

Identification Of Disparities In Educational Facilities Among Indonesian Provinces Using K-Means Clustering

Ananda Desilia Putri¹, Harni Selasih Dewi²,
Edy Widodo^{3*}

^{1,2,3} Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia, Jl. Kaliurang km 14.5 Yogyakarta 55584, Indonesia

¹23611034@students.uii.ac.id, ²23611027@students.uii.ac.id, ³edywidodo@uii.ac.id

*Corresponding author

ABSTRAK

Pendidikan merupakan salah satu indikator dalam mengukur kemajuan suatu negara, karena peran penting pendidikan dalam menghasilkan sumber daya manusia yang berkualitas. Pendidikan yang baik tentunya didukung oleh ketersediaan fasilitas pendidikan yang memadai. Namun, geografi wilayah di Indonesia yang beragam menjadi tantangan dalam pemerataan akses pendidikan. Provinsi yang masuk kategori wilayah 3T harus menghadapi kekurangan fasilitas pendidikan dan tenaga pengajar. Studi ini bertujuan mengelompokkan provinsi-provinsi di Indonesia berdasarkan indikator fasilitas pendidikan untuk periode tahun ajaran 2023/2024 dengan menggunakan *K-Means Clustering Analysis*. Data yang digunakan mencakup 38 provinsi dengan 18 indikator fasilitas pendidikan, yang dianalisis setelah melalui pra-pemrosesan data. Dari hasil penelitian ini diperoleh 3 kluster, di mana kluster pertama terdiri dari 1 provinsi dengan akses dan kondisi infrastruktur kurang baik, kluster kedua terdiri dari 17 provinsi dengan akses dan kondisi infrastruktur cukup baik, dan kluster ketiga terdiri dari 20 provinsi dengan akses dan kondisi infrastruktur sangat baik. Hasil klusterisasi dari penelitian ini diharapkan dapat menjadi rujukan bagi perumusan kebijakan pemerataan fasilitas pendidikan serta penentuan prioritas pembangunan pada sektor pendidikan di Indonesia.

Kata Kunci:

Indikator Fasilitas Pendidikan, *Principal Component Analysis*, Analisis *K-Means Clustering*

ABSTRACT

Education is one of the indicators used to measure a country's progress, due to its important role in producing quality human resources. Good education is certainly supported by the availability of adequate educational facilities. However, Indonesia's diverse geography poses a challenge in terms of equal access to education. Provinces classified as 3T regions face a shortage of educational facilities and teaching staff. This study aims to group provinces in Indonesia based on educational facility indicators for the 2023/2024 academic year using *K-Means Clustering Analysis*. The data used covers 38 provinces with 18 educational facility indicators, which were analyzed after data pre-processing. The results of this study obtained 3 clusters, where the first cluster consisted of 1 province with poor access and infrastructure conditions, the second cluster consisted of 17 provinces with fairly good access and infrastructure conditions, and the third cluster consisted of 20 provinces with very good access and infrastructure conditions. The clustering results from this study are expected to serve as a reference for the formulation of policies on the equitable distribution of educational facilities and the determination of development priorities in the education sector in Indonesia.

Keywords:

Educational Facility Indicators, *Principal Component Analysis*, *K-Means Clustering Analysis*

1. PENDAHULUAN

Pendidikan merupakan salah satu tolak ukur maju atau tidaknya suatu negara, karena peran penting dan keberfungsian dalam menghasilkan sumber daya manusia yang berkualitas, baik dalam aspek kemampuan kognitif, spiritual, dan keterampilan [1]. Salah satu indikator Indeks Pembangunan Manusia (IPM) adalah pendidikan. Hal ini memperkuat pentingnya kontribusi pendidikan dalam mengukur kemajuan suatu negara, di mana kemajuan yang

dimaksud direpresentasikan oleh tingkat kesejahteraan hidup penduduknya [2]. Pendidikan yang baik tentunya didukung oleh ketersediaan fasilitas pendidikan sebagai tubuh fisik yang merepresentasikan komitmen suatu negara terhadap masa depan generasinya. Fasilitas pendidikan menjadi indikator penting dalam mengidentifikasi kesenjangan atau kemajuan akses pendidikan di suatu wilayah [3].

Ketimpangan akses pendidikan di Indonesia antara wilayah maju dan tertinggal masih menjadi permasalahan dalam meningkatkan kualitas pendidikan secara inklusif [4]. Indonesia sendiri memiliki banyak wilayah terpencil yang sulit dijangkau dengan kondisi geografi yang menantang seperti pegunungan dan pulau-pulau kecil [5]. Kondisi tersebut menjadi salah satu faktor yang mengakibatkan keterbatasan sarana dan prasarana pendidikan, terutama daerah yang jauh dari pusat kota. Provinsi yang masuk dalam kategori wilayah 3T memiliki tingkat pendidikan yang lebih rendah karena kurangnya fasilitas pendidikan dan tenaga pengajar dibanding wilayah non-3T [6].

Sejumlah penelitian terdahulu umumnya menitikberatkan tolak ukur pada capaian pendidikan, seperti Angka Partisipasi Sekolah (APS), Angka Partisipasi Murni (APM), serta indikator capaian pendidikan lainnya [7]-[12]. Meskipun indikator-indikator tersebut penting dalam menggambarkan partisipasi dan keberhasilan pendidikan, penelitian yang secara khusus memetakan ketimpangan fasilitas pendidikan masih relatif terbatas. Ketersediaan fasilitas pendidikan merupakan aspek struktural yang memiliki peran penting dalam memenuhi akses pendidikan yang merata.

Penelitian ini menggunakan pendekatan klasterisasi dalam mengidentifikasi pola ketimpangan fasilitas pendidikan provinsi-provinsi di Indonesia. *K-Means clustering* memungkinkan klasterisasi provinsi berdasarkan kemiripan karakteristik yang tinggi, sehingga pola ketimpangan dapat diidentifikasi secara sistematis. Metode ini dipilih karena memiliki kemampuan dalam membentuk kelompok atau klaster yang tidak tumpang tindih, hal ini menunjukkan kemudahan dalam menginterpretasikan hasil klasterisasi [13]. Penelitian dilakukan pada tingkat provinsi karena level ini relevan dalam membantu perumusan dan memberikan masukan untuk evaluasi kebijakan pendidikan berskala makro di Indonesia.

2. METODE

2.1 DATA DAN VARIABEL PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan memanfaatkan data sekunder yang diperoleh dari situs resmi Portal Pendidikan Indonesia dan Badan Pusat Statistik (BPS). Data yang digunakan mencakup 40 pengamatan yang merepresentasikan kondisi fasilitas pendidikan pada setiap jenjang pendidikan formal, yaitu Sekolah Dasar (SD), Sekolah Menengah Pertama (SMP), dan Sekolah Menengah Atas (SMA) di seluruh Provinsi Indonesia.

Populasi dalam penelitian ini adalah seluruh provinsi di Indonesia dengan sampel penelitiannya adalah masing-masing provinsi di Indonesia. Data diteliti agar dapat mengidentifikasi dan menggambarkan ketimpangan fasilitas pendidikan antar jenjang pendidikan.

Variabel penelitian merupakan indikator fasilitas pendidikan yang mencerminkan ketersediaan serta kualitas sarana dan prasarana pendidikan pada masing-masing jenjang sekolah. Variabel-variabel tersebut meliputi rasio jumlah sekolah per 10000 penduduk usia sekolah, ketersediaan fasilitas dasar (toilet), kondisi ruang kelas, serta ketersediaan fasilitas penunjang pembelajaran (laboratorium IPA, perpustakaan, dan akses internet). Definisi operasional masing-masing variabel disajikan pada **Tabel 1**.

Tabel 1. Definisi Operasional Variabel

Kode Variabel	Nama Variabel	Definisi Operasional	Satuan
X1	RSD	Jumlah SD per 10000 penduduk usia sekolah	Per 10000 penduduk usia sekolah
X2	RSMP	Jumlah SMP per 10000 penduduk usia sekolah	Per 10000 penduduk usia sekolah
X3	RSMA	Jumlah SMA per 10000 penduduk usia sekolah	Per 10000 penduduk usia sekolah
X4	SD_Memiliki_Toilet	Persentase SD yang memiliki fasilitas toilet	Persen (%)
X5	SMP_Memiliki_Toilet	Persentase SMP yang memiliki fasilitas toilet	Persen (%)
X6	SMA_Memiliki Toilet	Persentase SMA yang memiliki fasilitas toilet	Persen (%)
X7	SD_Ruang Kelas_Rusak Berat	Persentase ruang kelas SD dengan kondisi rusak berat	Persen (%)
X8	SMP_Ruang Kelas_Rusak Berat	Persentase ruang kelas SMP dengan kondisi rusak berat	Persen (%)
X9	SMA_Ruang Kelas_Rusak Berat	Persentase ruang kelas SMA dengan kondisi rusak berat	Persen (%)
X10	SD_Lab IPA	Persentase SD yang memiliki laboratorium IPA	Persen (%)
X11	SMP_Lab IPA	Persentase SMP yang memiliki laboratorium IPA	Persen (%)
X12	SMA_Lab IPA	Persentase SMA yang memiliki laboratorium IPA	Persen (%)
X13	SD_Perpustakaan	Persentase SD yang memiliki perpustakaan	Persen (%)
X14	SMP_Perpustakaan	Persentase SMP yang memiliki perpustakaan	Persen (%)
X15	SMA_Perpustakaan	Persentase SMA yang memiliki perpustakaan	Persen (%)
X16	SD_Internet	Persentase SD yang memiliki akses internet untuk tujuan pembelajaran	Persen (%)
X17	SMP_Internet	Persentase SMP yang memiliki akses internet untuk tujuan pembelajaran	Persen (%)
X18	SMA_Internet	Persentase SMA yang memiliki akses internet untuk tujuan pembelajaran	Persen (%)

2.2 METODE ANALISIS STATISTIK

Metode analisis yang digunakan dalam penelitian ini adalah K-Means *clustering*. K-Means merupakan algoritma non hirarki yang berasal dari metode data klastering, yang bertujuan untuk mengelompokkan data ke dalam sejumlah klaster berdasarkan tingkat kemiripan tertentu. Metode ini digunakan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan karakteristik fasilitas pendidikan yang dimiliki.

Proses K-Means *clustering* diawali dengan pembentukan partisi klaster di awal, yang kemudian dilakukan proses iteratif untuk memperbaiki pembagian klaster hingga tidak terjadi perubahan yang signifikan hingga tercapai kondisi yang konvergen [14]. K-Means *clustering* dapat membagi data berdasarkan jarak antar data pada pusat klaster (*centroid*) yang telah ditetapkan. Algoritma ini bergantung pada fungsi untuk mengukur data yang memiliki ciri khas sama. Jarak itu sendiri dihitung menggunakan fungsi *Euclidean*. Proses ini terus diulang hingga posisi *centroid* stabil. Kemudian data dimasukkan dalam kelompok yang memiliki jarak terdekat [15].

2.2.1 PRA-PEMROSESAN DATA

Tahap pra-pemrosesan data dilakukan untuk memastikan kualitas data sebelum analisis lebih lanjut. Pada tahap ini dilakukan pemeriksaan *missing value* serta pengecekan *outlier* menggunakan jarak *Euclidean* untuk mengidentifikasi pengamatan dengan nilai ekstrem yang berpotensi memengaruhi hasil analisis. Berikut adalah rumus untuk perhitungan jarak *Euclidean*.

$$d_{(x_j, c_j)} = \sqrt{\sum_{j=1}^n (x_j - c_j)^2} \quad (2.1)$$

dengan $d_{(x_j, c_j)}$ merupakan jarak antara objek x dan pusat kluster (*centroid*), x_j menyatakan nilai variabel ke- j pada objek x , c_j merupakan nilai variabel ke- j pada *centroid*, n adalah jumlah variabel yang digunakan dalam analisis, dan j menunjukkan indeks data yang dianalisis.

Selanjutnya, data distandarisasi menggunakan metode *Z-Score* guna mengatasi perbedaan skala antar variabel sehingga setiap variabel memiliki kontribusi yang seimbang dalam proses pengelompokan. Berikut adalah rumus *Z-Score*.

$$Z_i = \frac{x_i - \bar{x}}{s} \quad (2.2)$$

dengan Z_i menyatakan nilai hasil standarisasi untuk observasi ke- i , x_i merupakan nilai asli variabel pada observasi ke- i , $\bar{x}$ menyatakan nilai rata-rata variabel, dan s merupakan standar deviasi variabel.

2.2.2 ANALISIS KORELASI

Analisis korelasi dilakukan untuk mengetahui hubungan antar variabel indikator fasilitas pendidikan. Tujuan dari analisis ini adalah untuk mengidentifikasi adanya hubungan yang kuat atau multikolinearitas antar variabel, yang dapat memengaruhi hasil pengelompokan. Hasil analisis korelasi digunakan sebagai dasar pertimbangan dalam penerapan metode reduksi dimensi.

2.2.3 PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal Component Analysis (PCA) diterapkan sebagai metode reduksi dimensi untuk mengatasi permasalahan multikolinearitas antar variabel. Kelayakan penerapan PCA diuji menggunakan *Bartlett's Test of Sphericity* dan *Kaiser-Meyer-Olkin* (KMO). Komponen utama dipilih berdasarkan kriteria nilai eigenvalue lebih dari satu serta pola penurunan pada *scree plot*. Skor komponen utama yang terpilih selanjutnya digunakan sebagai *input* dalam analisis *klustering*. Berikut adalah rumus menghitung nilai KMO.

$$KMO = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} p_{ij}^2} \quad (2.3)$$

dengan r_{ij} adalah koefisien sederhana antara variabel ke- i dan ke- j , sementara p_{ij} adalah koefisien parsial antara variabel ke- i dan ke- j .

2.2.4 PENENTUAN JUMLAH KLASSTER

Penentuan jumlah kluster optimal dilakukan untuk memperoleh hasil pengelompokan yang paling representatif. Metode yang digunakan adalah *Elbow Method*, dengan menganalisis nilai *Within Sum of Squares* (WSS) pada berbagai jumlah kluster. Titik siku (*elbow point*) pada grafik WSS digunakan sebagai acuan dalam menentukan jumlah kluster yang optimal.

2.2.5 ANALISIS K-MEANS CLUSTERING

Analisis *K-Means Clustering* dilakukan dengan menggunakan jumlah kluster optimal yang telah ditentukan sebelumnya. Proses pengelompokan dilakukan berdasarkan skor komponen utama hasil PCA. Algoritma K-Means bekerja secara iteratif dengan memperbarui pusat kluster (*centroid*) hingga tidak terjadi perubahan signifikan pada keanggotaan kluster atau hingga tercapai kondisi konvergen.

2.2.6 VALIDASI KLASSTER

Validasi kluster dilakukan untuk mengevaluasi kualitas hasil pengelompokan. Metode validasi yang digunakan adalah koefisien Silhouette, yang mengukur tingkat kesamaan objek dalam satu kluster dibandingkan dengan kluster lainnya. Nilai koefisien Silhouette yang lebih tinggi menunjukkan hasil pengelompokan yang lebih baik, dengan tingkat homogenitas dalam kluster dan pemisahan antar kluster yang optimal.

3. HASIL DAN PEMBAHASAN

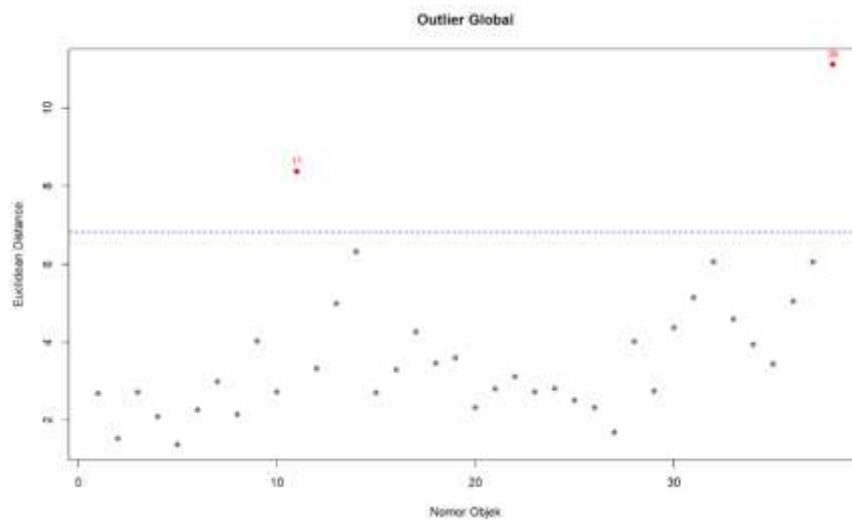
3.1 STATISTIK DESKRIPTIF

Data yang digunakan dalam analisis ini adalah data fasilitas pendidikan pada 38 Provinsi di Indonesia pada tahun ajaran 2023/2024. Proyeksi dari data tersebut dapat dilihat pada **Tabel 2**.

	Tabel 2. Analisis Deskriptif		
	Mean	Minimum	Maksimum
X1	65.82	23.16	72.28
X2	41.26	20.82	72.28
X3	14.71	5.12	30.45
X4	75.56	29.32	96.91
X5	83.10	48.22	96.92
X6	87.16	53.23	97.17
X7	12.31	1.98	22.68
X8	8.62	0.27	22.48
X9	6.34	0.12	14.72
X10	7.11	0.99	61.56
X11	78.13	43.15	105.73
X12	50.94	17.74	78.47
X13	76.99	31.13	100.96
X14	91.84	59.39	104.21
X15	97.86	77.42	121.39
X16	75.73	32.99	96.70
X17	67.21	15.79	97.26
X18	83.25	43.55	105.41

3.2 PRA-PEMROSESAN DATA

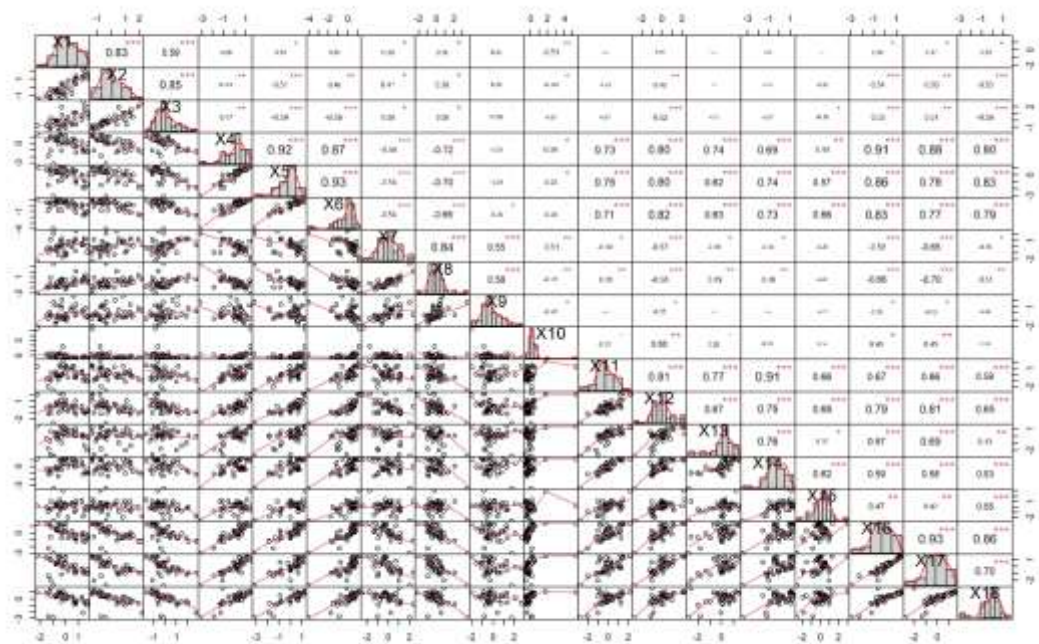
Dilakukan standarisasi data dengan metode *Z-Score* untuk menangani skala yang berbeda dari masing-masing variabel. Pengecekan *outlier* global menggunakan jarak Euclidean yang menunjukkan adanya beberapa *outlier* pada provinsi dengan jarak ekstrem, namun tetap dipertahankan dalam analisis karena merepresentasikan keadaan wilayah secara nyata. Hasil dari pemeriksaan *outlier* dapat dilihat pada **Gambar 1**.



Gambar 1. Pengecekan Data *Outlier*

3.3 UJI MULTIKOLINEARITAS

Pengujian multikolinearitas dengan melihat nilai koefisien korelasi, hasil menunjukkan bahwa terdapat multikolinearitas seperti pada **Gambar 2**.



Gambar 2. Korelasi Antar Variabel

Berdasarkan **Gambar 2**, terdapat beberapa variabel yang memiliki nilai koefisien korelasi lebih dari 0.8. Hal ini mengindikasikan terdapat multikolinearitas tinggi pada variabel independen, sehingga diperlukan langkah untuk mengatasi multikolinearitas dengan melakukan *Principal Component Analysis*.

3.4 PRINCIPAL COMPONENT ANALYSIS

PCA dilakukan karena terdapat multikolinearitas pada variabel independen. Pemeriksaan dan pengujian analisis komponen utama melalui pemeriksaan korelasi antar variabel dengan

Bartlett's Test of Sphericity dan kecukupan data dengan uji *Kaiser-Meyer-Olkin* (KMO). Hasil dari pemeriksaan korelasi antar variabel tersebut disajikan pada **Tabel 3**.

Tabel 3. Uji *Bartlett* dan Nilai KMO

KMO dan Uji <i>Bartlett</i>	
<i>Bartlett's Test of Sphericity</i> (Sig.)	< 0.001
Ukuran <i>Kaiser Meyer Olkin</i>	0.79

Dari **Tabel 3**, didapatkan hasil uji *Bartlett's Test of Sphericity* yang menunjukkan nilai signifikansi < 0.001, mengindikasikan bahwa matriks korelasi bukan matriks identitas. Uji KMO menunjukkan nilai *overall* MSA sebesar 0.79 yang di mana lebih besar dari 0.5, sehingga data dikatakan cukup untuk dilakukan analisis lebih lanjut yaitu PCA. Selanjutnya, hasil pengujian *Measure of Sampling Adequacy* (MSA) untuk masing-masing variabel disajikan pada **Tabel 4**.

Tabel 4. Nilai MSA

Variabel	Nilai MSA
X1	0.59
X2	0.64
X3	0.69
X4	0.92
X5	0.88
X6	0.85
X7	0.75
X8	0.75
X9	0.77
X10	0.81
X11	0.77
X12	0.94
X13	0.73
X14	0.88
X15	0.67
X16	0.78
X17	0.80
X18	0.82

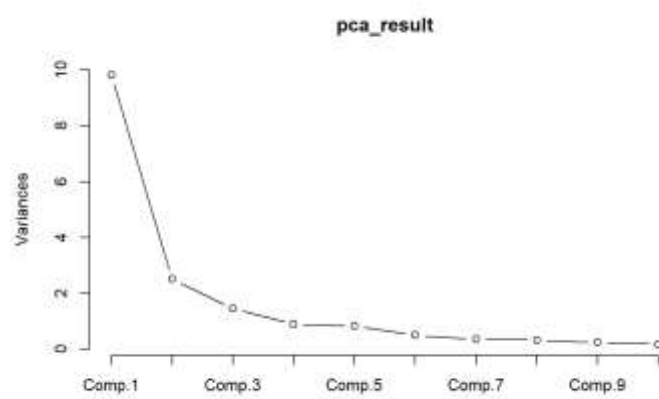
Dari **Tabel 4**, diperoleh nilai MSA dari masing-masing variabel adalah lebih dari 0.5 sehingga seluruh variabel dikatakan memiliki kelayakan yang cukup untuk diikutkan dalam analisis PCA. Pada proses analisis PCA, hasil menunjukkan terbentuk 18 komponen dengan keterangan seperti pada **Tabel 5**.

Tabel 5. Hasil Ekstraksi PCA

Komponen	Nilai Eigen	Persentase Varians (%)	Akumulasi Varians (%)
Comp.1	9.82	0.56	0.56
Comp.2	2.51	0.14	0.70
Comp.3	1.45	0.08	0.79
Comp.4	0.88	0.05	0.84
Comp.5	0.82	0.05	0.88
Comp.6	0.49	0.03	0.91
Comp.7	0.35	0.02	0.93
Comp.8	0.32	0.02	0.95
Comp.9	0.23	0.01	0.96
Comp.10	0.18	0.01	0.97
Comp.11	0.13	0.01	0.98
Comp.12	0.09	0.01	0.99
Comp.13	0.07	0.00	0.99

Komponen	Nilai Eigen	Persentase Varians (%)	Akumulasi Varians (%)
Comp.14	0.06	0.00	0.99
Comp.15	0.05	0.00	1.00
Comp.16	0.03	0.00	1.00
Comp.17	0.02	0.00	1.00
Comp.18	0.01	0.00	1.00

Berdasarkan **Tabel 5**, terdapat 3 komponen utama yang memiliki nilai eigen lebih dari 1. Sesuai dengan ketentuan dalam PCA, komponen yang dianggap signifikan adalah komponen yang memiliki nilai eigen lebih dari 1. Oleh karena itu, didapatkan 3 komponen utama yang terbentuk dalam analisis ini. Ketiga komponen tersebut mampu menjelaskan sebesar 79% total keragaman dari seluruh variabel independen. Hal ini diperkuat dengan tampilan grafik *scree plot* pada **Gambar 3** yang menunjukkan titik belok (*elbow*) setelah komponen ketiga.

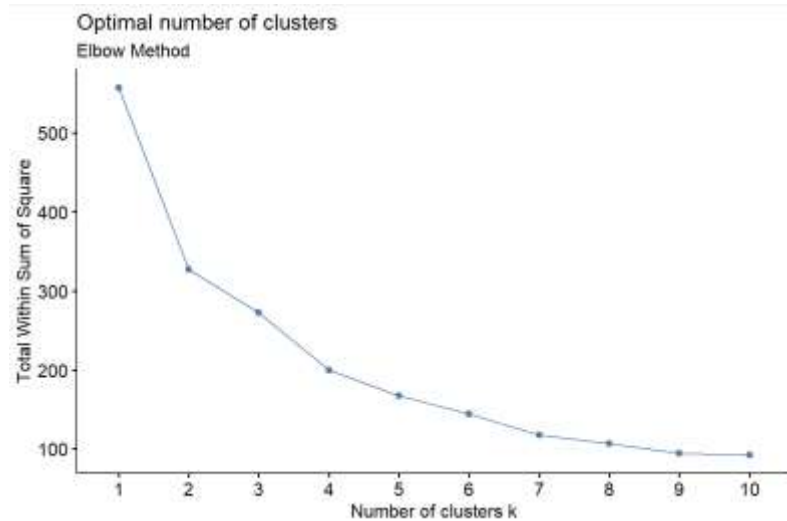


Gambar 3. *Scree Plot*

Pada **Gambar 3**, *Scree Plot* menunjukkan nilai eigen dari 18 komponen yang terbentuk. Hasil dari grafik ini, terlihat bahwa terdapat 3 komponen yang menurun tajam dengan nilai eigen lebih besar dari 1. Berdasarkan hasil analisis, maka ditentukan banyaknya komponen utama dengan komponen pertama sebagai faktor utama atau yang paling berpengaruh dibandingkan dengan komponen yang lain.

3.5 ANALISIS *K-MEANS CLUSTERING*

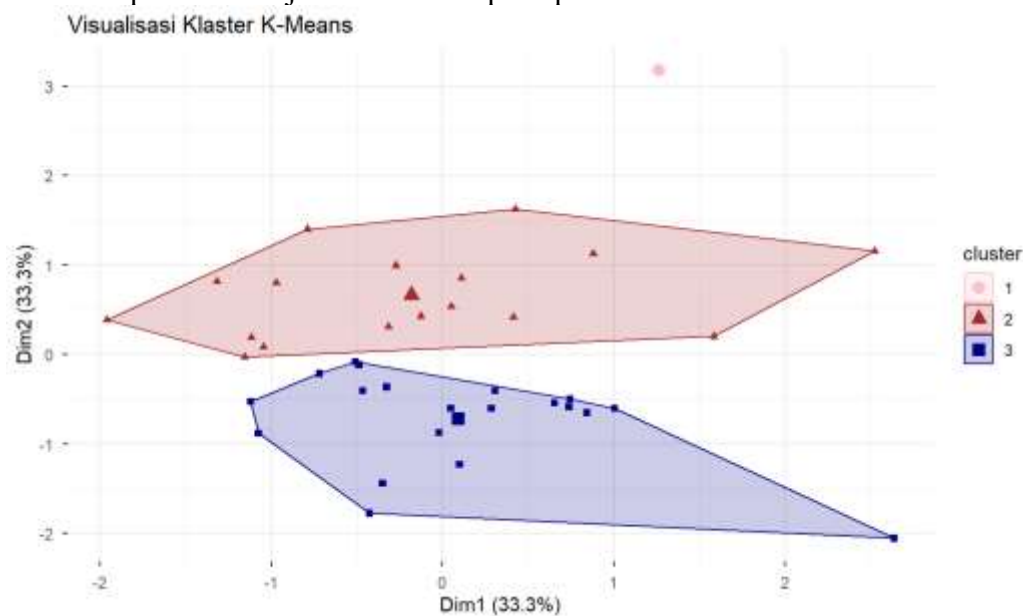
Metode yang digunakan dalam penelitian ini adalah *K-Means clustering* untuk mengelompokkan Provinsi di Indonesia berdasarkan indikator fasilitas pendidikan. Data yang digunakan adalah data indikator fasilitas pendidikan pada jenjang SD hingga SMA seluruh Provinsi di Indonesia tahun ajaran 2023/2024. Dalam menentukan banyaknya nilai kelompok (*k*) yang optimal, digunakan metode *elbow* atau *Within Sum of Square* (WSS) seperti pada **Gambar 4**.



Gambar 4. Penentuan Jumlah Kluster (Elbow)

Berdasarkan **Gambar 4.**, metode *lbow* dengan nilai k yang memiliki titik belok tajam adalah pada $k = 3$. Oleh karena itu, dipilih nilai $k = 3$ sebagai jumlah kluster optimal untuk mengelompokkan masing-masing provinsi berdasarkan indikator fasilitas pendidikan.

Klasterisasi provinsi dilakukan berdasarkan tiga komponen utama hasil analisis PCA yang kemudian dikelompokkan menjadi 3 kluster seperti pada **Gambar 5**.



Gambar 5. Hasil Klasterisasi Dengan *K-Means*

Dari **Gambar 5**, terdapat kluster yang hanya beranggotakan 1 provinsi. Hal ini bukan mengindikasikan *outlier* pada data cukup ekstrem sehingga mempengaruhi hasil klasterisasi, tetapi karena nilai tersebut menggambarkan kondisi secara nyata pada provinsi tersebut. Dari hasil klasterisasi, didapatkan 3 kluster dengan masing-masing anggota kluster seperti pada **Tabel 5**.

Tabel 5. Hasil Kluster Provinsi-provinsi Indonesia

Kluster	Provinsi	Jumlah Anggota
1	Papua Pegunungan	1
2	Aceh, NTB, NTT, Kalimantan Barat, Kalimantan Tengah, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Barat,	17

Klaster	Provinsi	Jumlah Anggota
3	Maluku, Maluku Utara, Papua Barat, Papua Barat Daya, Papua, Papua Selatan, Papua Tengah Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Kalimantan Selatan, Kalimantan Timur, Sulawesi Selatan, Gorontalo	20

Seperti pada **Tabel 5**, didapatkan jumlah anggota paling banyak ada pada klaster 3 dan klaster yang hanya memiliki satu anggota yaitu ada pada klaster 1. Setelah didapatkan hasil klasterisasi, kemudian dilanjutkan profilisasi klaster menggunakan nilai rata-rata variabel independen di masing-masing klaster seperti pada **Tabel 6**.

Tabel 6. Profilisasi Klaster

Variabel	Klaster 1	Klaster 2	Klaster 3
X ₁	47.77711	77.24970	56.99969
X ₂	35.35027	51.90626	32.50985
X ₃	14.05513	18.76392	11.30408
X ₄	29.32000	66.81176	85.31550
X ₅	48.22000	76.36647	90.56950
X ₆	53.23000	82.81059	92.55750
X ₇	21.23000	14.32412	10.15550
X ₈	22.48000	10.05765	6.69700
X ₉	11.780000	7.074706	5.434000
X ₁₀	7.670000	3.274118	10.333500
X ₁₁	43.15000	72.61882	84.56550
X ₁₂	17.74000	43.83941	58.64050
X ₁₃	31.13000	71.06118	84.32600
X ₁₄	59.3900	89.0000	95.8715
X ₁₅	77.42000	95.74294	100.69100
X ₁₆	32.99492	66.36294	85.83084
X ₁₇	15.78947	54.24571	80.80164
X ₁₈	43.54839	77.12120	90.44778

Berdasarkan hasil profilisasi pada **Tabel 6**, klaster 1 dicirikan oleh nilai yang relatif lebih rendah pada sebagian besar indikator dibandingkan dengan klaster lain. Persentase sekolah yang memiliki toilet, laboratorium IPA, perpustakaan, dan akses internet di semua jenjang pendidikan adalah yang paling rendah dari klaster lainnya. Klaster ini juga menunjukkan tingkat rasio satuan pendidikan per 10000 penduduk usia sekolah yang relatif rendah, mengindikasikan akses fisik terhadap layanan pendidikan masih terbatas. Selain itu, persentase ruang kelas dengan kondisi rusak berat relatif tinggi di semua jenjang pendidikan pada klaster ini. Secara keseluruhan, klaster ini merepresentasikan provinsi dengan kondisi akses dan infrastruktur pendidikan yang kurang baik dan perlu ditingkatkan.

Pada klaster 2, rasio satuan pendidikan per 10000 penduduk usia sekolah berada pada tingkat sedang/menengah yang menunjukkan bahwa akses pendidikan cukup memadai. Ketersediaan sarana dan prasarana pendidikan relatif berada pada kategori sedang, di mana proporsi sekolah yang memiliki toilet, laboratorium IPA, perpustakaan, dan akses internet lebih baik dibandingkan dengan klaster 1 meskipun tidak lebih tinggi dari klaster 3. Persentase ruang kelas dengan kondisi rusak berat relatif lebih rendah dibandingkan dengan klaster 3, menandakan kondisi fisik bangunan sekolah yang cukup layak. Secara

keseluruhan, klaster ini merepresentasikan provinsi dengan tingkat pembangunan infrastruktur pendidikan yang cukup baik tetapi masih perlu peningkatan agar lebih baik.

Pada klaster 3 dicirikan dengan rasio satuan pendidikan per 10000 penduduk usia sekolah yang paling tinggi pada semua jenjang pendidikan dibandingkan dengan klaster lain, menunjukkan akses pendidikan yang luas dan merata. Ketersediaan sarana dan prasarana pada klaster ini berada pada tingkat yang sangat tinggi. Sebaliknya, persentase ruang kelas dengan kondisi fisik bangunan yang rusak relatif rendah dibandingkan dengan klaster lain. Secara keseluruhan, klaster ini merepresentasikan provinsi dengan akses dan kualitas infrastruktur pendidikan yang sangat baik dibandingkan provinsi pada klaster lain.

3.6 VALIDASI KLASSTER

Validasi klaster yang telah terbentuk menggunakan koefisien Silhouette untuk menilai kualitas klasterisasi berdasarkan jarak antar klaster. Hasil analisis menunjukkan bahwa nilai Silhouette rata-rata berada pada kategori sedang dan didominasi dengan nilai positif, mengindikasikan bahwa sebagian besar provinsi telah dikelompokkan secara tepat sesuai karakteristik infrastruktur dan indikator fasilitas pendidikan lainnya. Klaster dengan jumlah anggota terbanyak menunjukkan tingkat homogenitas antar objek dalam klaster yang lebih baik dibandingkan dengan *klaster* lainnya, sedangkan klaster dengan jumlah anggota terbanyak kedua memperlihatkan pemisahan yang cukup memadai. Klaster yang hanya terdiri dari 1 anggota memiliki nilai Silhouette mendekati nol, yang menunjukkan adanya provinsi dengan karakteristik yang sangat berbeda dan berdiri sendiri.

4. KESIMPULAN

Berdasarkan hasil pengelompokan provinsi menggunakan metode *K-Means clustering*, pemerintah pusat dan pemerintah daerah disarankan untuk memfokuskan intervensi kebijakan pendidikan pada provinsi yang tergolong dalam klaster dengan kondisi fasilitas pendidikan rendah. Intervensi tersebut dapat diarahkan pada peningkatan rasio satuan pendidikan terhadap jumlah penduduk usia sekolah, pemenuhan fasilitas dasar dan penunjang yang masih terbatas, serta rehabilitasi ruang kelas dengan tingkat kerusakan tinggi pada jenjang SD, SMP, dan SMA.

Selain itu, komponen utama hasil PCA yang paling dominan dalam menjelaskan keragaman data dapat digunakan sebagai indikator prioritas dalam perencanaan dan evaluasi pembangunan fasilitas pendidikan antarprovinsi. Dengan demikian, alokasi anggaran dan program peningkatan infrastruktur pendidikan dapat difokuskan pada variabel-variabel yang secara statistik memiliki kontribusi terbesar terhadap ketimpangan fasilitas pendidikan.

Untuk pengembangan penelitian selanjutnya, disarankan untuk melakukan analisis lanjutan terhadap dinamika perubahan klaster antar tahun guna mengevaluasi efektivitas kebijakan pendidikan, serta membandingkan hasil analisis metode *K-Means* dengan metode klasterisasi lain seperti *Hierarchical Clustering* atau DBSCAN untuk menguji kestabilan struktur klaster yang terbentuk.

5. REFERENSI

- [1] P. Ananda *et al.*, Revolusi Pendidikan Indonesia: Mencetak Generasi Cerdas di Era Digital, 1st ed. Bandung, Indonesia: Uwais Inspirasi Indonesia, 2025. [Online]. Available: www.penerbituwais.com.
- [2] A. Rahman, M. Syafii, dan S. H. Hakim, Kesejahteraan dan Indikator Kunci Pembangunan, hlm. 113–118. Bandung, Indonesia: CV Merdeka Kreasi Group, 2021.

- [3] Sumarni *et al.*, *Komparasi Sistem Pendidikan Indonesia dan Negara Maju Benua Asia, Eropa dan Amerika*, hlm. 37–38. Yogyakarta, Indonesia: CV Idebuku, 2025.
- [4] Y. S. Wahyuni and Y. Miaz, *Pendidikan Global Multikultural dan Resolusi Konflik*, Padang, Indonesia: UMSB Press, 2024.
- [5] J. A. Banks, *Diversity, Transformative Knowledge, and Civic Education: Selected Essays*, New York, NY, USA: Routledge, 2020. doi: 10.4324/9781003018360. :contentReference[oaicite:0]{index=0}
- [6] N. Feriyanto, *Ekonomi Sumber Daya Manusia: Teori dan Evaluasi Kebijakan Pengentasan Kemiskinan dan Pengangguran di Indonesia*, hlm. 128–141. Yogyakarta, Indonesia: Deepublish Digital, 2025.
- [7] N. R. Handitia dan A. Sofro, “Analisis Klaster Kecamatan di Kota Surabaya Berdasarkan Data Pendidikan Tahun 2022–2023,” *Jurnal Gaussian*, vol. 13, no. 2, pp. 351–362, 2024, doi: 10.14710/j.gauss.13.2.351-362.
- [8] Y. A. Priambodo dan Y. J. Prasetyo, “Pemetaan Penyebaran Guru di Provinsi Banten dengan Menggunakan Metode Spatial Clustering K-Means,” *Indonesian Journal of Computing and Modeling*, vol. 5, 2018.
- [9] S. Hanifah dan A. H. Primandari, “Implementasi Metode K-Means Clustering dalam Pengelompokan Kabupaten/Kota di Provinsi NTB Berdasarkan Indikator Pendidikan,” *Emerging Statistics and Data Science Journal*, vol. 1, no. 3, 2023.
- [10] N. Lenama, M. A. Kleden, dan I. G. Pasangka, “K-Means Clustering Analysis pada Pengelompokan Kabupaten/Kota di Provinsi Nusa Tenggara Timur Berdasarkan Indikator Pendidikan,” *Jurnal Cakrawala Ilmiah*, vol. 2, no. 2, 2023.
- [11] A. P. Fialine, D. A. Alodia, D. Endriani, dan E. Widodo, “Implementasi Metode K-Medoids Clustering untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan,” *SEPREN: Journal of Mathematics Education and Applied*, vol. 2, no. 2, 2021.
- [12] R. R. Inkiriwang, R. Singal, dan J. V. Roeroe, “Kewajiban Negara dalam Penyediaan Fasilitas Pendidikan kepada Masyarakat Menurut Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional,” vol. 8, no. 2, 2020.
- [13] A. Angreini, R. A. Sugianto, A. N. Indahsari, R. M. Sari, dan M. R. Romahdoni, *Machine Learning dengan Python: Langkah Demi Langkah untuk Pemula*, hlm. 113–123. Jakarta, Indonesia: PT Serasi Media Teknologi, 2025.
- [14] M. R. Muttaqin dan M. Defriani, “Algoritma K-Means untuk Pengelompokan Topik Skripsi Mahasiswa,” *ILKOM Jurnal Ilmiah*, vol. 12, no. 2, pp. 121–129, 2020, doi: 10.33096/ilkom.v12i2.542.121-129.
- [15] M. Wahyudi, Mashita, R. Saragih, dan Solikhun, *Data Mining: Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering*, 1st ed. Medan, Indonesia: Yayasan Kita Menulis, 2020.