



Pengelompokan Provinsi Berdasarkan Kelayakan Ruang Kelas dan Tenaga Kependidikan Sekolah Dasar Menggunakan Algoritma K-Means: Analisis Data Periode 2023-2024

Maidy Tri Wardani¹, Varla Octavia Ramadhani², Namira Anggreani³, Sumanto⁴, Andi Diah Kuswanto⁵

^{1,2,3,4,5} Ilmu Komputer, Fakultas Teknik Informatika, Universitas Bina Sarana Informatika

meyditriwardani@gmail.com, varlaoctavia05@gmail.com, anggreaninamira@gmail.com, sumanto@bsi.ac.id,
andi.ahk@bsi.ac.id

Abstract

This study analyzes the disparity in primary education quality across Indonesian regions using K-Means clustering methodology to group 38 provinces based on classroom adequacy indicators and teaching staff availability for the 2023-2024 period. Adopting the CRISP-DM methodology and utilizing datasets from the Ministry of Education, Culture, Research, and Technology, this research reveals significant educational gaps across Indonesian regions. The analysis results show three clusters reflecting educational conditions: Cluster 1 (low category) encompasses 23 provinces (67.6%) dominated by Eastern Indonesia regions such as Papua, Maluku, and other remote areas; Cluster 2 (medium category) consists of 13 provinces (38.2%) including South Sumatra, Riau, and DKI Jakarta; and Cluster 3 (high category) contains only 3 provinces: East Java, West Java, and Central Java. Clustering validity is confirmed through silhouette coefficient with the highest value of 0.795 for Cluster 2. These findings identify structural inequality between Java and outer Java regions, providing empirical foundation for implementing differentiated budget allocation policies with 150% above-normal funding for low-category provinces, targeted teacher recruitment incentives through 100% salary bonuses for remote areas, and establishment of center of excellence programs in high-performing provinces to mentor underperforming regions. The policy framework recommends immediate implementation of affirmative education funding, mandatory 10-year teaching contracts with scholarship programs for local teacher candidates, and sister school partnerships between Java-based and outer Java provinces to accelerate educational equity and sustainable national development.

Keywords: K-Means clustering, primary education, regional disparity, educational infrastructure, teaching personnel.

Abstrak

Penelitian ini menganalisis disparitas kualitas pendidikan dasar di seluruh wilayah Indonesia menggunakan metode *K-Means clustering* untuk mengelompokkan 38 provinsi berdasarkan indikator kelayakan ruang kelas dan ketersediaan tenaga kependidikan periode 2023-2024. Dengan mengadopsi metodologi *CRISP-DM* dan memanfaatkan *dataset* dari Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, penelitian ini mengungkap kesenjangan pendidikan yang signifikan antar wilayah di Indonesia. Hasil analisis menunjukkan tiga *cluster* yang mencerminkan kondisi pendidikan: *Cluster 1* (kategori rendah) mencakup 23 provinsi (67,6%) yang didominasi wilayah Indonesia Timur seperti Papua, Maluku, dan daerah terpencil lainnya; *Cluster 2* (kategori sedang) terdiri dari 13 provinsi (38,2%) termasuk Sumatera Selatan, Riau, dan DKI Jakarta; serta *Cluster 3* (kategori tinggi) hanya terdapat 3 provinsi yaitu Jawa Timur, Jawa Barat, dan Jawa Tengah. Validitas *clustering* dikonfirmasi melalui *silhouette coefficient* dengan nilai tertinggi 0.795 untuk *Cluster 2*. Temuan ini mengidentifikasi ketimpangan struktural antara Jawa dan luar Jawa, memberikan dasar empiris untuk implementasi kebijakan alokasi anggaran berdiferensiasi dengan dana afirmasi 150% di atas standar normal untuk provinsi kategori rendah, insentif rekrutmen guru melalui tunjangan 100% gaji pokok untuk daerah terpencil, dan pembentukan program *center of excellence* di provinsi berprestasi tinggi untuk mentoring wilayah tertinggal. Kerangka kebijakan merekomendasikan implementasi segera dana afirmasi pendidikan, kontrak mengajar wajib 10 tahun dengan program beasiswa calon guru lokal, dan kemitraan *sister school* antara provinsi berbasis Jawa dan luar Jawa untuk mempercepat pemerataan pendidikan dan pembangunan nasional berkelanjutan.

Kata kunci: K-Means Clustering, Pendidikan Dasar, Disparatis Regional, Infrastruktur Pendidikan, Tenaga Kependidikan.



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

1. Pendahuluan

Pemerintah Indonesia telah berkomitmen untuk meningkatkan kecerdasan bangsa melalui jaminan hak pendidikan bagi seluruh warga negara. Komitmen ini dituangkan dalam Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional yang mengamanatkan penyediaan akses pendidikan berkualitas secara merata untuk seluruh rakyat Indonesia. Meskipun demikian, realitas di lapangan menunjukkan bahwa masih terdapat banyak daerah terpencil yang belum menikmati standar pendidikan yang setara dengan perkotaan [1].

Pendidikan dapat didefinisikan sebagai proses pembelajaran sistematis yang mencakup pengembangan pengetahuan, keterampilan, dan nilai-nilai moral untuk mengoptimalkan potensi individu dalam aspek fisik, mental, dan spiritual. Dalam skala nasional, sistem pendidikan Indonesia menghadapi tantangan substansial dalam peningkatan kualitas yang masih tertinggal dibandingkan beberapa negara lain [2]. Pendidikan adalah sebuah rangkaian kegiatan yang dibutuhkan untuk mencapai keharmonisan dan kematangan dalam pertumbuhan pribadi seseorang serta komunitas secara keseluruhan [3].

Dalam ilmu pendidikan, perlu dipahami perbedaan antara pedagogi dan pedagogik. Pedagogi merupakan praktik "mendidik" itu sendiri, sementara pedagogik adalah "ilmu yang mempelajari pendidikan". Istilah pedagogi berasal dari kata pedagogos yang bermakna membimbing anak untuk mencapai kemandirian dan tanggung jawab. Aktivitas mendidik mencakup seluruh aspek perkembangan manusia, mulai dari kesehatan fisik, kemampuan kognitif, emosional, kehidupan sosial, hingga perkembangan spiritual. Pendidikan merupakan kebutuhan fundamental bagi setiap individu [4].

Disparitas kualitas pendidikan antarprovinsi dan wilayah terpencil menunjukkan ketimpangan dalam implementasi sistem pendidikan yang memadai. Permasalahan utama meliputi rendahnya standar mutu pendidikan nasional, keterbatasan kompetensi dan kualifikasi tenaga pendidik, hambatan aksesibilitas pendidikan karena faktor ekonomi, serta implementasi profesionalisme guru yang belum optimal. Faktor-faktor tersebut menjadikan kualitas sistem pendidikan Indonesia sebagai indikator penting untuk mengukur tingkat kemajuan bangsa [5].

Kondisi infrastruktur pendidikan Indonesia saat ini menunjukkan tren yang mengkhawatirkan. Menurut [6] dalam artikel yang berjudul "Ruang Kelas Rusak Bertambah BRIN Riset Jumlah Ruang Kelas di Indonesia yang Rusak Berdasarkan Data Kemendikdasmen 2022-2024", kondisi ruang kelas di jenjang Sekolah Dasar mengalami deteriorasi dalam tiga tahun terakhir (2022-2024). Yuyun Libriyanti, peneliti

dari Pusat Riset Pendidikan Badan Riset dan Inovasi Nasional (BRIN), dalam diskusi daring memaparkan bahwa jumlah kelas dengan kerusakan sedang mengalami peningkatan dari 193.720 unit pada tahun 2022 menjadi 261.968 unit atau meningkat 35% pada tahun 2024. Demikian pula dengan kelas yang mengalami kerusakan berat, meningkat dari 97.471 unit pada 2022 menjadi 127.108 unit pada 2024.

Yuyun menekankan bahwa kondisi infrastruktur sekolah saat ini mengindikasikan perlunya pergeseran fokus pembangunan pendidikan dari penambahan kuantitas ruang kelas menuju peningkatan kualitas fasilitas. Untuk mengatasi hal ini, Yuyun merekomendasikan implementasi program rehabilitasi sekolah yang komprehensif dan terintegrasi. Pendekatan holistik ini mencakup tiga dimensi: perbaikan aspek teknis infrastruktur bangunan, penyesuaian dengan kebijakan pendidikan nasional, dan keterlibatan aktif masyarakat lokal sebagai pemangku kepentingan. Program rehabilitasi sekolah dipandang sebagai investasi jangka panjang untuk membangun ekosistem pembelajaran yang memenuhi empat kriteria esensial: keamanan struktural, ramah anak, inklusif bagi semua peserta didik, serta mampu menghadapi tantangan eksternal seperti bencana alam dan dampak perubahan iklim. Melalui pendekatan ini, rehabilitasi sekolah diharapkan dapat menjadi fondasi kuat untuk transformasi pendidikan di Indonesia [6].

Untuk menganalisis kompleksitas permasalahan pendidikan Indonesia, diperlukan pendekatan yang sistematis dan berbasis data. Algoritma K-Means clustering telah menunjukkan efektivitas yang terbukti dalam berbagai penelitian analisis data di Indonesia, memperlihatkan konsistensi dan superioritas dalam pengelompokan data multidimensional. Penelitian [7] membuktikan keunggulan K-Means dibandingkan Fuzzy C-Means dalam pengelompokan daerah penyebaran Covid-19, dengan K-Means menghasilkan akurasi yang lebih tinggi dalam menangani data berdimensi besar. Sementara itu, penelitian [8] mengonfirmasi efektivitas yang sama dalam konteks analisis risiko bencana, dengan implementasi K-Means menggunakan model CRISP-DM untuk identifikasi daerah rawan kebakaran yang menghasilkan struktur cluster berkualitas tinggi (strong structure). Kedua penelitian ini mendemonstrasikan fleksibilitas dan reliabilitas algoritma K-Means yang dapat diadaptasi secara optimal untuk analisis kesehatan masyarakat maupun manajemen risiko bencana, menegaskan posisinya sebagai metode clustering yang robust dan versatile untuk berbagai domain aplikasi di Indonesia.

Penelitian ini mengadopsi metode analisis cluster dengan algoritma K-Means Clustering untuk menganalisis data infrastruktur pendidikan di Indonesia. K-Means merupakan algoritma pengelompokan data

dimana konstanta K merepresentasikan jumlah cluster yang diinginkan, sedangkan "Means" merujuk pada nilai rata-rata dalam kelompok data yang ditetapkan sebagai pusat cluster. Metode ini termasuk dalam kategori clustering sebagai teknik analisis data atau data mining yang melakukan pemodelan tanpa pengawasan menggunakan sistem partisi untuk pengelompokan data. Tujuan utama algoritma K-Means adalah membagi data yang ada menjadi beberapa kelompok yang homogen [9].

Clustering atau analisis klaster merupakan teknik Unsupervised Classification yang digunakan untuk mengelompokkan data tanpa label berdasarkan kesamaan karakteristik tertentu [10] [11]. Metode K-Means dipilih karena kemampuannya dalam mengklasifikasikan data ke dalam kelompok-kelompok homogen dengan efisiensi komputasi yang tinggi. Data mining merupakan teknik analisis data yang bertujuan mengidentifikasi pola-pola tersembunyi dalam kumpulan data besar. Metode ini memungkinkan transformasi data mentah menjadi pengetahuan baru yang berguna untuk pengambilan keputusan strategis di masa depan. Proses data mining melibatkan pemeriksaan data dari berbagai perspektif untuk menghasilkan informasi bernilai yang dapat meningkatkan efektivitas pengambilan keputusan. Secara praktis, data mining berfungsi sebagai alat untuk mengidentifikasi hubungan dan pola dalam dataset yang sangat besar dengan ratusan hingga ribuan variabel. Berdasarkan uraian tersebut, data mining dapat didefinisikan sebagai proses ekstraksi informasi dari dataset besar dan kompleks untuk menemukan pola-pola yang tidak terlihat secara langsung [12].

Berdasarkan uraian di atas, penelitian ini bertujuan untuk mengelompokkan dan menganalisis kelayakan ruang kelas serta kualitas tenaga kependidikan di setiap provinsi di Indonesia menggunakan metode K-Means clustering. Dengan demikian, penelitian ini dapat memberikan gambaran komprehensif mengenai kesenjangan kondisi pendidikan antarwilayah dan menjadi dasar untuk pengambilan kebijakan pemerataan pendidikan nasional

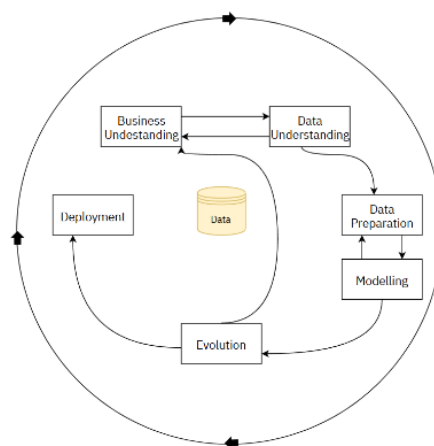
2. Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah *CRISP-DM*. *CRISP-DM* merupakan metodologi yang menjelaskan langkah-langkah sistematis untuk melakukan penambangan data secara menyeluruh. Meskipun penelitian ini mengadopsi metode *CRISP-DM*, namun dalam implementasinya menghasilkan model yang lebih spesifik dan disesuaikan dengan kebutuhan penelitian [12]. Dalam fase ini, metodologi yang diterapkan adalah Cross Industry Standard for Data Mining (*CRISP-DM*). *CRISP-DM* merupakan kerangka kerja yang mengadopsi model proses pengembangan

data yang sering dimanfaatkan oleh praktisi untuk menyelesaikan berbagai permasalahan [13].

Pemilihan algoritma K-Means dalam penelitian ini didasarkan pada beberapa pertimbangan metodologis yang sistematis. K-Means dipilih karena memiliki efisiensi komputasi yang tinggi dengan kompleksitas waktu $O(n \times k \times i)$, dimana n adalah jumlah data, k adalah jumlah cluster, dan i adalah jumlah iterasi, sehingga sangat cocok untuk mengolah dataset berukuran besar. Selain itu, algoritma ini memiliki kesederhanaan dalam implementasi dengan langkah-langkah yang jelas dan deterministik, serta menghasilkan output yang mudah diinterpretasi karena setiap cluster memiliki centroid yang jelas sebagai representasi karakteristik cluster tersebut. K-Means juga sesuai dengan tujuan penelitian untuk mengidentifikasi pola dalam data numerik dengan asumsi cluster berbentuk spherical dan memiliki ukuran yang relatif seimbang.

Model untuk mengidentifikasi pola dalam penelitian ini terbagi menjadi 5 tahapan utama. Tahapan pertama adalah memahami masalah bisnis (*Business Understanding*), kemudian memahami karakteristik data yang akan digunakan (*Data Understanding*), selanjutnya menyiapkan dan mengolah data (*Data Preparation*), membangun model (*Modelling*), dan terakhir mengevaluasi hasil serta mengimplementasikannya (*Evaluation* dan *Deployment*). Setiap tahapan ini saling berkaitan dan memiliki peran penting untuk menghasilkan analisis data yang berkualitas dan dapat diaplikasikan dalam praktik nyata [12].



Gambar 1. Model CRISP-DM

Business Understanding

Dalam fase ini, terdapat sejumlah aktivitas penting yang perlu dijalankan, antara lain: mengenali kebutuhan dan sasaran dari perspektif organisasi, mentransformasikan pemahaman tersebut menjadi rumusan permasalahan dalam konteks penambangan data, serta menyusun rancangan dan pendekatan strategis untuk meraih target yang ditetapkan dalam proyek data mining [14].

menjadi dasar pertimbangan dalam perumusan kebijakan alokasi sumber daya pendidikan yang lebih tepat sasaran dan berkeadilan. Hasil pengelompokan akan divisualisasikan dalam bentuk peta tematik dan diagram interpretif untuk memudahkan analisis spasial distribusi kualitas pendidikan di Indonesia. Gambar di atas menampilkan dataset penelitian yang diperoleh dari Portal Data Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia. Dataset ini menyajikan berbagai variabel secara menyeluruh yang mempengaruhi kualitas sistem pendidikan Indonesia. Data mencakup beragam indikator kunci, antara lain: populasi peserta didik, tingkat pengulangan kelas (*grade repetition*), tingkat putus sekolah (*dropout rate*), jumlah tenaga kepala sekolah dan pendidik, kuantitas kelompok belajar (rombongan belajar), serta kondisi infrastruktur ruang pembelajaran.

sudah tersedia secara utuh tanpa adanya nilai kosong maupun data yang tidak konsisten. Data sudah dipastikan tidak ada lagi data yang kosong sehingga dapat dilihat pada gambar.

Provinsi	Jumlah Siswa	Tingkat Pengulangan
1. Aceh	23286	0.2247
2. Sumatra Utara	23286	0.2247
3. Sumatra Barat	23286	0.2247
4. Riau	23286	0.2247
5. Kepulauan Riau	23286	0.2247
6. Jawa Barat	23286	0.2247
7. Jawa Tengah	23286	0.2247
8. Jawa Timur	23286	0.2247
9. Kalimantan Barat	23286	0.2247
10. Kalimantan Tengah	23286	0.2247
11. Kalimantan Selatan	23286	0.2247
12. Kalimantan Timur	23286	0.2247
13. Sulawesi Utara	23286	0.2247
14. Sulawesi Tengah	23286	0.2247
15. Sulawesi Selatan	23286	0.2247
16. Sulawesi Tenggara	23286	0.2247
17. Maluku	23286	0.2247
18. Maluku Utara	23286	0.2247
19. Papua	23286	0.2247
20. Papua Barat	23286	0.2247
21. Nusa Tenggara Barat	23286	0.2247
22. Nusa Tenggara Timur	23286	0.2247
23. Bali	23286	0.2247
24. DI Yogyakarta	23286	0.2247
25. DKI Jakarta	23286	0.2247
26. Banten	23286	0.2247
27. Lampung	23286	0.2247
28. Bengkulu	23286	0.2247
29. Kepulauan Bangka Belitung	23286	0.2247
30. Gorontalo	23286	0.2247
31. Sulawesi Barat	23286	0.2247
32. Kalimantan Utara	23286	0.2247

Gambar 4. Preprocessing Data

Cakupan data meliputi seluruh 38 provinsi di Indonesia beserta institusi pendidikan Indonesia yang berada dalam pengawasan Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. Dataset ini memberikan gambaran menyeluruh tentang kondisi pendidikan dasar di Indonesia pada periode 2023-2024, yang menjadi dasar analisis clustering untuk pengelompokan provinsi berdasarkan karakteristik pendidikan.

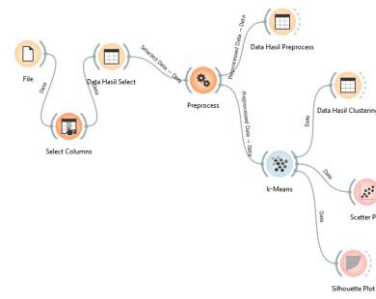
c. Penerapan Metode K-Means Clustering

Pada Tahap ini dilakukan penerapan perhitungan *K-Means Clustering* dengan 40 jumlah data, kemudian dilanjutkan proses memasukan data kedalam *Tools Orange* untuk mendapatkan hasil *cluster*.

a. Data Selection

Provinsi	Jumlah Siswa	Tingkat Pengulangan
1. Aceh	23286	0.2247
2. Sumatra Utara	23286	0.2247
3. Sumatra Barat	23286	0.2247
4. Riau	23286	0.2247
5. Kepulauan Riau	23286	0.2247
6. Jawa Barat	23286	0.2247
7. Jawa Tengah	23286	0.2247
8. Jawa Timur	23286	0.2247
9. Kalimantan Barat	23286	0.2247
10. Kalimantan Tengah	23286	0.2247
11. Kalimantan Selatan	23286	0.2247
12. Kalimantan Timur	23286	0.2247
13. Sulawesi Utara	23286	0.2247
14. Sulawesi Tengah	23286	0.2247
15. Sulawesi Selatan	23286	0.2247
16. Sulawesi Tenggara	23286	0.2247
17. Maluku	23286	0.2247
18. Maluku Utara	23286	0.2247
19. Papua	23286	0.2247
20. Papua Barat	23286	0.2247
21. Nusa Tenggara Barat	23286	0.2247
22. Nusa Tenggara Timur	23286	0.2247
23. Bali	23286	0.2247
24. DI Yogyakarta	23286	0.2247
25. DKI Jakarta	23286	0.2247
26. Banten	23286	0.2247
27. Lampung	23286	0.2247
28. Bengkulu	23286	0.2247
29. Kepulauan Bangka Belitung	23286	0.2247
30. Gorontalo	23286	0.2247
31. Sulawesi Barat	23286	0.2247
32. Kalimantan Utara	23286	0.2247

Gambar 3. Data Selection



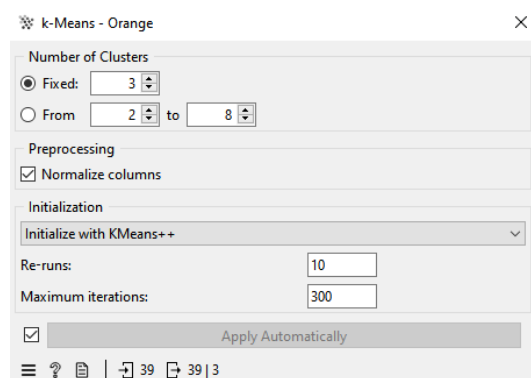
Gambar 5. Proses Metode K-Means Clustering

Pada tahap awal dilakukan seleksi atribut menggunakan *Select Columns* untuk menentukan variabel yang akan digunakan dalam proses pengelompokan. Dalam penelitian ini, hanya 3 atribut utama yang dipilih sebagai input clustering, yaitu: provinsi sebagai identifier wilayah, tenaga kependidikan sebagai indikator sumber daya manusia, dan kelayakan ruang kelas sebagai indikator infrastruktur pendidikan.

K-Means Clustering merupakan algoritma yang digunakan untuk mengelompokkan data ke dalam cluster berdasarkan karakteristik yang serupa. Dalam penelitian ini nilai K ditentukan Sebanyak 3 cluster yang dapat dilihat pada gambar.

b. Data Processing

Pada tahap *preprocessing* data, umumnya diperlukan langkah-langkah pembersihan data seperti proses cleaning yang mencakup pengelolaan nilai yang hilang. Akan tetapi, untuk dataset yang digunakan dalam penelitian ini, seluruh data



Gambar 6. Proses Penentuan Cluster

Provinsi	Cluster	Silhouette	Tenaga Kependidikan (SM)	
Prov. Aceh	C1	0,79386	0,72587	
Prov. Sumatera Utara	C1	0,79375	1,29943	1,88778
Prov. Sumatera Tengah	C1	0,79375	1,66174	1,74574
Prov. Sumatera Barat	C1	0,79375	0,28736	0,28242
Prov. Riau	C1	0,79375	1,39183	0,93584
Prov. Jambi	C1	0,79375	0,81887	0,97308
Prov. Sumatera Selatan	C1	0,79375	1,42804	1,28990
Prov. Lampung	C1	0,79375	0,61171	0,51882
Prov. Kalimantan Barat	C1	0,65242	0,42942	0,62217
Prov. Kalimantan Tengah	C1	0,65242	0,29982	0,68771
Prov. Kalimantan Selatan	C1	0,65242	0,68708	0,62114
Prov. Kalimantan Timur	C1	0,65242	0,68708	0,62114
Prov. Sulawesi Utara	C1	0,65242	0,68708	0,62114
Prov. Sulawesi Tengah	C1	0,65242	0,68708	0,62114
Prov. Sulawesi Selatan	C1	0,65242	0,68708	0,62114
Prov. Sulawesi Tenggara	C1	0,65242	0,68708	0,62114
Prov. Maluku	C1	0,65242	0,68708	0,62114
Prov. Maluku Utara	C1	0,65242	0,68708	0,62114
Prov. Papua	C1	0,65242	0,68708	0,62114
Prov. Papua Barat	C1	0,65242	0,68708	0,62114
Prov. Nusa Tenggara Barat	C1	0,65242	0,68708	0,62114
Prov. Nusa Tenggara Timur	C1	0,65242	0,68708	0,62114
Prov. Bali	C1	0,65242	0,68708	0,62114
Prov. Jawa Barat	C1	0,65242	0,68708	0,62114
Prov. Jawa Tengah	C1	0,65242	0,68708	0,62114
Prov. Jawa Timur	C1	0,65242	0,68708	0,62114
Prov. DI Yogyakarta	C1	0,65242	0,68708	0,62114
Prov. Aceh	C2	0,79386	0,72587	
Prov. Sumatera Utara	C2	0,79375	1,29943	
Prov. Sumatera Tengah	C2	0,79375	1,66174	
Prov. Sumatera Barat	C2	0,79375	0,28736	
Prov. Riau	C2	0,79375	1,39183	
Prov. Jambi	C2	0,79375	0,81887	
Prov. Sumatera Selatan	C2	0,79375	1,42804	
Prov. Lampung	C2	0,79375	0,61171	
Prov. Kalimantan Barat	C2	0,65242	0,42942	
Prov. Kalimantan Tengah	C2	0,65242	0,29982	
Prov. Kalimantan Selatan	C2	0,65242	0,68708	
Prov. Kalimantan Timur	C2	0,65242	0,68708	
Prov. Sulawesi Utara	C2	0,65242	0,68708	
Prov. Sulawesi Tengah	C2	0,65242	0,68708	
Prov. Sulawesi Selatan	C2	0,65242	0,68708	
Prov. Sulawesi Tenggara	C2	0,65242	0,68708	
Prov. Maluku	C2	0,65242	0,68708	
Prov. Maluku Utara	C2	0,65242	0,68708	
Prov. Papua	C2	0,65242	0,68708	
Prov. Papua Barat	C2	0,65242	0,68708	
Prov. Nusa Tenggara Barat	C2	0,65242	0,68708	
Prov. Nusa Tenggara Timur	C2	0,65242	0,68708	
Prov. Bali	C2	0,65242	0,68708	
Prov. Jawa Barat	C2	0,65242	0,68708	
Prov. Jawa Tengah	C2	0,65242	0,68708	
Prov. Jawa Timur	C2	0,65242	0,68708	
Prov. DI Yogyakarta	C2	0,65242	0,68708	
Prov. Aceh	C3	0,79386	0,72587	
Prov. Sumatera Utara	C3	0,79375	1,29943	
Prov. Sumatera Tengah	C3	0,79375	1,66174	
Prov. Sumatera Barat	C3	0,79375	0,28736	
Prov. Riau	C3	0,79375	1,39183	
Prov. Jambi	C3	0,79375	0,81887	
Prov. Sumatera Selatan	C3	0,79375	1,42804	
Prov. Lampung	C3	0,79375	0,61171	
Prov. Kalimantan Barat	C3	0,65242	0,42942	
Prov. Kalimantan Tengah	C3	0,65242	0,29982	
Prov. Kalimantan Selatan	C3	0,65242	0,68708	
Prov. Kalimantan Timur	C3	0,65242	0,68708	
Prov. Sulawesi Utara	C3	0,65242	0,68708	
Prov. Sulawesi Tengah	C3	0,65242	0,68708	
Prov. Sulawesi Selatan	C3	0,65242	0,68708	
Prov. Sulawesi Tenggara	C3	0,65242	0,68708	
Prov. Maluku	C3	0,65242	0,68708	
Prov. Maluku Utara	C3	0,65242	0,68708	
Prov. Papua	C3	0,65242	0,68708	
Prov. Papua Barat	C3	0,65242	0,68708	
Prov. Nusa Tenggara Barat	C3	0,65242	0,68708	
Prov. Nusa Tenggara Timur	C3	0,65242	0,68708	
Prov. Bali	C3	0,65242	0,68708	
Prov. Jawa Barat	C3	0,65242	0,68708	
Prov. Jawa Tengah	C3	0,65242	0,68708	
Prov. Jawa Timur	C3	0,65242	0,68708	
Prov. DI Yogyakarta	C3	0,65242	0,68708	

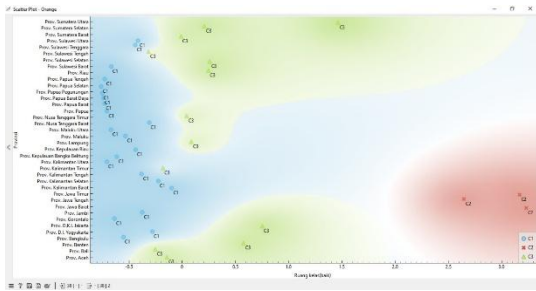
Gambar 7. Hasil Pembagian Cluster

Gambar diatas merupakan Hasil cluster dari ruang kelas dan tenaga kependidikan di setiap provinsi di Indonesia, tahun 2023-2024 yang telah di proses di software orange *data mining* yang terdiri dari cluster C1, C2, C3.

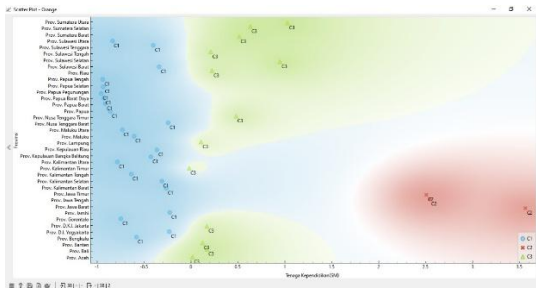
d. Evaluasi Hasil

Pada tahap evaluasi merupakan bagian penting dalam proses *data mining* dengan pendekatan *CRISP-DM*. Tujuan utamanya adalah menilai kualitas hasil klusterisasi serta keterkaitannya dengan tujuan analisis, yaitu mengelompokkan provinsi di Indonesia berdasarkan kelayakan ruang kelas dan ketersediaan tenaga pendidik di jenjang Sekolah Dasar.

Pengujian dilakukan menggunakan dua pendekatan utama, yaitu visualisasi hasil *clustering* melalui *scatter plot* dan pengukuran validitas *cluster* menggunakan *silhouette coefficient*.



Gambar 8. Hasil Scatter Plot Ruang Kelas (Baik)



Gambar 9. Hasil Scatter Plot Tenaga Kependidikan

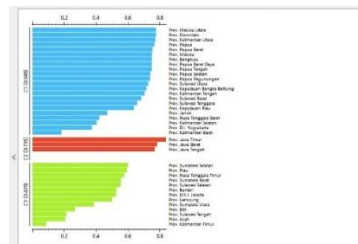
Visualisasi *Scatter plot* untuk melihat bagaimana provinsi – provinsi tersebar pada tiap kluster terhadap dua variabel utama yaitu:

1. Tenaga Kependidikan (SM) : digunakan untuk melihat ketersediaan SDM pendidikan
2. Ruang Kelas (Baik) : Digunakan untuk menilai kondisi infrastruktur fisik sekolah.

Dalam *scatter plot* gambar 8 dan gambar 9 setiap provinsi direpresentasikan dengan simbol:

1. Warna biru (C1) menunjukkan provinsi dengan kondisi menengah, baik dari sisi jumlah guru maupun kualitas ruang kelas.
2. Warna merah (C2) merepresentasikan provinsi dengan kondisi terbaik, di mana tenaga kependidikan dan fasilitas ruang belajar tergolong sangat baik.
3. Warna hijau (C3) menunjukkan provinsi dengan kondisi terendah, yang masih menghadapi tantangan besar dalam hal infrastruktur dan ketersediaan guru.

e. Evaluasi Validitas Cluster (Silhouette Plot)



Gambar 10. Silhouette Plot

Setelah hasil klusterisasi divisualisasikan, langkah berikutnya adalah melakukan evaluasi secara kuantitatif menggunakan *Silhouette plot*. Metrik ini membantu menilai seberapa baik setiap provinsi masuk ke dalam klusternya masing-masing, yaitu apakah provinsi tersebut benar-benar mirip dengan anggota lain dalam satu kluster (kohesi), dan seberapa berbeda dari kluster lainnya (separasi).

Nilai *Silhouette* berada pada rentang -1 hingga 1, dengan interpretasi sebagai berikut:

1. Mendekati 1.0: kluster sangat kuat dan jelas terpisah—hasil clustering sangat baik.
2. Sekitar 0.5–0.7: hasil masih tergolong baik dan bisa diterima.
3. Di bawah 0.4: kluster kurang baik, kemungkinan terjadi tumpang tindih antar kelompok.

Berdasarkan visualisasi *silhouette plot*:

1. *Cluster C2* mencatat nilai rata-rata tertinggi yaitu 0.795, yang berarti provinsi dalam kelompok ini sangat cocok dan konsisten dengan karakteristik klasternya.
2. *Cluster C3* memperoleh nilai 0.647, yang masih dalam kategori baik, menunjukkan bahwa provinsi-provinsi dalam klaster ini cukup seragam.
3. *Cluster C1* memiliki skor rata-rata 0.440, yang tergolong rendah. Hal ini menandakan bahwa sebagian provinsi dalam kelompok ini mungkin memiliki kemiripan dengan klaster lain atau bahwa perbedaan dalam klaster ini tidak begitu jelas.

Berikut adalah tabel hasil analisis cluster berdasarkan provinsi:

Tabel 1. Hasil Cluster

No	Nama Provinsi
1	Sumatra Utara
2	Gorontalo
3	Kalimantan Utara
4	Papua
5	Papua Barat
6	Papua Barat Daya
7	Papua Tengah
8	Papua Selatan
9	Papua Pengunungan
10	Maluku
11	Bengkulu
12	Sulawesi Utara
13	Kepulauan Bangka Belitung
14	Kalimantan Tengah
15	Sulawesi Barat
16	Sulawesi Tenggara
17	Kepulauan Riau
18	Jambi
19	Nusa Tenggara Barat
20	Kalimantan Selatan
21	D.I. Yogyakarta
22	Kalimantan Barat

Tabel 2. Cluster 2 (Hijau) Kategori Rendah

No	Nama Provinsi
1	Sumatera Selatan

2	Riau
3	Nusa Tenggara Timur
4	Sumatera Barat
5	Sulawesi Selatan
6	Banten
7	DKI Jakarta
8	Lampung
9	Sumatera Utara
10	Bali
11	Sulawesi Tengah
12	Aceh
13	Kalimantan Timur

Tabel 3. Cluster 1 (Biru) Kategori Sedang

Cluster	Warna	Jumlah Provinsi	Karakteristik Posisi
Cluster 2	Hijau	23 Provinsi	Rendah pada kedua dimensi
Cluster 1	Biru	13 Provinsi	Rendah pada kedua dimensi
Cluster 2	Merah	3 Provinsi	Tinggi pada kedua dimensi

Tabel 3. Cluster 2 (Merah) Kategori Tinggi

NO	Nama Provinsi
1	Jawa Timur
2	Jawa Barat
3	Jawa tengah

Klasterisasi 38 provinsi di Indonesia berdasarkan kelayakan ruang kelas dan rasio tenaga kependidikan menghasilkan tiga kelompok dengan karakteristik berbeda:

1. Cluster C2 - Kinerja Unggul Provinsi dengan proporsi ruang kelas layak tinggi dan rasio guru-siswa optimal. Memiliki infrastruktur pendidikan yang memadai dan kondisi terbaik secara keseluruhan.
2. Cluster C1 - Kinerja Menengah Provinsi dengan kondisi pendidikan cukup memadai namun masih memerlukan peningkatan. Memiliki potensi pengembangan yang baik untuk infrastruktur dan tenaga pendidik.
3. Cluster C3 - Tantangan Tinggi Provinsi dengan keterbatasan signifikan dalam fasilitas pendidikan layak dan rasio tenaga pendidik yang belum optimal. Membutuhkan prioritas intervensi.

Rekomendasi Kebijakan

Cluster C3 (Prioritas Utama):

- a. Program pembangunan dan rehabilitasi ruang kelas
- b. Penambahan kuota dan redistribusi tenaga pendidik
- c. Alokasi anggaran prioritas

Cluster C1 (Prioritas Menengah):

- a. Peningkatan kualitas infrastruktur existing
- b. Optimalisasi distribusi tenaga pendidik
- c. Penguatan kapasitas kelembagaan

Cluster C2 (Pemeliharaan):

- a. Mempertahankan standar kualitas
- b. Pengembangan program unggulan
- c. Berbagi praktik baik dengan provinsi lain

Hasil klusterisasi ini memberikan dasar empiris bagi pemerintah pusat untuk alokasi anggaran yang tepat sasaran, membantu dinas pendidikan daerah mengidentifikasi area prioritas, dan mendukung lembaga kebijakan dalam penyusunan strategi pendidikan yang lebih efektif dan responsif terhadap kondisi regional.

4. Kesimpulan

Penelitian ini berhasil mengidentifikasi pola distribusi kualitas pendidikan dasar di Indonesia melalui analisis klusterisasi menggunakan algoritma K-Means dengan dua indikator kunci yaitu kelayakan ruang kelas dan ketersediaan tenaga kependidikan. Hasil analisis menunjukkan bahwa 34 provinsi di Indonesia terbagi menjadi tiga kelompok yang mencerminkan disparitas kondisi pendidikan yang signifikan antar wilayah, dengan mayoritas provinsi (23 provinsi atau 67,6%) berada dalam kategori rendah, 13 provinsi (38,2%) dalam kategori sedang, dan hanya 3 provinsi dalam kategori tinggi.

Temuan penelitian mengungkapkan kesenjangan pendidikan struktural antara Jawa dan luar Jawa, dimana kualitas pendidikan terbaik hanya terdapat pada tiga provinsi Jawa yaitu Jawa Timur, Jawa Barat, dan Jawa Tengah. Mayoritas provinsi dengan kondisi pendidikan rendah didominasi oleh wilayah Indonesia Timur seperti Papua dan Maluku yang menghadapi tantangan serius dalam infrastruktur pendidikan dan ketersediaan tenaga pendidik. Validitas hasil klusterisasi dikonfirmasi melalui evaluasi silhouette coefficient dengan nilai tertinggi pada Cluster 2 (0.795) dan Cluster 3 (0.647), sedangkan Cluster 1 memperoleh nilai 0.440 yang masih dapat diterima.

Provinsi kategori rendah memerlukan alokasi anggaran khusus 150% dari standar normal dengan fokus pembangunan ruang kelas baru, rekrutmen tenaga pendidik melalui skema insentif tunjangan daerah terpencil 100% gaji pokok, dan program beasiswa calon guru dengan ikatan dinas 10 tahun. Provinsi kategori sedang perlu program optimalisasi melalui hibah peningkatan kualitas dengan alokasi 110-120% standar

nasional untuk pelatihan guru berkelanjutan dan modernisasi sarana pembelajaran. Provinsi kategori tinggi mendapat dukungan program center of excellence sebagai model rujukan nasional dan implementasi sister school untuk mentoring provinsi kategori rendah.

Dengan demikian, analisis klusterisasi ini menjadi instrumen strategis dalam upaya pemerataan kualitas pendidikan nasional yang didukung roadmap implementasi jelas, target terukur, dan komitmen anggaran jangka panjang untuk memastikan sustainability program peningkatan kualitas pendidikan di seluruh Indonesia.

Daftar Rujukan

- [1] S. Indah Elisabet Tambun, G. Sirait, dan J. Simamora, "TENTANG SISTEM PENDIDIKAN NASIONAL MENCAKUP BAB IV PASAL 5 MENGENAI HAK DAN KEWAJIBAN WARGA NEGARA, ORANG TUA DAN PEMERINTAH," 2003.
- [2] A. Rahman, S. Mnandar Asri, A. Fitriani, Y. Karlina, dan Yumriani, "7757-24249-1-PB," *Kajian Pendidikan Islam*, vol. 2, hlm. 1-8, Jun 2022.
- [3] N. Doktor dkk., "PENDIDIKAN DALAM UPAYA MEMAJUKAN TEKNOLOGI," 2013.
- [4] F. Nur dan A. Kurniawati, "MENINJAU PERMASALAHAN RENDAHNYA KUALITAS PENDIDIKAN DI INDONESIA DAN SOLUSI," 2022.
- [5] Safiq Maulido, Popi Karmijah, dan Vinanda Rahmi, "Upaya Meningkatkan Pendidikan Masyarakat Di Daerah Terpencil," *Jurnal Sadewa : Publikasi Ilmu Pendidikan, pembelajaran dan Ilmu Sosial*, vol. 2, no. 1, hlm. 198-208, Des 2023, doi: 10.61132/sadewa.v2i1.488.
- [6] P. A. Mawardi Upi, "Ruang Kelas Rusak Bertambah BRIN riset jumlah ruang kelas di Indonesia yang rusak berdasarkan data Kemendikdasmen 2022-2024.," <https://validnews.id/nasional/ruang-kelas-rusak-bertambah>.
- [7] M. R. Putri, G. Satya Nugraha, dan R. Dwiyanaputra, "Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan Menggunakan Metode K-Means Clustering Grouping Provinces in Indonesia Based on Education Indicators Using the K-Means Clustering Method," Jun 2023. [Daring]. Tersedia pada: <http://jcosine.if.unram.ac.id/>
- [8] F. N. Dhewayani dkk., "Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM," *Jurnal Teknologi dan Informatika*, vol. 12, hlm. 64-77, Mar 2022, doi: 10.34010/jati.v12i1.
- [9] T. Amalina, D. Bima, A. Pramana, dan B. N. Sari, "Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food," *Jurnal Ilmiah Wahana Pendidikan*, vol. 8, no. 15, hlm. 574-583, 2022, doi: 10.5281/zenodo.7052276.
- [10] A. Az-Zahra dan A. W. Wijayanto, "Tinjauan Kesejahteraan di Daerah Perbatasan Republik Indonesia Tahun 2021: Penerapan Analisis Kluster K-Means dan Hierarki," *Jurnal Sistem dan Teknologi Informasi (JustIN)*, vol. 12, no. 1, hlm. 55, Jan 2024, doi: 10.26418/justin.v12i1.69040.
- [11] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *Jurnal Ilmiah Informatika dan Ilmu Komputer (JIMA-ILKOM)*, vol. 3, no. 1, hlm. 46-56, Mar 2024, doi: 10.58602/jima-ilkom.v3i1.26.
- [12] I. Budiman, T. Prahasto, dan Y. Christyono, "DATA CLUSTERING MENGGUNAKAN METODOLOGI CRISP-DM UNTUK PENGENALAN POLA PROPORSI PELAKSANAAN TRIDHARMA," 2012.
- [13] D. Astuti, A. Rahmat Iskandar, dan A. Febrianti, "Penentuan Strategi Promosi Usaha Mikro Kecil Dan Menengah

- [14] (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering,” vol. 1, no. 2, hlm. 60–072, 2019, doi: 10.20895/INISTA.V1I2.
- [15] Y. Pitaloka Anggriani *dkk.*, “IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING DALAM MENENTUKAN BLOK TANAMAN SAWIT PRODUKTIF PADA PT ARTA PRIGEL,” 2024. Pastika Bening Puan, “Dataset Pendidikan SD Indonesia Tahun 2023-2024.”