p-ISSN : 2406-7733
e-ISSN : 2597-9922

KLASTERISASI *K-MEANS* TERHADAP PARTISIPASI PENDIDIKAN PADA WILAYAH SISWA SEKOLAH TERTINGGI DI INDONESIA

Muhamad Nawawi1*[1], Dony Waluya Firdaus[2], Ridwan Zulkifli[3], Adam Mukharil Bachtiar[4], Hanhan Maulana[5], Bobi Kurniawan[6], Ednawati Rainarli[7]
Program Studi Doktor Sistem Informasi, Fakultas Pascasarjana, Universitas Komputer Indonesia
Jl. Dipati Ukur No.112-116, Bandung
E-mail: *nawawi.75725028@mahasiswa.unikom.ac.id[1], dony.75725004@mahasiswa.unikom.ac.id[2], ridwan.75725009@mahasiswa.unikom.ac.id[3] adam@email.unikom.ac.id[4], hanhan@email.unikom.ac.id[5], bobi@email.unikom.ac.id[6], ednawati.rainarli@email.unikom.ac.id[7]

Abstrak - Pendidikan berfungsi sebagai pilar utama pembangunan nasional, namun ketimpangan akses dan partisipasi antar wilayah masih menjadi tantangan kritis di Indonesia. Tujuan dari penelitian ini adalah untuk melakukan pemodelan klasterisasi tingkat partisipasi 3 provinsi dengan tingkat angka sekolah tertinggi di Indonesia yakni Jawa Barat, Jawa Tengah dan Jawa Timur dengan jumlah 100 kabupaten dan kota. Penilaian menggunakan variabel Angka Partisipasi Sekolah (APS), Angka Partisipasi Murni (APM), dan Angka Partisipasi Kasar (APK) yang diturunkan menjadi 9 indikator di setiap jenjang yakni Sekolah Dasar/sederajat (SD), Sekolah Menengah Pertama/sederajat (SMP), dan Sekolah Menengah Atas/sederajat (SMA). Data sekunder yang digunakan berasal dari Badan Pusat Statistik (BPS) Indonesia untuk tahun 2023 dengan tools pemodelan menggunakan aplikasi Orange Data Mining. Metode *K-Means Clustering* diterapkan dan jumlah klaster yang optimal ditentukan melalui analisis *Silhouette Coefficient*, yang menghasilkan dua klaster dengan nilai rata-rata *Silhouette* lebih dari 50%, menandakan kualitas klasterisasi baik. Penelitian ini memetakan partisipasi pendidikan di tiga provinsi di Pulau Jawa menggunakan algoritma *K-Means Clustering* pada 100 kabupaten/kota. Hasil analisis menunjukkan bahwa ketimpangan pendidikan tidak lagi terkonsentrasi pada jenjang dasar, melainkan bergeser ke transisi pendidikan menengah, khususnya SMA. Hal ini tercermin dari perbedaan signifikan antara dua klaster, di mana wilayah prima memiliki APS SMA mendekati 80%, sementara wilayah atensi hanya mencapai 66,80%. Sebaliknya, partisipasi pada jenjang SD dan SMP relatif merata, menandakan keberhasilan pendidikan dasar namun masih lemahnya pencapaian program Wajib Belajar 12 Tahun.

Kata Kunci: Jawa Barat, Jawa Tengah, Jawa Timur, Klasterisasi, *K-Means*, Partisipasi Pendidikan

I. PENDAHULUAN

Dalam era globalisasi dan transformasi sosial yang berlangsung cepat, pendidikan tidak lagi dipandang sekadar sebagai sarana transfer pengetahuan, melainkan sebagai instrumen strategis dalam membentuk daya saing bangsa dan keberlanjutan pembangunan nasional. Negara-negara dengan sistem pendidikan yang kuat cenderung memiliki kapasitas adaptasi yang lebih baik terhadap perubahan ekonomi, teknologi, dan sosial. Oleh karena itu, pembangunan pendidikan menjadi agenda prioritas dalam kebijakan publik, khususnya di negara berkembang seperti Indonesia yang memiliki karakteristik wilayah, demografi, dan sosial-ekonomi yang sangat beragam. Keragaman tersebut menuntut adanya pendekatan analitis yang mampu menggambarkan kondisi pendidikan secara komprehensif dan berbasis data, sehingga kebijakan yang dirumuskan dapat menjawab kebutuhan nyata di setiap wilayah.

Pendidikan adalah aspek mendasar dalam membangun suatu negara, kemajuan suatu bangsa sangat dipengaruhi oleh kualitas pendidikannya. Meningkatnya tingkat pendidikan di suatu negara berdampak langsung pada perkembangan ekonomi, kualitas sumber daya manusia, dan mutu hidup masyarakat. Pendidikan dasar (*primary education*) berperan sebagai pondasi awal yang mendukung upaya peningkatan kualitas pendidikan di Indonesia. Untuk mencapai kemajuan, penguasaan pendidikan dasar merupakan persyaratan minimum bagi setiap warga negara Hammam & Utomo (2023), (Thoriq Basalamah & Setyadi, n.d.).

Meskipun demikian, kualitas pendidikan di Indonesia saat ini masih menghadapi masalah ketidakmerataan (Robbani & Fatah, 2024), Disparitas pendidikan yang signifikan persisten terjadi antar wilayah, terutama antara daerah perkotaan dan pedesaan, serta antara kelompok sosial-ekonomi yang berbeda,. Kesenjangan ini mencakup isu distribusi sumber daya, alokasi anggaran, kualitas guru, dan aksesibilitas infrastruktur pendidikan (Sidauruk dkk, 2025). Untuk mengukur tingkat keberhasilan pembangunan pendidikan, digunakan indikator partisipasi, seperti Angka Partisipasi Sekolah (APS), Angka Partisipasi Kasar (APK), dan Angka Partisipasi Murni (APM) (Putri dkk., 2023). Data menunjukkan bahwa presentase partisipasi cenderung menurun seiring dengan kenaikan jenjang pendidikan, ketidakseimbangan ini menyulitkan pengukuran kualitas pendidikan secara komprehensif (Arfan & Pekei, 2025).

p-ISSN : 2406-7733
e-ISSN : 2597-9922

Menanggapi tantangan ketidakmerataan ini, teknik klasterisasi data dapat digunakan untuk mengidentifikasi pola dan mengelompokkan wilayah berdasarkan indikator pendidikan (Robbani & Fatah, 2024). Algoritma *K-Means Clustering* adalah metode yang banyak digunakan untuk mengelompokkan data yang besar berdasarkan kesamaan fitur (Iswahyudi & Rahamwati, 2025), digunakan karena merupakan algoritma yang robust (tangguh) untuk mengidentifikasi pola heterogenitas pendidikan di Indonesia, memungkinkan deteksi dini terhadap daerah-daerah yang memerlukan intervensi kebijakan prioritas berdasarkan karakteristik data aktual. Dengan menggunakan klasterisasi, pemerintah dapat memperoleh gambaran yang lebih jelas mengenai distribusi kualitas pendidikan (Manik & Ariesta, 2023), dan merancang kebijakan intervensi yang lebih tepat sasaran untuk setiap kelompok wilayah (Arfan & Pekei, 2025).

Tujuan dari penelitian ini adalah untuk mengidentifikasi dan memetakan pola partisipasi pendidikan pada tingkat kabupaten/kota di Pulau Jawa dengan menggunakan algoritma *K-Means Clustering* berdasarkan indikator Angka Partisipasi Sekolah (APS), Angka Partisipasi Kasar (APK), dan Angka Partisipasi Murni (APM). Penelitian ini bertujuan mengelompokkan wilayah ke dalam klaster-klaster yang memiliki karakteristik partisipasi pendidikan yang serupa, sehingga dapat mengungkap tingkat ketimpangan antarwilayah secara lebih rinci. Selain itu, penelitian ini diarahkan untuk menentukan jenjang pendidikan yang menjadi titik kritis ketimpangan serta menyediakan dasar empiris bagi perumusan kebijakan pendidikan yang lebih terarah, berbasis data, dan sesuai dengan kebutuhan spesifik masing-masing klaster wilayah.

Penelitian ini berkontribusi dalam memperkaya studi demografi pendidikan melalui pemanfaatan teknik data mining untuk membedah disparitas wilayah secara granular. Jika penelitian sebelumnya seringkali terbatas pada analisis agregat tingkat provinsi yang menyembunyikan variasi lokal, penelitian ini menggunakan algoritma *K-Means Clustering* pada 100 kabupaten dan kota untuk mendekonstruksi pola ketimpangan spasial di Pulau Jawa. Kontribusi utama dari studi ini adalah teridentifikasinya partisipasi pendidikan yang nyata dari Provinsi Jawa Barat yang sedikit tertinggal dari Provinsi Jawa Timur. Temuan ini melengkapi literatur yang ada dengan memberikan bukti empiris bahwa tantangan pendidikan di wilayah padat penduduk tidaklah homogen, sehingga menuntut strategi penanganan yang tersegregasi berdasarkan profil klaster wilayah masing-masing.

II. TINJAUAN PUSTAKA

Data mining adalah proses pencarian pola yang menarik dan tersembunyi dari kumpulan data yang berukuran besar (Manik & Ariesta, 2023), konseptualisasi ini sering digunakan secara bergantian dengan proses penggalian informasi bernilai dari basis data yang melimpah (Basalamah & Setyadi, 2023). Klasterisasi adalah metode penemuan data (data mining) yang bersifat unsupervised (tanpa arahan), yang bertujuan untuk mengelompokkan sekumpulan objek atau data ke dalam kelompok (klaster) berdasarkan kemiripan karakteristik (Iswahyudi & Rahamwati, 2025), objek di dalam satu klaster akan memiliki kemiripan yang tinggi satu sama lain, namun berbeda signifikan dengan objek di klaster lain (Yolanda & Eriani, 2022).

K-Means adalah algoritma clustering berbasis partisi yang populer (Arfan & Pekei, 2025). Algoritma ini dipilih karena memiliki keunggulan dalam efisiensi untuk data yang besar dan mampu meningkatkan kecepatan proses pengelompokan (Iswahyudi & Rahamwati, 2025).

Tahapan kerja algoritma K-Means secara umum meliputi (Qori'atunnadyah, 2023):

1) Inisialisasi: Menentukan jumlah klaster (K) yang diinginkan
2) Penentuan *Centroid*: Memilih K titik pusat *(centroid*) awal, yang dapat dilakukan secara acak
3) Langkah Penugasan: Menghitung jarak setiap data ke setiap *centroid*, umumnya menggunakan *Euclidean Distance* dan menetapkan setiap data ke klaster dengan jarak terdekat
4) Pembaruan *Centroid*: Menghitung ulang *centroid* baru sebagai nilai rata-rata dari semua titik data dalam klaster tersebut
5) Iterasi: Proses penugasan dan pembaruan centroid diulang hingga posisi *centroid* stabil, atau keanggotaan klaster tidak berubah (Qori'atunnadyah, 2023).

Kualitas klasterisasi yang dihasilkan dapat dievaluasi menggunakan metrik seperti *Silhouette Coefficient* (SC) atau *Davies Bouldin Index* (DBI) (Rais dkk, 2020). Nilai SC antara 0.5 dan 0.7 menunjukkan struktur klaster yang tergolong standar (Putri dkk., 2023). DBI digunakan untuk memilih jumlah cluster terbaik, di mana nilai DBI yang lebih kecil (mendekati 0) mengindikasikan hasil klasterisasi yang lebih optimal (Nurahman dkk, 2022).

Indikator Partisipasi Pendidikan (APS, APM, APK) merupakan alat kunci untuk menilai peluang akses pendidikan dan tingkat keberhasilan pembangunan pendidikan (Sumarah & Wulandari, 2023) :

p-ISSN : 2406-7733
e-ISSN : 2597-9922

1) Angka Partisipasi Sekolah (APS): Digunakan untuk menilai peluang akses pendidikan bagi suatu kelompok usia tertentu. Penelitian lain juga berfokus pada analisis APS di berbagai jenjang usia, (Sumarah & Wulandari, 2023).
2) Angka Partisipasi Murni (APM): Merefleksikan seberapa besar proporsi anak dalam kelompok usia sekolah yang sesuai jenjang pendidikan tertentu benar-benar bersekolah di jenjang tersebut (Silviana dkk, 2025). Di beberapa daerah, APM pendidikan dasar masih menjadi tantangan serius dan belum mencapai 100 persen (Utomo, 2023).
3) Angka Partisipasi Kasar (APK): Menghitung rasio jumlah siswa pada jenjang pendidikan tertentu (usia berapapun) terhadap jumlah penduduk dalam kelompok usia sekolah resmi untuk jenjang tersebut (Putri dkk., 2023).

Penerapan algoritma K-Means dalam klasterisasi sektor pendidikan telah dilakukan secara luas (Mannan, 2025):

1) Studi klasterisasi provinsi di Indonesia berdasarkan indikator pendidikan pada tahun 2020 menunjukkan bahwa kualitas klaster terbaik adalah pada K=3 dengan nilai Silhouette Coefficient sebesar 0.6308 (Putri dkk., 2023).
2) Penelitian mengenai tingkat putus sekolah (SD, SMP, SMA, SMK) di Indonesia menggunakan *K-Means* menemukan hasil terbaik pada K=2 dengan nilai *Silhouette* tertinggi yaitu 0.722. Klasterisasi ini menghasilkan kelompok daerah dengan tingkat putus sekolah tinggi (40 daerah) dan rendah (474 daerah) (Mannan, 2025).
3) Studi klasterisasi tingkat penyelesaian pendidikan (SD, SMP, SMA) dengan K=3 menemukan bahwa setiap kenaikan jenjang pendidikan (dari SD ke SMP dan SMA), persentase tingkat penyelesaian mengalami penurunan. Provinsi seperti Papua sering tergabung dalam klaster dengan tingkat penyelesaian rendah (Basalamah & Setyadi, 2023).
4) Klasterisasi wilayah juga dapat didasarkan pada faktor pendukung pendidikan seperti jumlah sekolah, guru, dan fasilitas. Beberapa studi yang mengelompokkan provinsi berdasarkan fasilitas menunjukkan pembagian yang jelas antara klaster tinggi (misalnya Jawa Barat, Jawa Tengah, Jawa Timur) dan klaster rendah (31 provinsi lainnya) (Nurahman dkk, 2022).

III. METODE PENELITIAN

Penelitian ini menggunakan metode kuantitatif deskriptif dengan pendekatan *Data Mining*. Metode clustering yang digunakan adalah *K-Means Clustering* untuk mempartisi data (Basalamah & Setyadi, 2023).

Data yang digunakan adalah data sekunder yang data bersumber dari publikasi resmi Badan Pusat Statistik (BPS) Indonesia. Unit Analisis: 100 kabupaten kota dari 3 provinsi di Indonesia sesuai dengan yang ditunjukkan oleh tabel 1, Variabel Penelitian: Variabel yang dianalisis meliputi indikator partisipasi pendidikan pada jenjang SD, SMP, dan SMA di tahun 2023 yaitu:

1) Angka Partisipasi Sekolah (APS)
2) Angka Partisipasi Kasar (APK)
3) Angka Partisipasi Murni (APM)

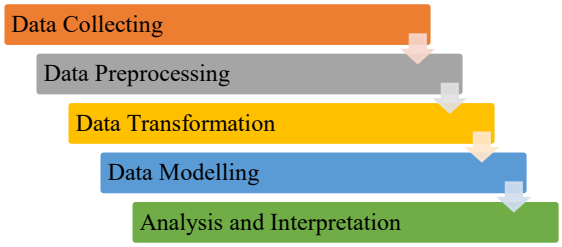
Tabel 1. Sumber Data

No.	Provinsi	Jumlah Kabupaten/ Kota	Website
1	Jawa Barat	27	https://jabar.bps.go.id/id
2	Jawa Tengah	35	https://jateng.bps.go.id/id
3	Jawa Timur	38	https://jatim.bps.go.id/id

Tabel 1 menyajikan cakupan wilayah penelitian yang meliputi tiga provinsi di Pulau Jawa, yaitu Jawa Barat, Jawa Tengah, dan Jawa Timur, beserta jumlah kabupaten/kota yang dijadikan sebagai unit analisis. Provinsi Jawa Barat terdiri atas 27 kabupaten/kota, Provinsi Jawa Tengah mencakup 35 kabupaten/kota, sedangkan Provinsi Jawa Timur memiliki jumlah wilayah administratif terbanyak, yaitu 38 kabupaten/kota. Data penelitian diperoleh dari situs resmi Badan Pusat Statistik (BPS) pada masing-masing provinsi, yang menyediakan data statistik pendidikan secara terstandar dan dapat dipertanggungjawabkan. Penyajian informasi ini bertujuan untuk menegaskan ruang lingkup penelitian serta memastikan bahwa analisis yang dilakukan merepresentasikan kondisi pendidikan pada tingkat kabupaten/kota di wilayah Pulau Jawa secara komprehensif dan sistematis.

Variabel APS, APM, dan APK tersebut diturunkan menjadi 9 indikator untuk setiap jenjang yakni SD, SMP, dan SMA dengan jumlah data sebanyak 100 sehingga terdapat 900 nilai dengan rentang minimal 43,96 yang merupakan APM Kabupaten Sampang dan maksimal pada 118,3 yang merupakan APK Kota Batu dimana keduanya terletak di Provinsi Jawa Timur.

p-ISSN : 2406-7733
e-ISSN : 2597-9922



Gambar 1. Tahapan penelitian

Tahapan penelitian yang dilakukan dapat dilihat pada gambar 1 mengacu pada langkah-langkah dalam proses data mining *atau Knowledge Discovery in Database* (KDD), (Nurahman dkk, 2022).

1) Pengumpulan Data (*Data Acquisition*): Data variabel APS, APM, dan APK dengan 9 indikator untuk setiap jenjang yakni SD, SMP, dan SMA dari 3 provinsi dengan angka sekolah tertinggi di Indonesia pada tahun 2023 yakni Jawa Barat, Jawa Tengah dan Jawa Timur dengan jumlah 100 kabupaten dan kota dikumpulkan dari BPS (Musthofiyah & Isnainiyah, 2020).
2) Pra-pemrosesan Data (*Data Preprocessing*): Melakukan pembersihan data dan normalisasi. Normalisasi data (misalnya menggunakan *standard scaler*) bertujuan untuk menyamakan nilai fitur, yang diperlukan untuk meningkatkan kualitas model *K-Means* (Putri dkk., 2023). Penggabungan data yang terpisah menjadi satu file berisi keseluruhan indikator sebelum dilakukan proses klasterisasi. Kemudian dilakukan pembersihan (*cleaning*) data terkait *record* yang tidak sesuai dalam hal ini berisi nilai APS, APM dan APK Indonesia.
3) Transformasi data dengan pemilihan Jumlah Klaster (K): Jumlah klaster yang optimal ditentukan melalui pengujian berulang (misalnya dari K=2 hingga K=7) (Putri dkk., 2023). Di tahap ini ditemukan bahwa 2 klaster ideal yang memilki nilai *silhouette* tertinggi.
4) Pemodelan Klaster (*Clustering*): Algoritma K-Means diterapkan menggunakan perhitungan jarak *Euclidean Distance* (Basalamah & Setyadi, 2023). Proses pengolahan data dilakukan menggunakan perangkat lunak *Orange Data Mining* (Manik & Ariesta, 2023). Pemodelan data dilakukan dengan beberapa *widget* visual yang tersedia di aplikasi orange, diantaranya *Silhouette Plot, Data Table* serta *Geo Map*. Di tahap ini sudah ditemukan sejumlah 3 klaster ideal yang dilakukan di tahapan sebelumnya.
5) Evaluasi dan Interpretasi Hasil: Kinerja klasterisasi dievaluasi dengan metrik seperti *Silhouette Coefficient* (Putri dkk., 2023) atau Davies Bouldin Index (DBI) (Nurahman dkk, 2022). Pola yang dihasilkan diinterpretasikan untuk memberikan wawasan yang mudah dipahami bagi pihak berkepentingan (Nurahman dkk, 2022). Hasil klasterisasi divisualisasikan menggunakan *Geo Map* dengan menambahkan data *latitude* dan *longitude* masing-masing kabupaten dan kota untuk melihat sebaran klaster.

IV. HASIL DAN PEMBAHASAN

Hasil

Analisis *elbow method* pada nilai *within-cluster sum of* squares (WCSS) untuk k = 2 sampai k = 8 menunjukkan titik siku yang cukup jelas pada k = 2. Nilai *silhouette coefficient* untuk k = 2 sebesar 0,217 merupakan yang paling tinggi diantara yang lain,. Berdasarkan indikator tersebut diputuskan menggunakan tiga klaster (k = 2).

Berdasarkan hasil pemodelan klasterisasi menggunakan *K-Means* menghasilkan nilai rata-rata *silhouette* tiap klaster > 50% artinya kualitas klasterisasi baik, masing-masing nilai *silhouette* dapat dilihat dari tabel 2 berikut :

Tabel 2. Nilai rata-rata *silhouette*

Klaster	Nilai rata-rata *Silhouette*
C1	0,58
C2	0,56

Selanjutnya dilakukan interpretasi per-klaster sebagai berikut:

Tabel 3. Nilai rata-rata klaster 1

Klaster 1	SD	SMP	SMA
APM	98,43	84,57	68,94
APS	99,54	98,28	79,24
APK	104,23	94,18	97,33

Berdasarkan tabel 3 dapat dilihat bahwa klaster 1 memiliki angka serapan yang baik di jenjang SD dan SMP serta transisi ke SMA cukup baik dengan APS 79,24%. APM hampir 100% menunjukkan angka tepat usia jenjang SD namun transisi ke SMP dan SMA menunjukkan penurunan angka tepat usia sekolah. Angka APK 104,23% menunjukkan semua anak tertampung di jenjang SD meskipun masih ada indikasi keterlambatan ataupun masuk usia sekolah terlalu cepat, angka APK jenjang SMP dan SMA cukup baik di angka lebih dari 90%. Klaster ini merupakan wilayah prima dengan tingkat partisipasi yang baik ditunjukkan dengan rata-rata APS 92,35%.

Tabel 4 Nilai rata-rata klaster 2

Klaster 2	SD	SMP	SMA
APM	98,32	81,54	56,51
APS	99,40	95,60	66,80
APK	104,45	93,75	77,45

p-ISSN : 2406-7733
e-ISSN : 2597-9922

Berdasarkan tabel 4 dapat dilihat bahwa pada klaster 2 terdapat masalah keterserapan partisipasi pendidikan saat transisi ke SMA, ditunjukkan dengan APS 66,8%. APK 77,45% menunjukkan ada sekitar 22,55% anak yang tidak tertampung untuk melanjutkan sekolah ke jenjang SMA. Masalah ketepatan usia sekolah juga terpengaruh dengan hanya 56,51% siswa SMA yang tepat sekolah sesuai usia. Klaster di wilayah ini perlu atensi penuh terkait keberlanjutan studi ke jenjang SMA.

Tabel 5 Perbandingan rata-rata (*Centroid*)

Indikator	C1	C2	Selisih
APM SD/sederajat	98,43	98,32	0,11
APM SMP/sederajat	84,57	81,54	3,03
APM SMA/sederajat	68,94	56,51	12,42
APS SD/sederajat	99,54	99,40	0,14
APS SMP/sederajat	98,28	95,60	2,68
APS SMA/sederajat	79,24	66,80	12,44
APK SD/sederajat	104,23	104,45	-0,22
APK SMP/sederajat	94,18	93,75	0,43
APK SMA/sederajat	97,33	77,45	19,88

Tabel 5 menunjukkan bahwa klaster 1 memiliki nilai centroid lebih baik untuk semua indikator dibandingkan klaster 2. Hal ini sekaligus mempertegas hasil analisis setiap klaster yang telah dijelaskan sebelumnya, terutama rendahnya transisi melanjutkan studi ke jenjang SMA, yang ditunjukkan dengan selisih yang tinggi di setiap indikatornya baik itu APM, APS, dan APK.



Gambar 2. *Geo Map* klasterisasi tingkat partisipasi pendidikan di Indonesia

Gambar 2 menjelaskan bahwa klaster 2 didominasi oleh wilayah yang berada di Provinsi Jawa Barat dengan jumlah 21 kabupaten dan kota atau 75% dari keseluruhan kabupaten dan kota di Jawa Barat. Ini mengindikasikan bahwa tingkat partisipasi pendidikan terutama pada jenjang SMA di wilayah tersebut paling rendah diantara dua provinsi lainnya. Detail terkait jumlah dan presentase wilayah dapat dilihat pada tabel 6.

Tabel 6 Jumlah dan presentase klaster per-provinsi

Provinsi	Jumlah C1	Jumlah C2	Total	Presentase C1	Presentase C2
Jabar	7	21	28	25%	75%
Jateng	18	17	35	52%	48%
Jatim	26	12	38	68%	32%

Temuan lain dari hasil klasterisasi ini adalah mengenai angka partisipasi pendidikan di Ibukota Provinsi yang menghasilkan Kota Surabaya berada pada klaster 2, angka APS hanya 65% lebih kecil dibanding dengan nilai centroid klaster 2 yakni 66,8%. Ini adalah temuan yang sangat menarik dalam studi demografi dan pendidikan. Fakta bahwa Surabaya yang merupakan Ibukota & Kota Metropolitan justru masuk ke klaster 2 (wilayah perlu atensi), sementara Provinsi Jawa Timur secara umum didominasi oleh klaster 1 (wilayah prima), bisa dinarasikan sebagai sebuah anomali yang membutuhkan pendekatan khusus. Detail terkait hasil klasterisasi Ibukota Provinsi dapat dilihat pada tabel 7.

Tabel 7. Hasil klaster Ibukota Provinsi

Ibukota Provinsi	Cluster	APM SMA/sederajat	APS SMA/sederajat	APK SMA/sederajat
Kota Semarang (Jateng)	C1	65,55	72,93	88,18
Kota Bandung (Jabar)	C1	63,76	69,66	93,60
Kota Surabaya (Jatim)	C2	57,15	65,29	83,06

Hasil klasterisasi partisipasi pendidikan pada provinsi dengan angka sekolah tertinggi di Indonesia yakni Jawa Barat, Jawa Tengah dan Jawa Timur menggunakan algoritma *K-Means*, diperoleh dua kelompok wilayah/data yang kemudian dikategorikan menjadi wilayah prima dan wilayah perlu atensi. Kategori ini ditentukan berdasarkan rata-rata nilai setiap indikator pada masing-masing klaster. Klaster dengan nilai indikator paling tinggi dikategorikan sebagai C1 (wilayah prima) dan klaster dengan nilai terendah sebagai C2 (wilayah perlu atensi). Detail terkait kategori klaster dapat dilihat pada tabel 8.

Tabel 8. Kategori klaster

Cluster	Kategori Klaster
C1	Wilayah Prima
C2	Wilayah Perlu Atensi

p-ISSN : 2406-7733
e-ISSN : 2597-9922

Pembahasan

Penggunaan algoritma *K-Means Clustering* dalam penelitian ini terbukti baik dalam memetakan disparitas pendidikan yang tersembunyi. Metode *clustering* mampu mengelompokkan provinsi atau wilayah berdasarkan indikator pendidikan secara objektif untuk membantu pemerintah menentukan strategi pembangunan yang tepat sasaran. (Basalamah & Setyadi, 2023).

Temuan penelitian ini memperkuat argumen yang menyatakan perlunya identifikasi daerah tertinggal ("klaster atensi") (Robbani & Fatah, 2024). Namun, jika penelitian tersebut berfokus pada jenjang SD, penelitian ini memberikan kebaruan (*novelty*) dengan membuktikan bahwa diferensiasi klaster saat ini lebih ditentukan oleh jenjang pendidikan menengah (SMA/sederajat), mengingat jenjang dasar (SD) relatif sudah merata.

Dominasi Jawa Barat dalam Cluster 2 (wilayah perlu atensi) mengindikasikan adanya hambatan struktural yang masif. Mengacu pada penelitian segmentasi wilayah di Jawa Barat, provinsi ini memiliki kompleksitas demografi tertinggi di Indonesia yang seringkali menghambat pemerataan pembangunan manusia (termasuk pendidikan dan gender) (Iswahyudi & Rahamwati, 2025).

Fenomena ketertinggalan ini dapat dijelaskan melalui perspektif *Supply-Side* (Ketersediaan). Ketimpangan distribusi guru dan ketersediaan sekolah negeri masih menjadi isu utama. Di wilayah geografis yang luas dan berbukit seperti Jawa Barat bagian Selatan, akses fisik ke SMA/SMK menjadi kendala utama yang menyebabkan rendahnya Angka Partisipasi Kasar (APK), berbeda dengan Jawa Timur yang memiliki sebaran sekolah yang lebih merata hingga ke pelosok (Arfan & Pekei, 2025)

Temuan anomali di mana Kota Surabaya masuk ke dalam klaster 2 (wilayah perlu atensi) meskipun merupakan pusat ekonomi dan mendapat dukungan teoritis dari Sidauruk et al. Dalam artikelnya "*Indonesia's Struggle with Education Inequality*", Sidauruk menekankan bahwa investasi besar dan status wilayah urban tidak menjamin kesetaraan akses. Kesenjangan sosio-ekonomi (*socio-economic gaps*) di wilayah urban seringkali lebih tajam daripada di pedesaan (Sidauruk dkk, 2025)

Hal ini juga relevan dengan perspektif mengenai *educational disparity* dan keadilan sosial. Di Surabaya, tantangan pendidikan bukan lagi soal ketersediaan gedung sekolah (seperti di Jawa Barat), melainkan ketidakmampuan kelompok marginal untuk mengakses pendidikan lanjut karena tingginya biaya hidup dan tuntutan ekonomi (Khadafi dkk, 2025)

Imperatif Wajib Belajar 12 Tahun Hasil clustering yang menunjukkan kesenjangan tajam pada level SMA menjadi bukti empiris bagi argumen Khadafi et al. (2025) dan Faradiba & Nomleni (2024). Khadafi et al. yang menekankan bahwa perencanaan Wajib Belajar 12 Tahun membutuhkan strategi yang matang dan sistematis karena tantangannya jauh lebih berat dibanding Wajib Belajar 9 Tahun (Raspatiningrum dkk, 2025)

Analisis jangka panjang menunjukkan bahwa meskipun indikator pendidikan meningkat sejak 1994, perlambatan (stagnasi) sering terjadi pada transisi SMP ke SMA. Oleh karena itu, kebijakan pemerintah tidak bisa lagi dipukul rata. Wilayah klaster 2 membutuhkan intervensi spesifik berupa afirmasi pendidikan menengah untuk mencegah *lost generation* pada usia produktif (Faradiba & Nomleni, 2024)

V. KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini berhasil memetakan peta partisipasi pendidikan di tiga provinsi dengan angka sekolah tertinggi di Indonesia melalui penerapan algoritma *K-Means Clustering* pada 100 kabupaten dan kota. Hasil analisis menegaskan bahwa titik kritis ketimpangan pendidikan di Pulau Jawa saat ini tidak lagi terletak pada jenjang dasar, melainkan telah bergeser secara signifikan ke transisi jenjang pendidikan menengah terutama Sekolah Menengah Atas (SMA). Hal ini dibuktikan dengan terbentuknya dua profil klaster yang kontras, di mana klaster 1 (Wilayah Prima) mencatatkan rata-rata Angka Partisipasi Sekolah (APS) SMA mendekati 80%, sedangkan klaster 2 (Wilayah Atensi) tertinggal cukup jauh di angka 66,80%. Sementara itu, indikator partisipasi untuk jenjang SD dan SMP relatif telah mencapai saturasi yang merata di seluruh wilayah, menandakan keberhasilan program pendidikan dasar namun menyisakan pekerjaan rumah besar pada program Wajib Belajar 12 Tahun.

Secara geografis, temuan ini mengungkap pola yang tegas, di mana Provinsi Jawa Timur menunjukkan kinerja paling progresif dengan dominasi 68% wilayahnya berada di klaster 1, berbanding terbalik dengan Jawa Barat yang 75% wilayahnya dalam klaster 2. Selain ketimpangan antarprovinsi, penelitian ini juga menyoroti fenomena anomali pada Kota Surabaya. Meskipun berstatus sebagai ibu kota provinsi dan pusat ekonomi, Kota Surabaya justru masuk dalam kategori wilayah rentan bersama daerah tertinggal lainnya. Fakta ini mengindikasikan bahwa hambatan partisipasi di wilayah metropolitan tidak semata-mata disebabkan oleh ketersediaan fasilitas, melainkan didominasi oleh kompleksitas faktor sosial-ekonomi, seperti tingginya biaya peluang (*opportunity cost*) dan preferensi angkatan kerja muda untuk bekerja dibandingkan melanjutkan sekolah.

p-ISSN : 2406-7733
e-ISSN : 2597-9922

Implikasi praktis dari penelitian ini menuntut kebijakan pendidikan berupa strategi intervensi yang tersegregasi berbasis karakteristik klaster. Pemerintah perlu menerapkan kebijakan afirmatif ganda, yakni akselerasi pembangunan unit sekolah baru untuk wilayah pedesaan yang tertinggal khususnya di Jawa Barat, serta penguatan jaring pengaman sosial untuk mencegah anak putus sekolah di wilayah urban seperti Kota Surabaya. Dengan demikian, pemanfaatan metode data mining dalam studi ini terbukti efektif dalam mendekonstruksi data agregat menjadi wawasan granular yang krusial bagi perumusan kebijakan pembangunan manusia yang lebih presisi, efisien, dan berkeadilan.

Saran

Berdasarkan temuan penelitian ini, pemerintah daerah di wilayah yang masuk dalam klaster atensi, khususnya di Provinsi Jawa Barat, disarankan untuk memprioritaskan kebijakan afirmatif melalui akselerasi pembangunan unit sekolah baru dan perluasan akses fisik guna mengejar ketertinggalan angka partisipasi pada jenjang menengah. Sementara itu, untuk wilayah metropolitan seperti Kota Surabaya yang mengalami anomali, intervensi kebijakan tidak boleh lagi bertumpu pada pembangunan infrastruktur semata, melainkan harus menyasar pada penguatan jaring pengaman sosial dan pemberian insentif ekonomi bagi keluarga prasejahtera untuk menekan tingginya biaya peluang yang menyebabkan anak putus sekolah. Kementerian Pendidikan juga perlu mengkaji ulang efektivitas program Wajib Belajar 12 Tahun dengan menerapkan strategi intervensi yang berbeda antarwilayah, di mana fokus daerah rural ditekankan pada aksesibilitas dan daerah urban pada ketahanan ekonomi. Bagi peneliti selanjutnya, disarankan untuk memperdalam analisis melalui pendekatan kualitatif pada daerah-daerah anomali guna mengungkap faktor psikososial yang memengaruhi preferensi pendidikan, serta mempertimbangkan penggunaan variabel lain seperti kualitas tenaga pendidik dan ketersediaan fasilitas teknologi untuk mendapatkan gambaran pemetaan yang lebih komprehensif.

DAFTAR PUSTAKA

Arfan, U., & Pekei, Y. Evaluasi Distribusi Guru-Siswa dan Ketersediaan Sekolah untuk Mendukung Pembangunan Pendidikan Menggunakan K-Means Clustering. (2025). MALCOM: Indonesian Journal of Machine Learning and Computer Science. 5(4). 1508-1516. https://doi.org/10.47065/malcom.v5i4.4443

Basalamah, A. T., & Setyadi, R. Penerapan Algoritma K-Means Clustering Pada Tingkat Penyelesaian Pendidikan Di Provinsi Indonesia. (2023). DICOM (Jurnal Informatika dan Teknologi Komputer). 04(02). 114-121. https://doi.org/10.33059/jicom.v4i2.8105

Faradiba, F., & Nomleni, M. V. Analisis Indikator Pendidikan di Indonesia Periode 1994-2022. (2024). Jurnal Penelitian Inovatif (JUPIN). 4(1). 121-128. https://doi.org/10.54082/jupin.273

Iswahyudi, L. E., & Rahmawati, F. Segmentasi Wilayah melalui K-Means Cluster untuk Mencapai SDGs 5 “Kesetaraan Gender”. (2025). Jurnal Ilmu Sosial dan Humaniora. 14(2). 258-268. https://doi.org/10.23887/jish.v14i2.92647

Khadafi, A., Qomariyah, S., & Fajarwati, D. Perencanaan Wajib Belajar 12 Tahun. (2025). Katalis Pendidikan: Jurnal Ilmu Pendidikan dan Matematika. 2(1). 91-110. https://doi.org/10.62383/katalis.v2i1.1204

Mannan, A., Suparto., & Kusaeri. Practices and Challenges of the Validity of Exploratory Factor Analysis (EFA)-Based Assessment Instruments: A Systematic Literature Review 2020-2025. (2025). EDUKASIA: Jurnal Pendidikan dan Pembelajaran. 6(1). 611-626. https://doi.org/10.62775/edukasia.v6i1.1463

Musthofiyah, U., & Isnainiyah, L. What makes Indonesian Eastern students struggle on academic performance at urban college: A grounded theory. (2020). EnJourMe (English Journal of Merdeka): Culture, Language, and Teaching of English. 5(1). 16-30. https://doi.org/10.26905/enjourme.v5i1.4311

Putri, M. R., Nugraha, G. S., & Dwiyansaputra, R. Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan Menggunakan Metode K-Means Clustering. (2023). J-COSINE (Journal of Computer Science and Informatics Engineering). 7(1). 11-20. https://doi.org/10.29303/jcosine.v7i1.509

Robbani, K. R., & Fatah, Z. Klasterisasi Pendidikan SD untuk Mengetahui Daerah dengan Pendidikan Terendah menggunakan Algoritma K-Means. (2024). Jurnal Ilmiah Multidisiplin Ilmu (DENASYA). 1(5). 74-80. https://doi.org/10.69714/igmf7903

Sidauruk, J. M., Susilowati, M., & Akbar D, K. K. Indonesia's Struggle with Education Inequality: Is Reform the Answer?. (2025). Indonesia Discourse. 2(1). 59-84. https://doi.org/10.15294/indi.v2i1.23035

Manik, R. A., & Ariesta, A. Data Mining Untuk Klasterisasi Provinsi di Indonesia Berdasarkan Data Pokok Pendidikan Nasional. (2023). Jurnal TICOM: Technology of Information and Communication. 11(3). 159-164.

p-ISSN : 2406-7733
e-ISSN : 2597-9922

Nurahman, Purwanto, A., & Mulyanto, S. Klasterisasi Sekolah Menggunakan Algoritma K-Means berdasarkan Fasilitas, Pendidik, dan Tenaga Pendidik. (2022). Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer. 21(2). 337-350. https://doi.org/10.30812/matrik.v21i2.1411

Qori'atunnadyah, M. Pengelompokkan Wilayah Berdasarkan Rasio Guru-Murid Pada Jenjang Pendidikan Menggunakan Algoritma K-Means. (2023). Journal of Informatics Development (JID). 2(1). 1-12. https://doi.org/10.30741/jid.v1i1.898

Rais, Z., Annas, S., & Refaldy, M. Algoritma K-Prototype dalam Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat Tahun 2020. (2024). VARIANSI: Journal of Statistics and Its Application on Teaching and Research. 6(3). 144-151. https://doi.org/10.35580/variansiunm20

Raspatiningrum, L., Rubai'ah, S., Nugraha, R. S., Shodikin, R., & Gorni, R. L. Teacher Retention in Rural Indonesian Schools: An Interpretative Phenomenological Analysis of Career Disorientation and Commitment. (2025). Journal of Education and Teaching (JET). 6(3). 626-650. https://doi.org/10.51454/jet.v6i3.641

Silviana., Khoiron., & Rahmawati, S. D. Evaluasi Kebijakan Wajib Belajar Pendidikan Dasar 12 Tahun di Kecamatan Kedungkandang (Studi pada Dinas Pendidikan dan Kebudayaan Kota Malang). (2025). Jurnal Respon Publik. 19(4). 10-18.

Sumarah, J., & Wulandari, A. T. Analysis of the K-Means Algorithm for Clustering School Participation Rates in Central Java. (2023). KnE Social Sciences / ICASI Conference. 2023. 841–851. https://doi.org/10.18502/kss.v8i10.13454

Utomo, M. H. T. Clustering Data Siswa Putus Sekolah dengan Algoritma K-Means dan DBSCAN. (2023). Jurnal Komputer dan Informatika. 18(2). 150-159.

Yolanda, A. M., & Eriani, E. Optimalisasi Angka Partisipasi Sekolah Anak Usia Dini untuk Meningkatkan Indeks Pembangunan Manusia (Studi Kasus di Provinsi Nusa Tenggara Timur). (2022). JSTAR: Jurnal Statistika Terapan. 2(1). 20-35. https://doi.org/10.5300/JSTAR.V2I01.13F

.