

# ANALISIS PENGELOMPOKKAN DISTRIBUSI FASILITAS PENDIDIKAN DI INDONESIA DENGAN METODE KLASTERISASI

Oleh :

Annisa Nurfadhiilah<sup>1)</sup>, Brigitta Griselda Martens<sup>2)</sup>, Arief Wibowo<sup>3)</sup>

<sup>1,2,3</sup> Fakultas Ekonomi dan Bisnis, Universitas Budi Luhur

<sup>1</sup>email: 2431600648@student.budiluhur.ac.id

<sup>2</sup>email: 2431600663@student.budiluhur.ac.id

<sup>3</sup>email: arief.wibowo@budiluhur.ac.id

## Informasi Artikel

### Riwayat Artikel :

Submit, 12 September 2025

Revisi, 1 November 2025

Diterima, 14 Januari 2026

Publish, 27 Januari 2026

### Kata Kunci :

Klasterisasi,  
K-Means,  
Fasilitas Pendidikan,  
Distribusi Pendidikan,  
Davies-Bouldin Index.



## ABSTRAK

Penelitian ini bertujuan untuk menganalisis sebaran fasilitas pendidikan di seluruh wilayah provinsi di Indonesia pada tahun 2024, mencakup jenjang mulai dari Sekolah Dasar/Madrasah Ibtidaiyah hingga tingkat Perguruan Tinggi. Sumber data berasal dari Sakernas tahun 2024 yang dirilis oleh BPS sebagai sumber data utama. Dalam penelitian ini, digunakan pendekatan statistik deskriptif untuk melihat pola distribusi fasilitas pendidikan, serta memanfaatkan metode K-Means *clustering* guna membagi provinsi ke dalam kategori berdasarkan tingkat ketersediaan fasilitas tersebut. Hasil evaluasi menggunakan Davies-Bouldin Index mengindikasikan bahwa pemodelan dengan dua klaster memiliki tingkat pemisahan terbaik dengan nilai indeks 0,083, sementara model lima klaster memberikan pembagian wilayah yang lebih rinci dengan skor indek sebesar 0,088. Hasil analisis menunjukkan ketimpangan signifikan dalam akses pendidikan antarprovinsi dan antarjenjang, dengan fasilitas cenderung berkurang pada jenjang pendidikan yang lebih tinggi. Provinsi di Pulau Jawa menunjukkan dominasi dalam jumlah fasilitas, sedangkan provinsi di Indonesia bagian timur masih menghadapi kekurangan. Penelitian ini memberikan dasar bagi perumusan kebijakan pemerataan pendidikan di Indonesia.

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



## Corresponding Author:

Nama: Arief Wibowo

Afiliasi: Universitas Budi Luhur

Email: arief.wibowo@budiluhur.ac.id

## 1. PENDAHULUAN

Pendidikan merupakan fondasi utama dalam pembangunan suatu bangsa karena berperan penting dalam menciptakan sumber daya manusia yang unggul, adaptif, dan kompetitif. Keberhasilan pembangunan di berbagai sektor, termasuk ekonomi, kesehatan, dan teknologi, sangat ditentukan oleh kualitas pendidikan yang merata dan mudah dijangkau oleh semua kalangan masyarakat. Oleh karena itu, pemerataan fasilitas pendidikan dari jenjang dasar hingga perguruan tinggi menjadi syarat krusial dalam mendukung kesetaraan akses belajar di seluruh wilayah Indonesia. Namun demikian, realitas di lapangan menunjukkan bahwa ketimpangan

distribusi fasilitas pendidikan antar provinsi masih menjadi tantangan besar.

Ketimpangan ini tidak hanya terjadi antara kawasan barat dan timur Indonesia, tetapi juga antara wilayah perkotaan dan pedesaan dalam satu provinsi. Faktor-faktor seperti kondisi geografis yang sulit dijangkau, ketimpangan jumlah penduduk, serta disparitas pembangunan infrastruktur dan layanan publik turut memperparah kesenjangan ini. Akibatnya, banyak daerah tertinggal yang belum memiliki fasilitas pendidikan memadai, terutama untuk jenjang menengah dan atas, sehingga berpotensi memperlebar kesenjangan kualitas pendidikan antar wilayah di Indonesia. Penelitian ini memakai data yang bersumber dari publikasi tahun

2024 oleh BPS, yang mencatat jumlah wilayah administrasi setingkat desa dan kelurahan yang memiliki sarana pendidikan di tiap jenjang, yaitu SD/MI, SMP/MTs, SMA/MA, SMK, dan perguruan tinggi.

Data ini memberikan gambaran menyeluruh mengenai distribusi spasial fasilitas pendidikan formal di tingkat desa/kelurahan di seluruh provinsi di Indonesia. Analisis terhadap data ini menjadi sangat penting untuk memahami pola ketersediaan dan keterjangkauan layanan pendidikan antar wilayah. Melalui pemetaan daerah yang masih mengalami keterbatasan akses terhadap fasilitas pendidikan, baik dilihat dari jumlah mutlak maupun dibandingkan dengan jumlah penduduk dan luas wilayah, penelitian ini diharapkan berpotensi menjadi acuan dalam pembuatan kebijakan yang berbasis pada data dan fakta di lapangan (*evidence-based policy*). Pendekatan semacam ini diperlukan agar alokasi sumber daya dan program pembangunan pendidikan dapat dilakukan secara lebih efektif, merata, dan berkeadilan, terutama di wilayah-wilayah yang selama ini tertinggal dalam pembangunan infrastruktur pendidikan.

Metode yang sering dipakai dalam analisis data spasial salah satunya adalah teknik pengelompokan data, seperti algoritma K-Means. Pendekatan ini memungkinkan pengelompokan objek secara cepat dan efisien berdasarkan kesamaan nilai-nilai numerik yang dimilikinya. K-Means banyak diterapkan karena mampu menangani data berskala besar serta menghasilkan segmentasi yang jelas dan terstruktur. Evaluasi terhadap hasil pengelompokan dilakukan menggunakan Davies-Bouldin Index (DBI), sebuah metrik evaluasi internal yang menilai perbandingan antara jarak antar kelompok dan sebaran data di dalam masing-masing kelompok. Skor DBI yang lebih rendah mengindikasikan bahwa klaster yang terbentuk lebih baik, karena menunjukkan pemisahan yang jelas serta kekompakan dalam tiap klaster.

Metode K-Means terbukti mampu menyederhanakan data yang rumit menjadi sejumlah kelompok yang lebih terorganisir, sehingga mempermudah proses interpretasi dan analisis. Dalam studi mereka mengenai klasifikasi kualitas manajemen pendidikan menengah di Indonesia, data dari 34 provinsi dikelompokkan ke dalam tiga klaster utama berdasarkan indikator manajemen mutu, seperti ketersediaan guru, kepala sekolah, sarana prasarana, serta pelatihan tenaga pendidik. Skor *Davies-Bouldin Index (DBI)* yang didapatkan dari hasil pengelompokan adalah 0,36, yang menunjukkan bahwa klaster kelompok yang terbentuk menunjukkan struktur yang padat dan memiliki batas pemisah yang jelas antar kelompok. Melalui visualisasi klaster dan analisis masing-masing kelompok, penelitian ini mampu mengidentifikasi wilayah dengan karakteristik manajerial yang berbeda, serta menyoroti disparitas antar wilayah

untuk mendukung formulasi kebijakan pendidikan yang berbasis data dan kontekstual. Temuan ini menegaskan bahwa pendekatan K-Means efektif dalam mengungkap ketimpangan pendidikan serta memberikan rekomendasi peningkatan yang bersifat lokal-spesifik.

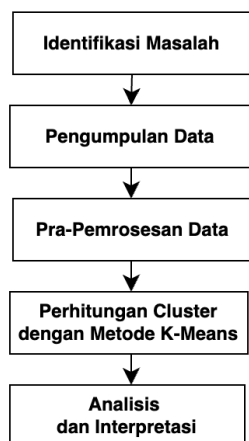
Di samping itu, penggunaan algoritma metode K-Means di dalam konteks bidang pendidikan sudah dilakukan pada berbagai studi, seperti Penelitian yang dilakukan oleh Setiawan dkk., penelitian tersebut mengelompokkan informasi calon murid baru dari Pusat Pengembangan Anak Fajar Baru Cengkareng berdasarkan karakteristik pribadi, seperti usia, jenis kelamin, dan minat belajar. Dengan menentukan nilai  $k$  sebanyak tiga klaster, proses pengelompokan menghasilkan tiga segmen utama siswa yang memiliki kecenderungan karakteristik yang berbeda. Hasil perhitungan menghasilkan skor Davies-Bouldin Index (DBI) sebesar 0,57, yang mengindikasikan bahwa klaster tersebut terbentuk cukup kompak dan terpisah dengan baik. Temuan ini menunjukkan bahwa algoritma K-Means tidak hanya efektif dalam menyederhanakan kompleksitas data, tetapi juga mampu memberikan gambaran segmentasi yang berguna sebagai dasar perencanaan strategi pembelajaran yang lebih tepat sasaran.

Dalam bidang pendidikan, algoritma K-Means telah banyak dimanfaatkan untuk mengelompokkan berbagai jenis data secara efektif. Pendekatan ini digunakan, antara lain, untuk mengelompokkan siswa berdasarkan nilai akademik, memetakan pembangunan pendidikan antarprovinsi, serta mengelompokkan wilayah administratif seperti kabupaten/kota berdasarkan indikator pendidikan. K-Means juga diterapkan untuk menganalisis performa belajar mahasiswa berdasarkan atribut akademik guna mendukung evaluasi dan peningkatan strategi pembelajaran. Selain itu, metode ini dimanfaatkan dalam pemetaan institusi pendidikan berdasarkan fasilitas dan tenaga pendidik. Secara umum, algoritma K-Means terbukti mampu menyederhanakan data yang kompleks menjadi klaster yang bermakna, sehingga memudahkan identifikasi pola dan mendukung perumusan kebijakan pendidikan yang lebih terarah dan berbasis data.

Berdasarkan keberhasilan metode K-Means dalam berbagai studi terdahulu, penelitian ini dilakukan untuk mengidentifikasi pola pengelompokan dalam sebaran fasilitas pendidikan di Indonesia dengan menerapkan metode K-Means, menggunakan data resmi dari BPS tahun 2024. Hasil klasterisasi ini diharapkan dapat memberikan gambaran yang jelas tentang pemerataan fasilitas pendidikan di seluruh provinsi, serta menjadi landasan dalam perumusan kebijakan pendidikan yang lebih adil, fokus, dan berdasarkan data.

## 2. METODE PENELITIAN

Penelitian ini memakai metode kuantitatif yang memanfaatkan algoritma *K-Means clustering* yang diimplementasikan melalui perangkat lunak RapidMiner. Tahapan pelaksanaan penelitian ditunjukkan pada Gambar 1 sebagai berikut:



Gambar 1. Tahapan Penelitian

### Tahap penelitian

Tahap awal dimulai dengan mengidentifikasi permasalahan utama, yaitu ketimpangan distribusi fasilitas pendidikan formal antarprovinsi di Indonesia. Ketimpangan ini mencerminkan adanya perbedaan signifikan dalam jumlah dan jenis fasilitas pendidikan—seperti SD/MI, SMP/MTs, SMA/MA, SMK, hingga Perguruan Tinggi—yang tersedia di setiap provinsi, yang berpotensi menimbulkan ketidakseimbangan akses pendidikan antarwilayah serta berdampak pada kesenjangan kualitas sumber daya manusia di masa depan. Fokus analisis diarahkan pada pengelompokan provinsi berdasarkan tingkat ketersediaan fasilitas pendidikan untuk mengidentifikasi pola distribusi yang ada serta mengklasifikasikan wilayah-wilayah dengan kondisi serupa, sehingga dapat menunjukkan wilayah yang memiliki fasilitas pendidikan memadai dan wilayah yang masih membutuhkan perhatian lebih dalam pembangunan infrastruktur pendidikan.

### Sumber Data

Penelitian ini memanfaatkan data yang diambil dari Badan Pusat Statistik (BPS) tahun 2024, yang berisi informasi mengenai jumlah desa atau kelurahan di setiap provinsi di Indonesia yang memiliki fasilitas pendidikan formal. Fasilitas pendidikan yang dibahas dalam penelitian ini meliputi lima tingkatan, yaitu Sekolah Dasar/Madrasah Ibtidaiyah (SD/MI), Sekolah Menengah Pertama/Madrasah Tsanawiyah (SMP/MTs), Sekolah Menengah Atas/Madrasah Aliyah (SMA/MA), Sekolah Menengah Kejuruan (SMK), serta Perguruan Tinggi. Data ini menggambarkan kondisi pemerataan infrastruktur pendidikan di seluruh wilayah Indonesia dan menjadi dasar dalam analisis untuk mengidentifikasi

perbedaan tingkat ketersediaan fasilitas pendidikan antarprovinsi.

### Pra-Pemrosesan Data

Penelitian ini memanfaatkan data yang diambil dari Badan Pusat Statistik (BPS) tahun 2024, yang berisi informasi mengenai jumlah desa atau kelurahan di setiap provinsi di Indonesia yang memiliki fasilitas pendidikan formal. Fasilitas pendidikan yang dibahas dalam penelitian ini meliputi lima tingkatan, yaitu Sekolah Dasar/Madrasah Ibtidaiyah (SD/MI), Sekolah Menengah Pertama/Madrasah Tsanawiyah (SMP/MTs), Sekolah Menengah Atas/Madrasah Aliyah (SMA/MA), Sekolah Menengah Kejuruan (SMK), serta Perguruan Tinggi. Data ini menggambarkan kondisi pemerataan infrastruktur pendidikan di seluruh wilayah Indonesia dan menjadi dasar dalam analisis untuk mengidentifikasi perbedaan tingkat ketersediaan fasilitas pendidikan antarprovinsi.

### Perhitungan Cluster dengan Metode K-Means

Data yang telah dikumpulkan selanjutnya diproses melalui tahap pembersihan (cleaning) untuk memastikan kelayakan data sebelum dianalisis. Setelah data dinyatakan bersih, dilakukan proses normalisasi untuk menyamakan skala antar variabel, mengingat jumlah desa atau kelurahan dengan fasilitas pendidikan pada tiap jenjang memiliki rentang nilai yang berbeda. Normalisasi diperlukan agar tidak ada satu variabel pun yang mendominasi dalam proses pengelompokan.

### Perhitungan Cluster dengan Metode K-Means

Teknik K-Means clustering digunakan untuk membangun yang membagi provinsi-provinsi di Indonesia berdasarkan jumlah desa atau kelurahan dengan fasilitas pendidikan formal di setiap tingkat pendidikan, mulai dari SD/MI hingga Perguruan Tinggi. Pemodelan dilakukan menggunakan perangkat lunak RapidMiner, dengan memasukkan data yang telah dinormalisasi ke dalam model kluster. Normalisasi memastikan bahwa setiap variabel memiliki kontribusi yang setara dalam proses pembentukan kluster. Setelah proses pemodelan selesai, hasilnya divisualisasikan dalam bentuk scatter plot yang menampilkan distribusi provinsi berdasarkan kesamaan karakteristik fasilitas pendidikan. Visualisasi ini membantu dalam mengidentifikasi pola-pola keterkaitan antarprovinsi dan mempermudah interpretasi hasil analisis terhadap kelompok yang terbentuk.

### Analisis dan Interpretasi

Kualitas hasil klasterisasi dinilai dengan menggunakan Davies-Bouldin Index (DBI), yang berperan sebagai alat untuk mengukur seberapa baik pemisahan serta kepadatan antar kluster. Angka DBI yang lebih kecil mengindikasikan kualitas pemodelan yang lebih maksimal, karena hal ini menandakan bahwa cluster yang terbentuk menunjukkan pemisahan yang jelas dengan anggota yang ada di setiap kelompok memiliki tingkat kemiripan yang

tinggi. Tahap ini juga mencakup interpretasi terhadap setiap cluster yang terbentuk untuk memahami pola ketersediaan fasilitas pendidikan di masing-masing kelompok provinsi berdasarkan hasil pemodelan yang telah dilakukan.

### K-Means

K-Means merupakan teknik yang dipakai untuk merangkum informasi dari data yang beragam dengan cara mengelompokkan objek-objek dengan karakteristik yang memiliki kesamaan ke dalam beberapa kategori yang lebih sederhana dan mudah dimengerti. Metode ini membantu dalam menata data agar lebih terstruktur dan memberikan gambaran umum terhadap hubungan antar elemen dalam kumpulan informasi

Dalam penerapannya, K-Means berguna untuk menyajikan data dalam bentuk yang lebih ringkas dan representatif. Hal ini memungkinkan pengguna melihat kecenderungan atau pola umum yang mungkin tidak terlihat secara langsung pada data mentah. Dengan pengelompokan yang sistematis, proses interpretasi data menjadi lebih efisien dan terarah.

K-Means juga dapat digunakan untuk meningkatkan pemahaman terhadap variasi karakteristik dalam sekumpulan informasi. Pendekatan ini tidak hanya mempermudah proses pengorganisasian data, tetapi juga membantu memperjelas perbedaan dan kesamaan yang terdapat di dalamnya.

Algoritma K-Means sering digunakan dalam analisis pemetaan data pendidikan dan sosial karena kemampuannya dalam menyederhanakan kompleksitas data dan memberikan visualisasi yang mudah dipahami. Implementasi algoritma ini juga dapat dilakukan menggunakan perangkat lunak seperti RapidMiner yang mempermudah proses pemodelan dan evaluasi cluster secara otomatis.

Dalam menilai kualitas klaster, metode ini kerap dikombinasikan dengan indeks validasi seperti Davies–Bouldin Index guna menentukan seberapa optimal pemisahan antar klaster yang dihasilkan. Metode ini telah banyak dimanfaatkan dalam berbagai penelitian, mulai dari pengelompokan sekolah, pemetaan fasilitas pendidikan, hingga analisis terkait pembangunan manusia.

## 3. HASIL DAN PEMBAHASAN

### Hasil Penelitian

Jumlah fasilitas pendidikan dari jenjang SD, SMP, hingga SMA di 38 provinsi di Indonesia menjadi dasar dalam proses klasterisasi. Data tersebut diambil dari Badan Pusat Statistik (BPS) serta telah melalui proses seleksi serta normalisasi agar siap untuk diolah lebih lanjut. Hasil representasi data yang akan diklasterisasikan ditampilkan pada Tabel 1.

Tabel 1. Jumlah Fasilitas Pendidikan Tahun 2024

No	Provinsi	SD	SMP	SMA	SMK	PT
----	----------	----	-----	-----	-----	----

No	Provinsi	SD	SMP	SMA	SMK	PT
1	Aceh	3382	1421	735	205	119
2	Sumatera Utara	5003	2319	1147	712	202
3	Sumatera Barat	1256	808	415	180	102
4	Riau	1811	1210	631	259	80
5	Jambi	1484	810	393	169	42
6	Sumatera Sel.	2996	1410	720	252	99
7	Bengkulu	1187	488	194	93	27
8	Lampung	2537	1437	695	405	96
9	Kep. BaBel	391	217	85	49	16
10	Kep. Riau	408	249	132	62	33
11	DKI Jakarta	264	255	219	207	129
12	Jawa Barat	5948	4314	2191	2070	477
13	Jawa Tengah	8423	3692	1343	1204	314
14	DI Yogyakarta	437	310	148	139	66
15	Jawa Timur	8442	4905	2587	1690	465
16	Banten	1544	1259	702	540	134
17	Bali	710	313	148	131	41
18	NTB	1157	924	587	287	79
19	NTT	3331	1699	667	341	60
20	Kalbar	2079	1153	437	189	58
21	Kalteng	1545	799	274	120	22
22	Kalsel	1878	777	323	111	56
23	Kaltim	1002	551	243	151	48
24	Kalut	312	158	63	31	13
25	Sulut	1564	716	250	177	70
26	Sulteng	1928	966	345	168	42
27	Sulsel	2972	1789	824	333	155
28	Sultra	1875	911	415	154	40
29	Gorontalo	658	363	109	55	16
30	Sulbar	631	406	168	112	23
31	Maluku	1104	626	274	100	45
32	Maluku Utara	1090	616	294	142	25
33	Papua Barat	501	147	59	20	16
34	Papua Brt Daya	460	164	78	32	18
35	Papua	655	239	106	45	27
36	Papua Selatan	545	110	32	24	10
37	Papua Tengah	430	135	50	36	16
38	Papua Peg.	530	169	48	15	8

### Pembahasan

Data jumlah fasilitas jenjang pendidikan dari sekolah dasar, menengah pertama, hingga menengah atas di 38 provinsi di Indonesia yang telah melalui proses seleksi dan normalisasi selanjutnya dieksplor untuk diolah menggunakan RapidMiner.



### Hasil Perhitungan Metode K-Means

Berdasarkan desain perhitungan klasterisasi menggunakan RapidMiner sebagaimana ditunjukkan pada gambar 2, Proses ini menghasilkan model klaster yang menggambarkan segmentasi provinsi menurut kemiripan jumlah fasilitas pendidikan dari jenjang SD, SMP, hingga SMA. Model klaster ini memungkinkan identifikasi pola ketimpangan maupun kesetaraan dalam distribusi fasilitas pendidikan antardaerah.

Proses pembentukan klaster dilakukan tanpa menetapkan jumlah kelompok secara manual, melainkan berdasarkan evaluasi kualitas pemisahan klaster yang optimal menggunakan metrik seperti *Davies-Bouldin Index*. Berikut merupakan hasil kelompok yang ditampilkan pada Tabel 2.

Tabel 2. Hasil Cluster

Jumlah Klaster	Davies Bouldin Index
Dua Klaster	0,083
Tiga Klaster	0,091
Empat Klaster	0,136
Lima Klaster	0,088

Berdasarkan pada tabel 2, didapatkan hasil yang menunjukkan bahwa Dua Klaster memiliki nilai indeks terendah yaitu 0,083, diikuti oleh Lima Klaster dengan 0,088, Tiga Klaster dengan 0,091, dan Empat Klaster dengan 0,136. Karena nilai Davies-Bouldin dengan nilai yang lebih kecil mencerminkan kualitas pengelompokan yang lebih optimal—yakni pemisahan antar *cluster* yang jelas dan kekompakan dalam *cluster*—maka posisi dua klaster menunjukkan hasil klasterisasi yang paling optimal secara statistik. Sementara lima klaster, sebagai urutan kedua dengan nilai indeks terendah, juga menunjukkan kualitas yang relatif baik dan memberikan segmentasi yang lebih rinci, sehingga tetap relevan untuk digunakan dalam analisis yang membutuhkan detail lebih spesifik. Hasil analisis klasterisasi ditunjukkan pada Tabel 3.

Tabel 3. Pemodelan Dua Klaster

Cluster	Kategori	Provinsi
A	Akses Pendidikan Berkembang	Aceh, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, DI Yogyakarta, Banten, Bali, NTB, NTT, 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, Papua Barat Daya, Papua, Papua Selatan, Papua Tengah, Papua Pegunungan
B	Akses Pendidikan Sangat Tinggi	Sumatera Utara, Jawa Barat, Jawa Tengah, Jawa Timur

Hasil pemodelan menggunakan dua klaster yang ditampilkan pada Tabel 3 menunjukkan pembagian dua kelompok utama akses pendidikan antarprovinsi di Indonesia. Klaster A, yang dikategorikan sebagai “Akses Pendidikan Berkembang”, mencakup sebagian besar provinsi, terutama yang berada di luar Pulau Jawa. Wilayah-

wilayah dalam cluster ini menunjukkan keterbatasan signifikan dalam ketersediaan fasilitas pendidikan, khususnya pada jenjang menengah dan tinggi. Hal ini mengindikasikan perlunya intervensi dan prioritas dalam pembangunan fasilitas pendidikan yang lebih merata guna menjamin kesetaraan akses pendidikan di seluruh wilayah Indonesia. Keterbatasan tersebut dapat disebabkan oleh berbagai faktor seperti hambatan geografis, kepadatan penduduk rendah, serta keterbatasan infrastruktur pendidikan. Sebaliknya, Klaster B dikategorikan sebagai “Akses Pendidikan Sangat Tinggi” dan mencakup provinsi-provinsi dengan kepadatan penduduk tinggi yang didominasi wilayah Pulau Jawa, yaitu Sumatera Utara, Jawa Barat, Jawa Tengah, dan Jawa Timur. Keempat provinsi ini menunjukkan tingkat ketersediaan fasilitas pendidikan yang relatif merata dari jenjang dasar hingga perguruan tinggi, mencerminkan tingkat pembangunan pendidikan yang lebih maju dan terstruktur. Kondisi ini menjadikan provinsi dalam klaster B sebagai model atau acuan dalam perumusan kebijakan pengembangan sarana pendidikan nasional. Dengan tingkat pemisahan cluster yang optimal berdasarkan nilai Davies-Bouldin Index sebesar 0,083. Hasil klasterisasi ini memberikan pemetaan yang jelas antara daerah dengan akses pendidikan yang sudah baik dan daerah yang masih membutuhkan peningkatan signifikan, sehingga dapat dijadikan dasar dalam menyusun strategi pemerataan pembangunan pendidikan yang lebih tepat sasaran.

Tabel 4. Pemodelan Lima Klaster

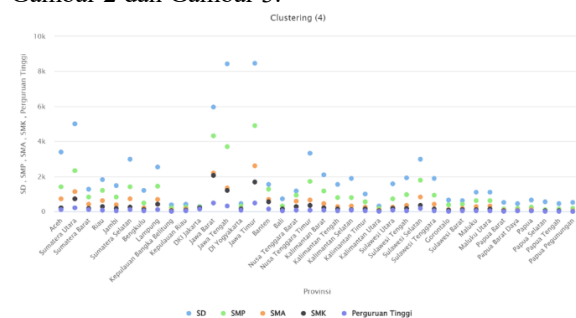
Cluster	Nama Cluster	Provinsi
C	Akses Pendidikan Tinggi	Jawa Tengah, Jawa Timur
D	Akses Pendidikan Menengah	Sumatera Barat, Jambi, Bengkulu, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, DI Yogyakarta, Bali, NTB, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua Barat Daya, Papua, Papua Selatan, Papua Tengah, Papua Pegunungan
E	Akses Pendidikan Cukup Tinggi	Sumatera Utara
F	Akses Pendidikan Berkembang	Aceh, Riau, Sumatera Selatan, Lampung, Banten, NTT, Kalimantan Barat, Kalimantan Selatan, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara
G	Akses Pendidikan Sangat Tinggi	Jawa Barat

Hasil pemodelan menggunakan lima klaster yang ditampilkan pada Tabel 4 menghasilkan segmentasi provinsi yang lebih rinci berdasarkan tingkat ketersediaan fasilitas pendidikan. Setiap cluster menggambarkan kelompok provinsi dengan karakteristik yang berbeda-beda, yang dapat menjadi dasar untuk perencanaan pembangunan pendidikan

yang lebih spesifik. Klaster C, yang dikategorikan sebagai “Akses Pendidikan Tinggi”, mencakup Jawa Tengah dan Jawa Timur yang menunjukkan ketersediaan fasilitas pendidikan yang luas dan merata di hampir semua jenjang. Sementara itu, Klaster D dengan label “Akses Pendidikan Menengah” terdiri dari 22 provinsi, yaitu Sumatera Barat, Jambi, Bengkulu, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, DI Yogyakarta, Bali, NTB, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua Barat Daya, Papua, Papua Selatan, Papua Tengah, dan Papua Pegunungan. Kelompok ini umumnya memiliki fasilitas pendidikan yang cukup memadai pada jenjang dasar dan menengah, tetapi masih memerlukan penguatan akses pada jenjang pendidikan tinggi. Klaster E, yaitu “Akses Pendidikan Cukup Tinggi”, hanya mencakup provinsi Sumatera Utara, yang memiliki kondisi akses pendidikan yang cukup baik namun belum mencapai kategori tertinggi. Selanjutnya, Klaster F dengan kategori “Akses Pendidikan Berkembang” meliputi provinsi Aceh, Riau, Sumatera Selatan, Lampung, Banten, NTT, Kalimantan Barat, Kalimantan Selatan, Sulawesi Tengah, Sulawesi Selatan, dan Sulawesi Tenggara, yang menunjukkan masih terbatasnya fasilitas pendidikan terutama pada jenjang menengah dan tinggi. Terakhir, Klaster G dikategorikan sebagai “Akses Pendidikan Sangat Tinggi”, yang hanya mencakup satu provinsi yaitu Jawa Barat. Provinsi ini menunjukkan performa tertinggi dalam ketersediaan fasilitas pendidikan di seluruh jenjang dan dapat dijadikan acuan dalam perencanaan pemerataan pendidikan nasional. Segmentasi ini memperlihatkan bahwa pendekatan cluster mampu memberikan gambaran yang lebih tajam terhadap ketimpangan akses pendidikan antarwilayah dan dapat dimanfaatkan untuk mendukung penyusunan kebijakan pendidikan yang lebih adil dan berbasis kebutuhan masing-masing provinsi.

#### Visualisasi Scatter Plot

Adapun visualisasi pada hasil pemodelan dari dua klaster dan lima klaster dapat dilihat pada Gambar 2 dan Gambar 3.



Gambar 3. Scatter Plot Lima Klaster

Pada gambar 3 tersebut menunjukkan grafik sebar (scatter plot) yang menunjukkan jumlah fasilitas pendidikan di berbagai provinsi Indonesia berdasarkan jenjang: SD, SMP, SMA, SMK, dan

Perguruan Tinggi. Warna-warna yang digunakan mewakili masing-masing jenjang pendidikan, dengan SD tampak paling dominan di sebagian besar provinsi, terutama di Jawa Barat, Jawa Tengah, dan Jawa Timur yang terdiri dari sejumlah SD sangat tinggi.

Sementara itu, jenjang seperti SMA, SMK, dan Perguruan Tinggi jumlahnya cenderung lebih sedikit dan merata di berbagai wilayah. Gambar ini menunjukkan ketimpangan distribusi fasilitas pendidikan antar jenjang dan antar provinsi, yang dapat dianalisis lebih lanjut untuk keperluan pemerataan pendidikan.

#### 4. KESIMPULAN

Analisis terhadap data fasilitas pendidikan di Indonesia tahun 2024 menunjukkan adanya ketimpangan signifikan antarprovinsi dan antarjenjang pendidikan. Secara umum, ditemukan bahwa semakin tinggi jenjang pendidikan, semakin terbatas fasilitas yang tersedia di tingkat desa/kelurahan, mencerminkan masih rendahnya akses pendidikan menengah dan tinggi di berbagai wilayah. Provinsi-provinsi di Pulau Jawa, khususnya Jawa Tengah, Jawa Timur, dan Jawa Barat, mendominasi ketersediaan fasilitas pendidikan formal, terutama pada jenjang dasar dan menengah, sementara provinsi di luar Jawa, terutama yang berada di wilayah Indonesia bagian timur, seperti Papua, Maluku, dan Nusa Tenggara Timur, masih mengalami kekurangan serius dalam hal fasilitas pendidikan, khususnya pada jenjang SMA/SMK dan perguruan tinggi. Untuk mengidentifikasi pola sebaran ini secara lebih sistematis, diterapkan algoritma K-Means yang mampu mengklasifikasikan provinsi mengacu pada kesamaan karakteristik tingkat akses pendidikan. Model dua klaster menghasilkan pemisahan paling optimal berdasarkan nilai angka Davies-Bouldin Index (DBI) yang mencapai 0,083, yang mengindikasikan kualitas pengelompokan yang efektif. Cluster pertama mencakup empat provinsi, yaitu Sumatera Utara, Jawa Barat, Jawa Tengah, dan Jawa Timur, yang digolongkan sebagai kelompok dengan “Akses Pendidikan Sangat Tinggi” karena menunjukkan pemerataan fasilitas pendidikan yang baik pada hampir semua jenjang. Cluster kedua terdiri dari 34 provinsi lainnya yang masuk dalam kategori “Akses Pendidikan Berkembang” dan membutuhkan intervensi kebijakan yang lebih kuat untuk meningkatkan fasilitas pendidikan, khususnya pada jenjang menengah dan tinggi.

Pada pemodelan lima klaster, dengan nilai DBI sebesar 0,088 memberikan segmentasi yang lebih terperinci, memungkinkan klasifikasi provinsi ke dalam lima kategori berbeda berdasarkan tingkat akses pendidikan, mulai dari sangat tinggi hingga sangat rendah. Segmentasi ini memberikan gambaran yang lebih rinci terhadap kebutuhan masing-masing daerah, sehingga dapat menjadi dasar penyusunan

strategi pembangunan yang lebih spesifik dan terarah. Temuan ini menegaskan pentingnya pendekatan berbasis data dalam perumusan kebijakan pemerataan akses pendidikan, serta dijadikan referensi oleh pemerintah pusat maupun daerah dalam mengalokasikan sumber daya secara lebih adil dan efisien guna meningkatkan kualitas dan keterjangkauan pendidikan di seluruh Indonesia.

## 5. REFERENSI

- Z. Mustakim and R. Kamal, "K-Means Clustering for Classifying the Quality Management of Secondary Education in Indonesia," *Cakrawala Pendidik.*, vol. 40, no. 3, pp. 725–737, 2021, doi: 10.21831/cp.v40i3.40150.
- Badan Pusat Statistik, "Jumlah desa/kelurahan yang memiliki fasilitas sekolah menurut provinsi dan tingkat pendidikan," 2024, [Online]. Available: <https://www.bps.go.id/id/statistics-table/3/ZURVeVVGRIjVWRMV25JeVEwVkrWa29yWldsR1p6MDkjMyMwMDAw/jumlah-desa-sup-1--sup--kelurahan-yang-memiliki-fasilitas-sekolah-menurut-provinsi-dan-tingkat-pendidikan.html>
- M. Risnasari, Nabila Aulia, and Laili Cahyani, "Clustering Of Student Learning Styles in the industri 4.0 Using KMeans Algorithm," *JTP - J. Teknol. Pendidik.*, vol. 24, no. 2, pp. 246–257, 2022, doi: 10.21009/jtp.v24i2.28029.
- K. Setiawan and Y. Y. A. Saputry, "Clustering Data Calon Siswa Baru Menggunakan Metode K-Means di Pusat Pengembangan Anak Fajar Baru Cengkareng," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 8, no. 1, pp. 75–83, 2024, doi: 10.35870/jtik.v8i1.1426.
- A. Yudistira and R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," *J. Artif. Intell. Technol. Inf.*, vol. 1, no. 1, pp. 20–28, 2023, doi: 10.58602/jaiti.v1i1.22.
- Nurfajriyani, Dentina Dewi Amaliana, and Sri Pingit Wulandari, "Pengelompokan Provinsi berdasarkan Aspek Pembangunan Pendidikan di Indonesia Tahun 2023 menggunakan Analisis Cluster," *Pentagon J. Mat. dan Ilmu Pengetah. Alam*, vol. 2, no. 4, pp. 134–153, 2024, doi: 10.62383/pentagon.v2i4.309.
- N. Lenama, M. A. Kleden, and I. G. Pasangka, "K-Means Clustering Analysis pada Pengelompokan Kabupaten/Kota di Provinsi Nusa Tenggara Timur Berdasarkan Indikator Pendidikan," *J. Cakrawala Ilm.*, vol. 2, no. 9, p. 3365, 2023, doi: 10.53625/jcijurnalcakrawalailmiah.v2i9.5653.
- S. J. Sultan Alalawi, I. N. Mohd Shahrane, and J. Mohd Jamil, "Clustering Student Performance Data Using K-Means Algorithms," *J. Comput. Innov. Anal.*, vol. 2, no. 1, pp. 41–55, 2023, doi: 10.32890/jcia2023.2.1.3.
- M. R. Sholihin and Rudiman, "Pemetaan dan klasterisasi sekolah Muhammadiyah di Kabupaten PPU berdasarkan fasilitas, pendidik, dan tenaga pendidik menggunakan metode K-Means clustering," *Explor. IT J. Keilmuan dan Apl. Tek. Inform.*, vol. 14, no. 2, pp. 45–51, 2022.
- A. J. M. S. Arockiam and E. S. Irudhayaraj, "Reclust: an efficient clustering algorithm for mixed data based on reclustering and cluster validation," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 29, no. 1, pp. 545–552, 2023, doi: 10.11591/ijeecs.v29.i1.pp545-552.
- Y. D. Putra, M. Sudarma, and I. B. Swamardika, "Clustering History Data Penjualan Menggunakan Algoritma K-Means," *Maj. Ilm. Teknol. Elektro*, vol. 20, no. 2, p. 195, 2021, doi: 10.24843/mite.2021.v20i02.p03.
- I. Fahmiyah and R. A. Ningrum, "Human Development Clustering in Indonesia: Using K-Means Method and Based on Human Development Index Categories," *J. Adv. Technol. Multidiscip.*, vol. 2, no. 1, pp. 27–33, 2023, doi: 10.20473/jatm.v2i1.45070.
- Springer, "KMeans algorithm with RapidMiner in clustering school participation rate in Indonesia," *Data Min. Appl.*, pp. 70–85, 2021, doi: doi.org/10.1007/978-981-33-4597-3\_70.
- Y. A. Wijaya, D. A. Kurniady, E. Setyanto, W. S. Tarihoran, D. Rusmana, and R. Rahim, "Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities," *TEM J.*, vol. 10, no. 3, pp. 1099–1103, 2021, doi: 10.18421/TEM103-13.
- M. Jelita, "Analisis Clustering Menggunakan Metode K-Means untuk Mengelompokkan Kabupaten/Kota di Indonesia berdasarkan UnsurUnsur Pembangun Literasi Masyarakat (UPLM)," *Semin. Nas. Off. Stat.*, vol. 2024, no. 1, pp. 701–710, 2024, doi: 10.34123/semnasoffstat.v2024i1.2188.
- J. Novaldi and A. W. Wijayanto, "Analisis Cluster Kualitas Pemuda di Indonesia pada Tahun 2022 dengan Agglomerative Hierarchical dan K-Means," *Komputika J. Sist. Komput.*, vol. 12, no. 2, pp. 91–99, 2023, doi: 10.34010/komputika.v12i2.10348.