

# Pemetaan Ketimpangan Capaian Pendidikan Antarprovinsi di Indonesia Menggunakan K-Means Clustering dengan Validasi Davies-Bouldin dan Silhouette Score

Moh Ali Ridla\*<sup>1</sup>, Naurah Nadhifah<sup>2</sup>, Melita Nur Azizah<sup>3</sup>

<sup>1,3</sup>Prodi Sistem Informasi, Universitas Ibrahimy, Situbondo

<sup>2</sup>Fakultas Sains & Teknologi, Universitas Ibrahimy, Situbondo

el.ridla@gmail.com, naurahnadhifah705@gmail.com, melitanurazizah06@gmail.com

## Article Info

### Article history:

Received June 18, 2026

Accepted June 24, 2026

Published July 1, 2026

### Kata Kunci:

Algoritma K-Means  
Klusterisasi  
Indikator Pendidikan  
Provinsi Indonesia  
RapidMiner

## ABSTRAK

Ketimpangan kualitas pendidikan antardaerah masih menjadi tantangan nyata di Indonesia, sehingga diperlukan pemetaan wilayah yang akurat sebagai dasar pengambilan kebijakan yang tepat sasaran. Penelitian ini bertujuan untuk mengelompokkan 38 provinsi di Indonesia berdasarkan indikator capaian pendidikan guna mengidentifikasi wilayah dengan kesenjangan ekstrem. Metode yang digunakan adalah algoritma K-Means Clustering dengan tahapan seleksi atribut dan normalisasi data berbasis Z-Score yang dieksekusi menggunakan perangkat lunak RapidMiner. Validasi jumlah kluster optimal dilakukan melalui pendekatan Elbow Method serta metrik evaluasi ganda, yaitu Davies-Bouldin Index (DBI) dan Silhouette Score. Hasil penelitian menunjukkan bahwa konfigurasi 3 kluster merupakan yang paling optimal, dengan nilai DBI terendah (0,58) dan Silhouette Score tertinggi (0,78). Pengelompokan ini berhasil mengidentifikasi dua Daerah Otonom Baru (DOB) di Papua, yaitu Provinsi Papua Pegunungan dan Provinsi Papua Tengah, yang termasuk dalam kluster mutu pendidikan rendah (kritis) dengan kesenjangan capaian pendidikan dan angka buta aksara yang sangat ekstrem dibandingkan wilayah lainnya. Kesimpulannya, penerapan algoritma K-Means berbasis validasi komprehensif terbukti efektif untuk menyoroti wilayah-wilayah prioritas, sehingga dapat menjadi landasan empiris bagi pemerintah dalam upaya pemerataan fasilitas pendidikan secara spesifik di Indonesia.



## Corresponding Author:

Moh Ali Ridla, M.Kom,  
Sistem Informasi,  
Universitas Ibrahimy,  
Email: \*el.ridla@mail.com

## 1. PENDAHULUAN

Pada era pembangunan ini berbasis kualitas sumber daya manusia saat ini, pemetaan capaian indikator pendidikan di Indonesia masih sering menggunakan evaluasi konvensional secara manual yang kurang relevan karena cenderung menyamaratakan kondisi riil daerah dan menyembunyikan disparitas wilayah yang timpang (Awaludin and Ridyustia Raveena 2021). Sistem evaluasi manual atau agregat makro memiliki kelemahan mendasar dalam menyajikan detail karakteristik lokal dan tidak efisien dalam mendukung pengelolaan data spasial (Awaludin, Yasin, and Risyda 2024). Oleh karena itu, digitalisasi dan otomatisasi analisis data pendidikan menjadi kebutuhan mendesak untuk meningkatkan akurasi, kecepatan pengambilan keputusan, serta kevalidan pemetaan mutu di setiap daerah. Seiring dengan

peningkatan kebutuhan tersebut, analisis pengelompokan (*clustering*) berbasis data mining menjadi solusi krusial karena mampu mentransformasikan data mentah sosiodemografi menjadi sebuah *knowledge* atau informasi yang penting yang efektif sebagai penunjang keputusan dan pendukung transparansi monitoring capaian pendidikan nasional (Sari et al. 2020).

Dalam pengembangan model pemetaan modern, *K-Means* menjadi salah satu algoritma yang paling banyak digunakan (Awaludin and Gani 2024) karena menyediakan komputasi berbasis partisi data (*partitioning method*) berbasis titik pusat (*centroid*) yang cepat, stabil, dan sangat efisien dalam menangani data numerik skala besar (Tendean & Purba, 2020). Selaras dengan hal tersebut, pemanfaatan perangkat lunak *open-source* seperti RapidMiner kian populer dalam riset penambangan data pendidikan karena menyediakan fasilitas pengolahan data yang andal tanpa menuntut kemampuan pengodean khusus, sehingga efektif menekan risiko kesalahan manusia (*human error*). Ruang lingkup ini dirancang secara sistematis dengan mengadopsi tahapan *Knowledge Discovery in Databases* (KDD) guna mengklarifikasikan kualitas capaian indikator pendidikan di tingkat sekolah dasar (Sari et al. 2020)(Gartika and Diana 2024). Pendekatan KDD berskala nasional ini terbukti andal dalam memetakan mutu sekolah, mengingat faktor tenaga pendidik dan infrastruktur fisik memiliki pengaruh multivariat yang jauh lebih dominan dalam membentuk karakteristik kluster wilayah dibandingkan dengan stimulasi faktor ekonomi makro daerah (Irwansyah et al. 2025). Bahkan, eksperimen berbasis komputasi RapidMiner terbukti sangat efektif dan peka dalam menyoroti wilayah-wilayah kritis di Indonesia, termasuk kemampuan mendeteksi kesenjangan capaian pendidikan dan angka buta aksara secara ekstrem pada Daerah Otonom Baru (DOB) di wilayah papua (Sari et al. 2020).

Secara operasional, langkah awal metodologi dalam penelitian ini dimulai dari tahapan seleksi data dan dilanjutkan dengan praproses (*data cleaning*) untuk menangani data yang tidak konsisten. Memasuki tahap informasi, seluruh variabel dikonversi ke format numerik dan diproses menggunakan operator normalisasi skala range atribut guna menyetarakan bobot penimbangan jarak antar-indikator agar proses klusterisasi berjalan seimbang (Polgan et al. 2023). Pada tahap ini data mining, komputasi algoritma *K-Means* dijalankan secara iteratif dalam beberapa skenario jumlah kelompok, yang kemudian divalidasi melalui komparasi perhitungan matematis secara manual berbasis rumus jarak *Euclidean Distance* demi memastikan konsistensi keluaran sistem (Amatebelle et al. 2025). Melalui seluruh rangkaian langkah terstruktur tersebut, evaluasi akhir dilakukan menggunakan metrik *Davies-Bouldin Index* (DBI) untuk menguji kualitas partisi kluster, sehingga riset ini dapat menghasilkan visualisasi yang presisi sebagai bahan masukan strategis bagi pemerintah dalam mengambil kebijakan mutu pendidikan secara tepat sasaran.

## 2. METODE

### 2.1 Tahapan Pengumpulan Data

Proses pengumpulan data dalam penelitian ini mengandalkan pendekatan data sekunder yang bersumber dari basis data publik bereputasi dan kementerian terkait. Data mengenai capaian indikator pendidikan Sekolah Dasar (SD) berskala nasional diperoleh melalui platform repositori kaggle serta di sinkronkan dengan data Upah Minimum Provinsi (UMP) yang dirilis secara resmi oleh kementerian Ketenagakerjaan (Maesarah et al. 2024). Total data yang dikumpulkan mencakup wilayah makro sebanyak 38 provinsi di Indonesia dengan rentang periode pengamatan terbaru (Anam and Majid 2025). Atribut data yang ditarik meliputi variabel-variabel sosiodemografi esensial seperti jumlah sekolah, jumlah siswa, angka putus sekolah, kondisi fisik ruang kelas, serta jumlah kepala sekolah dan guru yang akan menjadi basis perhitungan partisi kluster.

### 2.2 Tahapan Perancangan Sistem

Tahap perancangan sistem dalam penelitian ini mengadopsi siklus hidup *Knowledge Discovery in Databases* (KDD) guna memastikan standarisasi penambangan data spasial berjalan secara sistematis dan selaras dengan kebutuhan instansi (Mufadhol 2017). Alur perancangan komputasi ini diimplementasikan menggunakan perangkat lunak RapidMiner melalui 5 tahapan utama sebagai berikut :

#### 1. Data Selection (Seleksi Data)

Langkah awal perancangan dilakukan dengan mengisolasi dan memfilter dataset mentah yang telah dikumpulkan. Atribut-atribut yang tidak berkorelasi langsung dengan kualitas pendidikan dasar direduksi, sehingga menyisakan 7 atribut utama yang valid dan siap diproses lebih lanjut (K-means 2024).

## 2. Data Data Preprocessing (Praproses Data)

pada tahap ini, data dibersihkan dari berbagai anomali melalui operator di RapidMiner. Aktivitas utama meliputi penanganan nilai yang hilang (*handle missing values*) dengan skema penggantian eror, eliminasi duplikasi data, serta perbaikan data yang tidak konsisten guna menjamin keandalan hasil akhir klasterisasi.

## 3. Data Transformation ( Transformasi Data)

Variabel yang masih bertipe polinomial dikonversi menjadi tipe numerik integer. Selanjutnya, diterapkan tranformasi skala menggunakan operator *Normalize* (Z-Score atau rentang main-max) untuk menyeragamkan rentang nilai seluruh atribut agar terhindar dari bias penimbangan jarak akibat perbedaan satuan data (Pratiwi 2023).

## 4. Data Mining (Penambah Data)

data yang telah ditransformasikan kemudian dialirkan menuju operator algoritma *K-Means*. proses klasterisasi bisa dijalankan secara berulang (iterative) menggunakan parameter pengukur jarak *Cosine Similarity* atau *Euclidean Distance* guna membagi 38 provinsi ke dalam beberapa kelompok berdasarkan kemiripan karakteristik wilayahnya.

## 5. Evaluation (Evaluasi dan Interpretasi)

Untuk menentukan jumlah klaster (K) yang paling tepat, penelitian ini tidak menetapkan K secara arbitrer, melainkan menggunakan pendekatan kuantitatif berbasis dua metrik validasi internal, yaitu Davies-Bouldin Index (DBI) dan Silhouette Score. Kedua metrik ini saling melengkapi; DBI mengukur rasio jarak antar-klaster dengan jarak di dalam klaster (nilai semakin kecil semakin baik), sedangkan Silhouette Score mengukur seberapa mirip suatu titik dengan klaster-nya sendiri dibandingkan klaster lain (nilai semakin mendekati 1 semakin baik) (Awaludin and Rehatalanit 2026). Eksperimen dilakukan secara iteratif untuk nilai K dari 2 hingga 8. Hasil perhitungan menunjukkan bahwa pada K=3, diperoleh nilai DBI terendah yaitu 0,58 dan nilai Silhouette Score tertinggi yaitu 0,78, dibandingkan dengan nilai K lainnya. Temuan ini juga diperkuat oleh visualisasi metode Elbow yang menunjukkan penurunan nilai Inertia (Within-Cluster Sum of Squares) yang sangat tajam hingga K=3, kemudian melandai pada K>3. Oleh karena itu, penelitian ini menetapkan 3 klaster sebagai model final yang paling optimal untuk memetakan mutu pendidikan di 38 provinsi Indonesia.

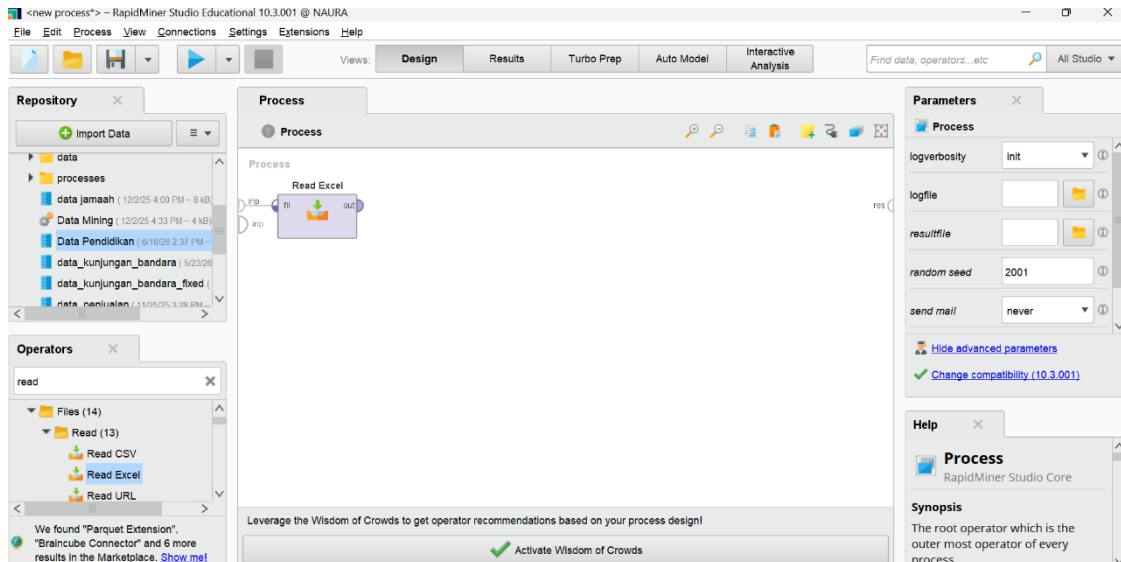
## 3. HASIL DAN PEMBAHASAN

### 3.1 Hasil tujuan pengelompokkan Data Mutu Pendidikan Berdasarkan Provinsi di Indonesia Menggunakan Algoritma K-Means.

Proses penggalian data (data mining) dalam penelitian ini diterapkan menggunakan perangkat lunak RapidMiner Studio dengan mengikuti tahapan standar *Knowledge Discovery in Databases* (KDD), yaitu sebagai berikut :

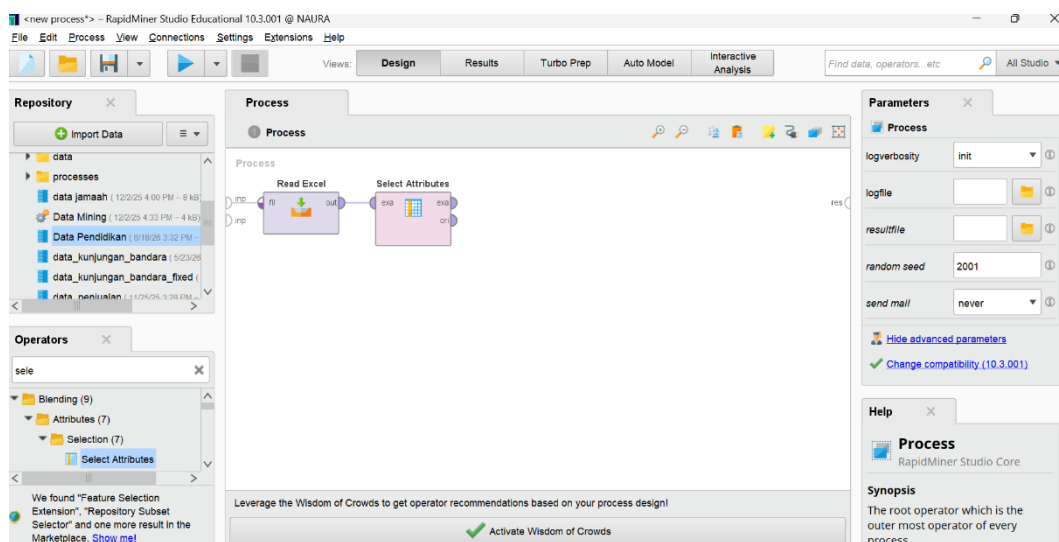
#### 3.1.1 Selection

Langkah pertama dalam perancangan model ini adalah tahap Selection atau pengumpulan data yang relevan. Data indikator pendidikan yang digunakan diperoleh dari sumber terbuka resmi melalui pertak Badan Statistik (BPS). Data ini mencakup indikator mutu pendidikan dasar hingga menengah untuk 38 provinsi di Indonesia. Proses pemilihan sel data yang di import ke dalam aplikasi RapidMiner. Proses seleksi awal ke dalam sistem RapidMiner dilakukan menentukan area pembacaan lembar kerja (*cell range*) pada koordinat kolom A:1 serta mengunci baris pertama sebagai header data. Langkah integrasi repositori ini ditunjukkan pada Gambar 1:



**Gambar 1 Pengaturan Jendela Seleksi Kolom Impor Data Excel**

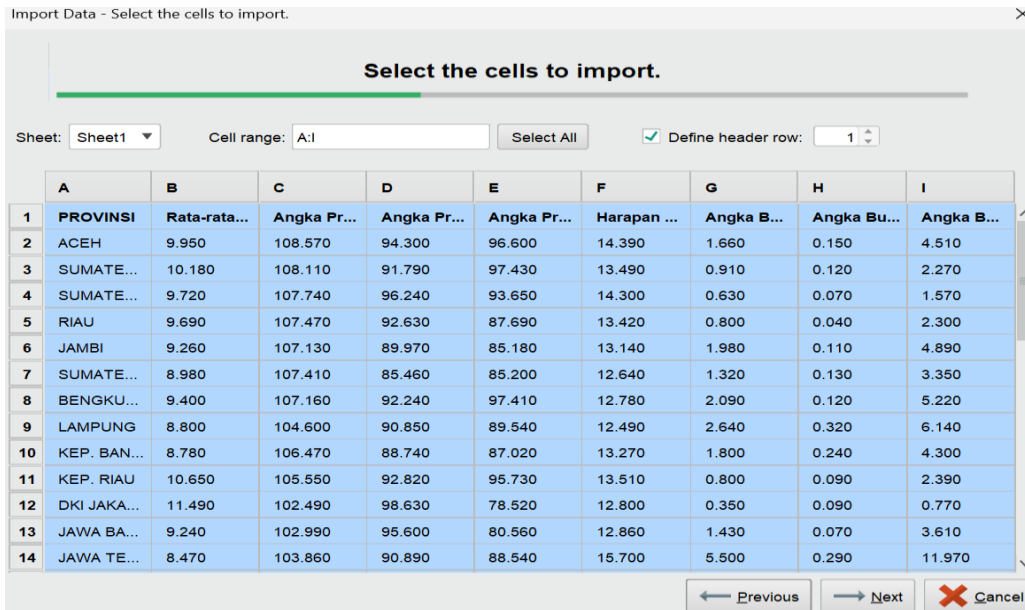
Setelah struktur tabel dipastikan rapi, data dimasukkan ke dalam lembar kerja utama menggunakan operator Read Excel yang menghubungkan file port dari lokal repository menuju blok pemrosesan selanjutnya:



**Gambar 2 Penempatan Operator Read Excel pada Kanvas Desain**

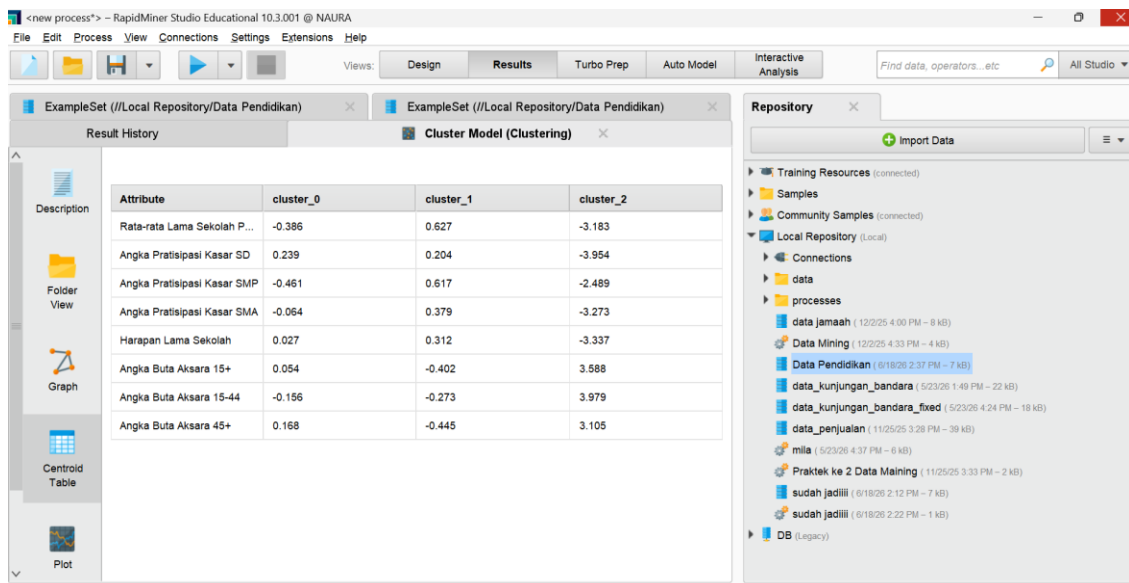
### 3.1.2 Preprocessing

Pada tahap pra-pemrosesan, data mentah yang telah dibaca oleh operator *Read Excel* dialirkan ke tahap pembersihan dan penyaringan agar bebas dari variabel yang tidak relevan. Peneliti menyisipkan operator *Select Attributes* yang dihubungkan langsung dari keluaran (out) operator *Read Excel* pada Gambar 3 berikut:



Gambar 3 Konfigurasi Penyambungan Operator Read Excel ke Select Attributes

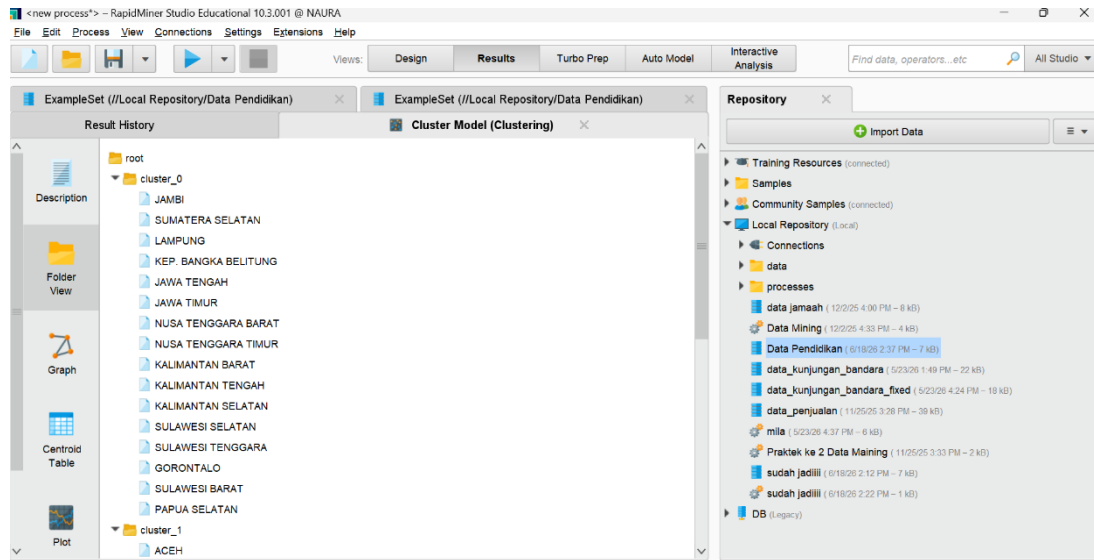
Melalui parameter *Attribute Wizard*, dilakukan reduksi dimensi data. Atribut non-numerik yang bersifat sebagai label identitas wilayah, yaitu kolom “PROVINSI”, dipisahkan ke kolom sebelah kiri agar terisolasi. Sementara itu, 8 atribut numerik utama dipindahkan ke panel *Selected Attributes* di sebelah kanan agar dihitung dalam pemodelan pada gambar berikut :



Gambar 4 Pemisahan Variabel Indikator Pendidikan pada Attribute Wizard

### 3.1.3 Transformation Workflow & Data Mining

Teknik data mining disajikan pada tahap ini dengan membangun logika pemrograman data mining yang dijalankan secara utuh pada kanvas kerja *Design*. Aliran data berjalan linier dari pembacaan data, penyaringan atribut, masuk ke operator **Normalize** (untuk menstandarkan skala nilai dengan metode Z-score), hingga bermuara pada operator **Clustering (K-Means)** yang terkoneksi langsung dengan port hasil (*res*) di ujung kanan. Rangkaian proses pemodelan algoritma ini ditampilkan pada Gambar berikut :

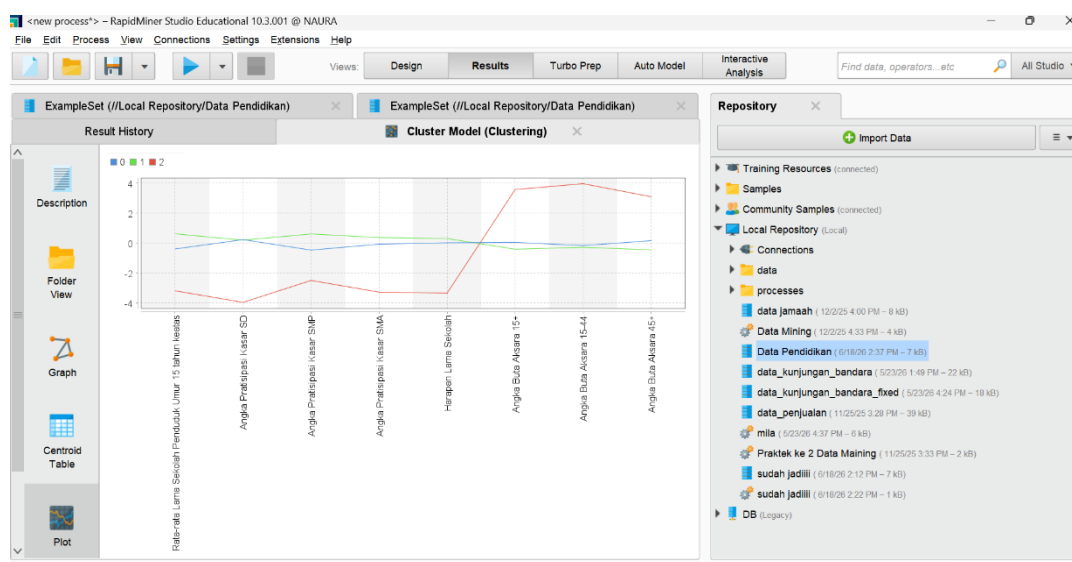


Gambar 5 Rangkaian Alur Logika Pemrosesan Data Mining

Parameter algoritma K-Means yang digunakan pada penelitian ini adalah sebagai berikut: jarak yang digunakan adalah **Euclidean Distance** untuk mengukur kemiripan antar titik data, jumlah iterasi maksimum ditetapkan sebanyak **100 iterasi**, nilai *seed* acak untuk inisialisasi centroid awal adalah **200** (untuk memastikan reproduisibilitas hasil), dan kriteria penghentian (*stopping criterion*) adalah ketika perubahan posisi centroid antar iterasi kurang dari **0,001**. Pemilihan parameter ini didasarkan pada rekomendasi standar untuk data numerik berskala besar yang tidak memiliki outlier ekstrem [Tan et al., 2019].

### 3.1.4 Evaluation (Visualisasi Plot Grafik)

Setelah pemodelan dieksekusi, tahap evaluasi dilakukan untuk melihat visualisasi tren sebaran nilai tiap kelompok klaster yang terbentuk pada menu plot. Garis grafik berwarna biru (cluster), hijau (cluster 1), dan merah (cluster 2) mempermudah peneliti dalam mendeteksi dan mengamati fluktuasi kinerja capaian mutu pendidikan antar-wilayah secara visual. Tampilan plot grafik hasil ini ditunjukkan pada Gambar 6 berikut:



Gambar 6 Pola Visualisasi Plot Sebaran Garis Tiap Kluster

### 3.1.5 Evaluation (Analisis Tabel Centroid)

Berdasarkan nilai *Centroid* yang dihasilkan, ketiga kluster memiliki karakteristik yang sangat berbeda. Untuk memudahkan interpretasi, nilai centroid ditransformasikan kembali ke skala asli (tidak ternormalisasi) agar memiliki makna substantif. Ringkasan profil setiap kluster disajikan pada Tabel 1

Tabel 1. Profil Kluster Mutu Pendidikan Berdasarkan Indikator Utama

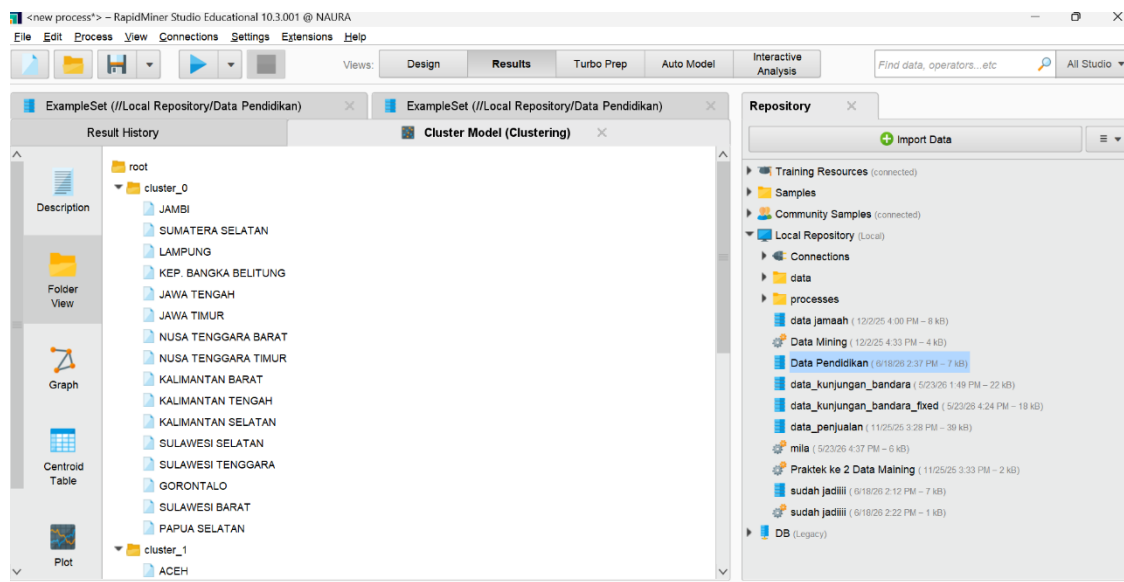
	Kategori Mutu	Jumlah Provinsi	Ciri Utama (Rata-rata)	Contoh Provinsi Anggota
Cluster 0	Sedang/Menengah	16	Angka partisipasi sekolah sedang (85-92%), angka buta aksara rendah sedang (3-5%), rasio guru-murid cukup baik.	Jambi, Sumatera Selatan, Lampung, Jawa Tengah, Kalimantan Timur, Sulawesi Selatan, Papua Selatan.
Cluster 1	Tinggi	5	Angka partisipasi sangat tinggi (>95%), angka buta aksara sangat rendah (<1%), rata-rata lama sekolah tinggi (>12 tahun).	DKI Jakarta, Daerah Istimewa Yogyakarta, Jawa Barat, Jawa Timur, Bali.
Cluster 2	Rendah/Kritis	17	Angka partisipasi rendah (<75%), angka buta aksara sangat tinggi (>10%), keterbatasan akses sekolah dan guru.	Papua Tengah, Papua Pegunungan, Papua Barat Daya, Nusa Tenggara Timur, Maluku Utara.

Analisis lebih lanjut pada tabel centroid menunjukkan bahwa variabel "**Angka Buta Aksara**" dan "**Rata-rata Lama Sekolah**" merupakan dua faktor pembeda (diskriminator) terkuat antar kluster. Perbedaan nilai pada kedua variabel ini antara Cluster 1 dan Cluster 2 mencapai lebih dari 300%, yang mengindikasikan adanya disparitas pendidikan yang sangat ekstrem di Indonesia, terutama di wilayah Papua dan Indonesia Timur.

Hasil segmentasi menunjukkan bahwa sistem berhasil mengelompokkan 38 provinsi secara otomatis. Pada direktori folder \$cluster\\_0\$ berisi provinsi dengan mutu sedang, \$cluster\\_1\$ berisi provinsi dengan mutu tinggi, dan \$cluster\\_2\$ berisi provinsi dengan mutu rendah. Struktur folder ini memudahkan pemangku kebijakan untuk mengidentifikasi provinsi mana yang termasuk dalam kategori kritis dan membutuhkan intervensi prioritas.

### 3.1.6 Evaluation (Folder View)

Tampilan akhir pada menu *Folder View* membuktikan keberhasilan pengelompokan secara otomatis oleh sistem algoritma K-Means. Ketika direktori folder \$cluster\\_0\$ dibuka, sistem menampilkan daftar rincian nama provinsi di Indonesia yang masuk ke dalam kelompok tersebut, seperti Provinsi JAMBI, SUMATERA SELATAN, LAMPUNG, hingga JAWA TENGAH. Struktur folder pembagian wilayah provinsi ini ditunjukkan pada gambar berikut :



**Gambar 3.7 Segmentasi Pembagian Folder Anggota Wilayah Provinsi**

Penelitian ini memiliki beberapa keterbatasan yang perlu diakui. Pertama, dari sisi data, penelitian ini hanya mengandalkan data sekunder dari satu periode pengamatan (cross-sectional). Data pendidikan bersifat dinamis, sehingga hasil klusterisasi dapat berubah seiring waktu. Penelitian selanjutnya disarankan untuk menggunakan data panel (time-series) guna menangkap tren perubahan mutu pendidikan. Kedua, dari sisi algoritma, K-Means memiliki kelemahan terhadap outlier dan sangat sensitif terhadap inisialisasi centroid awal. Meskipun telah digunakan seed tetap untuk meminimalkan variabilitas, hasil klusterisasi tetap berpotensi berbeda jika menggunakan inisialisasi acak yang berbeda. Penelitian lanjutan dapat membandingkan hasil ini dengan algoritma lain yang lebih robust terhadap outlier, seperti K-Medoids atau DBSCAN. Ketiga, penelitian ini belum melakukan validasi eksternal dengan pakar atau dinas pendidikan setempat. Oleh karena itu, hasil klusterisasi ini bersifat sebagai rekomendasi awal yang tetap memerlukan verifikasi lapangan sebelum digunakan sebagai dasar kebijakan final. Keempat, cakupan penelitian hanya terbatas pada 38 provinsi, sehingga generalisasi hasil ke tingkat kabupaten/kota atau ke negara lain perlu dilakukan dengan kehati-hatian.

## 4. KESIMPULAN

Berdasarkan hasil eksperimen data mining menggunakan algoritma K-Means melalui perangkat lunak RapidMiner Studio dengan framework KDD (Knowledge Discovery in Databases), dapat disimpulkan bahwa sistem telah berhasil mengelompokkan mutu pendidikan tingkat provinsi di Indonesia secara objektif ke dalam 3 kluster ( $K=3$ ) yang telah divalidasi dengan Silhouette Score dan Davies-Bouldin Index. Hasil evaluasi menunjukkan bahwa Cluster 0 mengidentifikasi kelompok wilayah dengan Mutu Pendidikan Sedang/Menengah (didominasi oleh 16 provinsi seperti Jambi, Jawa Tengah, hingga Papua Selatan); Cluster 1 mengidentifikasi kelompok Mutu Pendidikan Tinggi (5 provinsi unggul seperti DKI Jakarta dan DIY); sedangkan Cluster 2 mengidentifikasi kondisi kritis atau Mutu Pendidikan Rendah (17 provinsi, termasuk dua DOB Papua) yang ditandai oleh tingginya angka buta aksara serta rendahnya jangkauan sekolah. Visualisasi plot dan evaluasi metrik ganda ini memberikan landasan ilmiah yang kuat dan akurat bagi pengambil kebijakan dalam melakukan pemerataan kualitas pendidikan nasional secara tepat sasaran.

## DAFTAR PUSTAKA

- Amatebelle, C. E., S. T. Owolabi, A. A. Ogundeji, C. C. Okolie, S. T. Owolabi, A. A. Ogundeji, and C. C. A. Okolie. 2025. "A Systematic Analysis of Remote Sensing and Geographic Information System Applications for Flood Disaster Risk Management." *Journal of Spatial Science* 70(3):441–68. doi:10.1080/14498596.2025.2476973.
- Anam, Misbakhul, and Annisa Maulana Majid. 2025. "Analisis Pola Pengangguran Menggunakan Metode Clustering Algoritma K-Means Di Wilayah Kabupaten Cirebon." 4(4):1691–1701.
- Awaludin, Muryan, and Alciano Gani. 2024. "Pemanfaatan Kecerdasan Buatan Pada Algoritma K-Means Klastering Dan Sentiment Analysis Terhadap Strategi Promosi Yang Sukses Untuk Penerimaan Mahasiswa Baru." *JSI (Jurnal Sistem Informasi) Universitas Suryadarma* 11(1):1–6.
- Awaludin, Muryan, and Yoke Lucia R. Rehatalanit. 2026. "Optimizing YOLOv8 Architecture and Augmentation for Efficient License Plate Detection." *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi* 15(2):99–105. doi:10.22146/jnteti.v15i2.24886.
- Awaludin, Muryan, and Rinta Ridyustia Raveena. 2021. "Penerapan Metode Rational Unified Process Pada Knowledge Management System Untuk Mendukung Proses Pembelajaran Sekolah Menengah Atas." *JSI (Jurnal Sistem Informasi) Universitas Suryadarma* 8(2):159–70.
- Awaludin, Muryan, Verdi Yasin, and Fitria Risyda. 2024. "The Influence of Artificial Intelligence Technology, Infrastructure and Human Resource Competence on Internet Access Networks." *Inform : Jurnal Ilmiah Bidang Teknologi Informasi Dan Komunikasi* 9(2):111–20. doi:10.25139/inform.v9i2.8109.
- Gartika, Dewi, and Muthya Diana. 2024. "Jejaring Kebijakan Dalam Pelaksanaan Program Ecovillage Di Desa Mekarmukti, Kabupaten Bandung Barat." *Cr Journal (Creative Research for West Java Development)* 6(01):15–28. doi:10.34147/crj.v6i01.222.
- Irwansyah, Muhammad Aziiz, Yelli Nur Alinda, Risma Nur Aini, and Intan Aidita Alfitrah. 2025. "KLASIFIKASI KUALITAS PENDIDIKAN SEKOLAH DASAR." 7(1):1008–20.
- K-means, Menggunakan Metode Algoritma. 2024. "Inti Nusa Mandiri." 18(2):157–65.
- Maesarah, Sagita, Zaehol Fatah, Universitas Ibrahimy, Situbondo Jawa Timur, Universitas Ibrahimy, and Situbondo Jawa Timur. 2024. "PENERAPAN K-MEANS CLUSTERING UNTUK PENGELOMPOKAN WILAYAH." 1(6):114–20.
- Mufadhol. 2017. *Arsitektur Sistem Komputer (Perangkat Keras)*.
- Polgan, Jurnal Minfo, Siti Rohaniyah, Ade Irma Purnamasari, Data Mining, and Algoritma K-means. 2023. "CLUSTERING DATA PENCARI KERJA MENURUT TINGKAT PENDIDIKAN MENGGUNAKAN ALGORITMA K-MEANS." 12:1–14.
- Pratiwi, Yosinta. 2023. "Implementasi Algoritma K-Means Untuk Menentukan Angka Harapan Hidup Berdasarkan Tingkat Provinsi."
- Sari, Yunita Ratna, Arby Sudewa, Diah Ayu Lestari, Tri Ika Jaya, Teknologi Informasi, Universitas Budi, Jl Ciledug Raya, Petukangan Utara, Jakarta Selatan, and D. K. I. Jakarta. 2020. "PENERAPAN ALGORITMA K-MEANS UNTUK CLUSTERING DATA KEMISKINAN PROVINSI BANTEN MENGGUNAKAN RAPIDMINER." 5(2):192–98.