga.
2. Melakukan pengelompokan menggunakan K-Means Clustering
3. Mendapatkan anggota dari masing-masing cluster berdasarkan hasil pengelompokan
4. Melakukan profilisasi setiap kluster dan menginterpretasikannya dalam bentuk pemetaan visualisasi. Dalam penelitian ini, ketiga kluster yang diperoleh diklasifikasikan sebagai kluster rendah, kluster sedang, dan kluster tinggi.

Hasil penelitian disajikan dalam bentuk tabel, grafik, dan visualisasi peta kluster untuk memudahkan pemahaman terhadap pola determinan pendidikan di Indonesia. Visualisasi ini diharapkan dapat memberikan gambaran yang komprehensif mengenai distribusi kluster pendidikan antarwilayah serta menjadi dasar dalam perumusan kebijakan pendidikan yang lebih tepat sasaran.

Hasil dan Pembahasan

Dilakukan analisis pada data indikator faktor-faktor yang mempengaruhi pendidikan di Indonesia tahun 2024 dengan metode K-Means Cluster dengan

langkah awal yaitu menentukan jumlah cluster yang akan dibentuk adalah 3 cluster. Standardisasi data dengan transformasi ke bentuk Z-Score dilakukan terlebih dahulu karena terdapat perbedaan ukuran satuan yang besar pada data antar variabel agar tidak menyebabkan bias pada analisis cluster. Selanjutnya baru dilakukan analisis pada data yang telah di standardisasi menggunakan K-Means Cluster. Hasil analisis menggunakan metode K-Means Clustering menunjukkan bahwa determinan pendidikan di Indonesia dapat dikelompokkan ke dalam tiga klaster utama, yaitu klaster rendah, klaster sedang, dan klaster tinggi. Pengelompokan ini didasarkan pada kesamaan karakteristik provinsi terhadap indikator-indikator pendidikan yang meliputi Angka Melek Aksara, Angka Buta Aksara, Angka Partisipasi Sekolah, Angka Partisipasi Kasar, dan Angka Partisipasi Murni. Pembentukan klaster ini memberikan gambaran mengenai pola ketimpangan dan perbedaan kondisi pendidikan antarwilayah di Indonesia.

Hasil deskripsi klaster menunjukkan bahwa klaster rendah didominasi oleh provinsi-provinsi dengan capaian rendah pada sebagian besar indikator pendidikan, khususnya pada aspek literasi dasar. Tingginya Angka Buta Aksara pada klaster ini mengindikasikan masih terbatasnya akses dan kualitas pendidikan dasar di wilayah tersebut. Meskipun demikian, nilai sedang pada Angka Partisipasi Sekolah dan Angka Partisipasi Murni menunjukkan adanya upaya peningkatan akses pendidikan, namun belum diikuti dengan kualitas hasil pendidikan yang memadai. Klaster sedang memperlihatkan kondisi pendidikan yang relatif lebih baik dibandingkan klaster rendah, terutama ditandai oleh tingginya Angka Melek Aksara. Hal ini menunjukkan bahwa sebagian besar penduduk di wilayah ini telah memiliki kemampuan literasi dasar yang baik. Namun, tingkat partisipasi pendidikan yang masih berada pada kategori sedang mengindikasikan adanya tantangan dalam mempertahankan keberlanjutan pendidikan pada jenjang yang lebih tinggi. Dengan demikian, fokus kebijakan pendidikan pada klaster ini perlu diarahkan pada peningkatan partisipasi dan pemerataan akses pendidikan.

Sementara itu, klaster tinggi mencerminkan wilayah dengan kondisi pendidikan yang relatif maju. Rendahnya Angka Buta Aksara dan tingginya Angka Partisipasi Sekolah menunjukkan bahwa akses dan partisipasi pendidikan di wilayah ini telah berjalan dengan baik. Meskipun beberapa indikator masih berada pada kategori sedang, klaster ini secara umum menggambarkan daerah dengan sistem pendidikan yang lebih stabil dan berkelanjutan. Untuk memastikan perbedaan karakteristik antar klaster, dilakukan uji ANOVA terhadap seluruh variabel yang digunakan. Hasil uji ANOVA menunjukkan bahwa semua variabel memiliki nilai signifikansi $\text{Sig.} < 0,05$, yang berarti terdapat perbedaan yang signifikan secara statistik antar klaster. Hal ini menegaskan bahwa pengelompokan yang dihasilkan oleh metode K-Means Clustering bersifat valid dan mampu membedakan karakteristik determinan pendidikan antarwilayah secara nyata.

Tabel 2. Output Descriptive Statistics

	<i>Min</i>	<i>Max</i>	<i>Mean</i>	<i>Std. Deviation</i>
x1	90	100	99,32	2,128
x2	48	90	75,07	7,525
x3	0	27	3,45	4,889
x4	54	99	88,22	8,938
x5	35	81	65,47	9,458

Sumber: www.bps.go.id (diolah)

Pada Tabel 2 dapat dilihat nilai deskriptif statistik pada masing-masing indikator yaitu Angka Melek Aksara (X_1), Angka Partisipasi Sekolah (APS) (X_2), Angka Buta Aksara (X_3), Angka Partisipasi Kasar (APK) (X_4), dan Angka Partisipasi Murni (APM) (X_5) pada 38 Provinsi di Indonesia yang memuat rata-rata, standar deviasi, nilai maksimum dan minimum.

Tabel 3. Output Initial Cluster Center

	<i>Cluster</i>		
	1	2	3
Zscore(x1)	-4,108	0,286	0,300
Zscore(x2)	-2,522	-0,367	1,394
Zscore(x3)	4,913	-0,399	-0,538
Zscore(x4)	-2,679	-0,677	0,840
Zscore(x5)	-2,182	-0,638	1,649

Pada Tabel 3 dapat diketahui tampilan awal proses clustering yaitu hasil sementara proses pengelompokan data sebelum dilakukan proses iterasi yang menunjukkan nilai pusat atau *centroid cluster* awal dari data yang sebelumnya telah dilakukan standarisasi.

Tabel 4. Output Iteration History

<i>Iteration</i>	<i>Change in Cluster Centers</i>		
	1	2	3
1	1,578	0,734	1,008
2	0,000	0,000	0,000

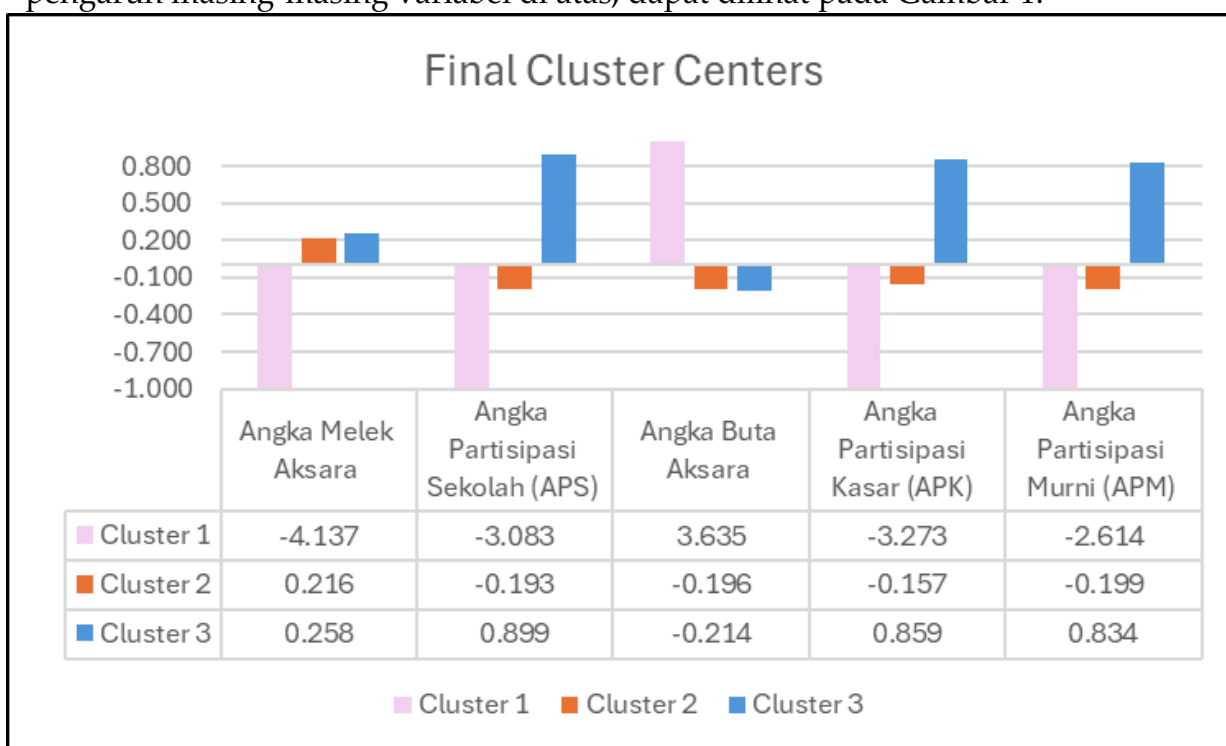
Pada Tabel 4 dapat diketahui bahwa proses iterasi dilakukan sebanyak 2 kali. Proses ini dilakukan untuk mendapatkan kluster yang tepat. Selanjutnya dapat diketahui bahwa jarak minimum antar pusat kluster yang terjadi dari hasil iterasi adalah 3,264.

Tabel 5. Output Final Cluster Centers

	Cluster		
	1	2	3
Zscore (x1)	-4,137	0,216	0,258
Zscore (x2)	-3,083	-0,193	0,899
Zscore (x3)	3,635	-0,196	-0,214
Zscore (x4)	-3,273	-0,157	0,859
Zscore (x5)	-2,614	-0,199	0,834

Sumber: www.bps.go.id (diolah)

Tabel 5 menyajikan hasil dari proses akhir dalam *clustering*, yang menghasilkan tiga klaster untuk masing-masing variabel. Variabel pada tabel *Final Cluster Centers* menunjukkan nilai yang telah di standarisasi. Angka negatif mengindikasikan bahwa data berada di bawah rata-rata total, sedangkan angka positif menunjukkan bahwa data berada di atas rata-rata total. Untuk memahami pengaruh masing-masing variabel di atas, dapat dilihat pada Gambar 1.



Gambar 1. Grafik Final Cluster Centers

Sumber: www.bps.go.id (diolah)

Berdasarkan Tabel 5 dan Gambar 1, diperoleh karakteristik masing-masing klaster yang terbentuk dari hasil analisis K-Means Clustering. Setiap klaster menunjukkan pola yang berbeda terkait determinan pendidikan di Indonesia berdasarkan variabel yang dianalisis. Klaster 1 (Klaster Rendah) terdiri atas

provinsi-provinsi yang memiliki nilai rendah pada hampir seluruh variabel pendidikan yang diteliti. Kondisi ini menunjukkan bahwa capaian pendidikan di wilayah yang termasuk dalam klaster ini relatif tertinggal dibandingkan klaster lainnya. Namun demikian, terdapat pengecualian pada beberapa variabel. Pada variabel Angka Partisipasi Murni (APM) (X_5) dan Angka Partisipasi Sekolah (APS) (X_2), klaster ini menunjukkan nilai pada kategori sedang. Sementara itu, klaster 1 memiliki nilai yang relatif tinggi pada variabel Angka Buta Aksara, yang mengindikasikan masih tingginya proporsi penduduk yang belum mampu membaca dan menulis.

Klaster 2 (Klaster Sedang) mencakup provinsi-provinsi dengan karakteristik nilai rendah pada sebagian besar variabel pendidikan, kecuali pada variabel Angka Melek Aksara (X_1). Pada klaster ini, nilai Angka Partisipasi Kasar (APK) (X_4) dan Angka Partisipasi Sekolah (APS) (X_2) berada pada kategori sedang, yang menunjukkan bahwa akses terhadap pendidikan sudah mulai terbuka, meskipun belum optimal. Klaster 2 memiliki nilai tinggi pada variabel Angka Melek Aksara (X_1), yang menandakan bahwa kemampuan literasi dasar penduduk relatif baik. Dengan demikian, tantangan utama pada klaster ini bukan lagi pada aspek literasi, melainkan pada peningkatan partisipasi dan keberlanjutan pendidikan.

Klaster 3 (Klaster Tinggi) terdiri atas provinsi-provinsi dengan kondisi pendidikan yang relatif lebih baik dibandingkan klaster lainnya. Klaster ini memiliki nilai rendah pada variabel Angka Buta Aksara, yang menunjukkan tingkat literasi masyarakat yang tinggi. Nilai sedang ditunjukkan pada variabel Angka Melek Aksara (X_1), Angka Partisipasi Kasar (APK) (X_4), dan Angka Partisipasi Murni (APM) (X_5). Sementara itu, nilai tinggi pada klaster ini terdapat pada variabel Angka Partisipasi Sekolah (APS) (X_2), yang mengindikasikan tingginya keterlibatan penduduk usia sekolah dalam sistem pendidikan formal. Hal ini mencerminkan kondisi pendidikan yang relatif maju dengan tingkat partisipasi yang baik.

Berdasarkan uraian karakteristik yang diperoleh dari Tabel dan Grafik *Final Cluster Centers*, dapat disimpulkan bahwa masing-masing klaster merepresentasikan determinan pendidikan yang dominan di Indonesia. Klaster 1 mencerminkan wilayah dengan permasalahan utama pada tingginya angka buta aksara, klaster 2 merepresentasikan wilayah yang ditandai oleh tingginya angka melek aksara, dan klaster 3 menggambarkan wilayah dengan tingginya angka partisipasi sekolah. Hasil pengelompokan ini memberikan gambaran yang komprehensif mengenai kesenjangan pendidikan antarwilayah di Indonesia dan dapat menjadi dasar bagi perumusan kebijakan pendidikan yang lebih terarah dan berbasis karakteristik wilayah. Sehingga berdasarkan rincian yang diperoleh dari Tabel dan Grafik *Final Cluster Centers* tersebut dapat dikatakan bahwa klaster 1 merupakan pengelompokan faktor-faktor yang mempengaruhi pendidikan di

Indonesia berdasarkan angka buta aksara, kluster 2 berdasarkan angka melek aksara dan kluster 3 berdasarkan angka partisipasi sekolah.

Tabel 6. Uji Anova

	<i>Cluster</i>		<i>Error</i>		F	Sig.
	<i>Mean Square</i>	df	<i>Mean Square</i>	df		
Zscore (x1)	18,070	2	0,025	35	734,626	0,000
Zscore (x2)	14,807	2	0,211	35	70,167	0,000
Zscore (x)	13,945	2	0,260	35	53,579	0,000
Zscore(x4)	15,429	2	0,176	35	87,913	0,000
Zscore (x5)	11,481	2	0,401	35	28,624	0,000

Sumber: www.bps.go.id (diolah)

Berdasarkan hasil analisis yang diperoleh, seluruh variabel memiliki nilai signifikansi (Sig. < 0,05), yang menunjukkan bahwa ketiga kluster yang terbentuk memiliki perbedaan yang signifikan secara statistik. Hasil uji ANOVA juga menunjukkan bahwa variabel Angka Melek Aksara (X_1) merupakan variabel yang paling membedakan anggota dari ketiga kluster, karena memiliki nilai F terbesar dibandingkan dengan variabel determinan pendidikan lainnya. Temuan ini mengindikasikan bahwa tingkat literasi penduduk menjadi faktor dominan dalam pengelompokan kondisi pendidikan di Indonesia.

Tabel 7. Cluster Membership

<i>Case Number</i>	Provinsi	<i>Cluster</i>	<i>Distance</i>
1	ACEH	3	0,243
2	SUMATERA UTARA	3	0,615
3	SUMATERA BARAT	3	0,641
4	RIAU	2	1,064
5	JAMBI	2	0,362
6	SUMATERA SELATAN	2	0,649
7	BENGKULU	3	0,451
8	LAMPUNG	2	0,376
9	KEP. BANGKA BELITUNG	2	0,462
10	KEP. RIAU	3	1,008
11	DKI JAKARTA	2	1,328
12	JAWA BARAT	2	0,844

<i>Case Number</i>	<i>Provinsi</i>	<i>Cluster</i>	<i>Distance</i>
13	JAWA TENGAH	2	0,633
14	DI YOGYAKARTA	3	1,340
15	JAWA TIMUR	2	0,753
16	BANTEN	2	0,890
17	BALI	3	0,582
18	NUSA TENGGARA BARAT	3	1,514
19	NUSA TENGGARA TIMUR	2	0,624
20	KALIMANTAN BARAT	2	0,883
21	KALIMANTAN TENGAH	2	0,828
22	KALIMANTAN SELATAN	2	0,734
23	KALIMANTAN TIMUR	3	0,464
24	KALIMANTAN UTARA	3	0,513
25	SULAWESI UTARA	2	0,626
26	SULAWESI TENGAH	2	0,773
27	SULAWESI SELATAN	2	0,758
28	SULAWESI TENGGARA	2	0,853
29	GORONTALO	2	0,436
30	SULAWESI BARAT	2	0,496
31	MALUKU	3	0,750
32	MALUKU UTARA	2	0,805
33	PAPUA BARAT	2	0,763
34	PAPUA BARAT DAYA	3	0,713
35	PAPUA	2	1,100
36	PAPUA SELATAN	2	3,268
37	PAPUA TENGAH	1	1,578
38	PAPUA PEGUNUNGAN	1	1,578

Tabel 7 menyajikan hasil akhir pengelompokan provinsi di Indonesia berdasarkan analisis K-Means Clustering terhadap indikator determinan pendidikan. Tabel ini menampilkan informasi mengenai nomor kasus, nama provinsi, klaster hasil pengelompokan, serta nilai jarak (*distance*) masing-masing provinsi terhadap pusat klaster (*centroid*). Nilai *distance* menunjukkan tingkat kedekatan karakteristik suatu provinsi dengan klaster tempat provinsi tersebut tergabung; semakin kecil nilai *distance*, semakin dekat karakteristik provinsi tersebut dengan pusat klaster. Berdasarkan hasil pengelompokan, provinsi-provinsi di Indonesia terbagi ke dalam tiga klaster, yaitu klaster rendah (klaster 1), klaster sedang (klaster 2), dan klaster tinggi (klaster 3).

Secara keseluruhan, hasil Cluster Membership pada Tabel 7 menunjukkan adanya variasi kondisi pendidikan antarprovinsi di Indonesia. Pengelompokan ini memberikan gambaran yang jelas mengenai pola determinan pendidikan berdasarkan tingkat literasi dan partisipasi pendidikan. Klaster rendah

mencerminkan wilayah dengan tantangan pendidikan yang kompleks, klaster sedang menggambarkan wilayah dengan capaian pendidikan yang cukup tetapi belum optimal, dan klaster tinggi menunjukkan wilayah dengan kondisi pendidikan yang relatif lebih baik. Dengan demikian, hasil pemetaan anggota klaster ini dapat dijadikan sebagai dasar dalam penyusunan kebijakan pendidikan yang lebih spesifik, terarah, dan berbasis wilayah, sehingga intervensi yang dilakukan dapat disesuaikan dengan karakteristik dan kebutuhan masing-masing klaster.



Gambar 2. Peta Clustering

Sumber: www.bps.go.id (diolah)

Kesimpulan

Pengelompokan faktor-faktor yang memengaruhi pendidikan di Indonesia pada tahun 2024 menggunakan metode K-Means menghasilkan tiga klaster. Klaster 1 (tingkat pendidikan rendah) terdiri dari Papua Tengah dan Papua Pegunungan. Klaster 2 (tingkat pendidikan sedang) mencakup Riau, Jambi, Sumatera Selatan, Lampung, Kepulauan Bangka Belitung, 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, Maluku Utara, Papua Barat, Papua, dan Papua Selatan. Klaster 3 (tingkat pendidikan tinggi) mencakup Aceh, Sumatera Utara, Sumatera Barat, Bengkulu, Kepulauan Riau, Yogyakarta, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Utara, Maluku, dan Papua Barat Daya. Setiap klaster memiliki karakteristiknya masing-masing yang dapat digunakan oleh pemerintah daerah maupun pusat sebagai acuan untuk mengoptimalkan faktor-faktor yang memengaruhi pendidikan di Indonesia.

References

- Abid, A., & Setiawan, R. P. (2023). Pemanfaatan Metode Clustering untuk Menganalisa Penduduk Kebumen yang Memiliki Keterampilan Teknologi Informasi dan Komunikasi (TIK). *Journal of Data Science Theory and Application*, 2(2), 36–41.
- Arifah, S., Swedia, E. R., & Septian, M. R. D. (2023). Analisis Perbandingan Algoritma Clustering Dalam Melakukan Segmentasi Warna Pada Citra Jajan Tradisional. *Sebatik*, 27(1), 70–76. <https://doi.org/10.46984/sebatik.v27i1.2273>
- Asroni, A., & Adrian, R. (2016). Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang. *Semesta Teknika*, 18(1), 76–82. <https://doi.org/10.18196/st.v18i1.708>
- Aulia, S. (2021). Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja). *Djtechno: Jurnal Teknologi Informasi*, 1(1), 1–5. <https://doi.org/10.46576/djtechno.v1i1.964>
- Auliya, D., Islam, A., & Marlina, E. (2024). Analisis Diskriminan Terhadap Faktor-Faktor Yang Mempengaruhi Prestasi Belajar Siswa Dalam Mata Pelajaran Matematika. *Jurnal Riset Matematika Dan Sains Terapan*, 4, 30–36.
- Azzam Al Fauzie, M., & Akhir Putra, J. (2023). Clustering Data Menggunakan Metode K-Means untuk Rekomendasikan Pembelajaran Akademik bagi Siswa Aktif dalam Ekstrakurikuler. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 4(1),

- 642–648. <https://doi.org/10.30865/klik.v4i1.1116>
- BPS. (2024). Statistik Pendidikan 2024. Jakarta: Badan Pusat Statistik.
- Cantika, R., Fauzi, A., & Sihombing, A. (2024). Penerapan Metode Clustering Untuk Mengetahui Kepatuhan Wajib Pajak Bumi Dan Bangunan Pada Desa Perkebunan Tanjung Keliling kemampuannya dalam memajukan kesejahteraan masyarakat . Untuk mencapai tujuan yang dan / atau perairan pedalaman dan / atau laut . A. *Jurnal Publikasi Sistem Informasi Dan Telekomunikasi*, 2(4).
- Fakhri, D. A., Defit, S., & Sumijan. (2021). Optimalisasi Pelayanan Perpustakaan terhadap Minat Baca Menggunakan Metode K-Means Clustering. *Jurnal Informasi Dan Teknologi*, 3, 160–166. <https://doi.org/10.37034/jidt.v3i3.137>
- Kasa, M. Y. S., Abolladaka, J., & Br Simanungkalit, E. F. (2022). Faktor-Faktor Yang Mempengaruhi Tingkat Pendidikan Anak Di Kelurahan Fatukbot Kecamatan Atambua Selatan Kabupaten Belu. *Journal Economic Education, Business and Accounting*, 1(2), 79–86. <https://doi.org/10.35508/jeeba.v1i2.8699>
- Lestari, W. (2019). Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi (Studi Kasus: STMIK Bina Bangsa Kendari). *Simkom*, 4(2), 35–48. <https://doi.org/10.51717/simkom.v4i2.37>
- Nur Afidah, N. (2023). Penerapan Metode Clustering dengan Algoritma K-means untuk Pengelompokkan Data Migrasi Penduduk Tiap Kecamatan di Kabupaten Rembang. *PRISMA, Prosiding Seminar Nasional Matematika*, 6, 729–738. <https://journal.unnes.ac.id/sju/index.php/prisma/>
- Safitri, W., Nopianti, H., & Widiono, S. (2024). Analisis Perbandingan Tren Angka Partisipasi Sekolah Berdasarkan Jenis Kelamin ; Laki-laki dan Perempuan di Kota Bengkulu Periode (2014-2023). *Jurnal Manajemen Pendidikan Dan Ilmu Sosial*, 5(6), 2651–2659. <https://creativecommons.org/licenses/by/4.0/%0AAalisis>
- Salsabila, A. (2020). Faktor-Faktor Yang Mempengaruhi Prestasi Belajar Siswa Sekolah Dasar. *Jurnal Pendidikan Dan Dakwah*, 2(2), 278–288.
- Tri Cahaya, D., Puspita, D., & Syahri, R. (2024). Penerapan Metode K-Means Clustering Untuk Pengelompokan Potensi Padi Di Kota Pagar Alam. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 2187–2193. <https://doi.org/10.36040/jati.v8i2.9432>
- Via Septiani. (2023). Types And Levels Of Education In The National Education System In Indonesia. *ATTAQWA: Jurnal Pendidikan Islam Dan Anak Usia Dini*, 2(3), 118–125. <https://doi.org/10.58355/attaqwa.v2i3.4>