Analisis Kesenjangan Infrastruktur Internet Sekolah di Indonesia melalui Klasterisasi K-Means

Aydil Harsy Putra1   
*Sistem Informasi  
Universitas Komputer Indonesia*Bandung, Indonesia  
aydil.10522021@mahasiswa.unikom.ac.id

Aslam Ainul Yaqin4  
*Sistem Informasi  
Universitas Komputer Indonesia*Bandung, Indonesia  
aslam.10522036@mahasiswa.unikom.ac.idMuhamad Farrel Alfarizi2  
*Sistem Informasi*  
*Universitas Komputer Indonesia*Bandung, Indonesia  
farrel.10522023@mahasiswa.unikom.ac.id

.Muhammad Dzikri Fathurramdan3  
*Sistem Informasi  
Universitas Komputer Indonesia*Bandung, Indonesia  
dzikri.10522027@mahasiswa.unikom.ac.id

*Abstrak*—Pemerataan akses internet pada sektor pendidikan masih menjadi tantangan di berbagai wilayah Indonesia. Penelitian ini bertujuan untuk mengidentifikasi kesenjangan digital antarprovinsi dengan memanfaatkan algoritma K-Means Clustering pada data sekolah jenjang dasar hingga menengah. Analisis dilakukan menggunakan perangkat lunak Orange Data Mining, dengan variabel utama Ratio\_Internet, yaitu rasio antara jumlah sekolah yang memiliki akses internet terhadap total sekolah di setiap provinsi. Proses klasterisasi dilakukan dengan jumlah klaster tetap (k = 3) dan menghasilkan tiga kelompok utama, yaitu provinsi dengan akses internet rendah, sedang, dan tinggi. Hasil analisis menunjukkan bahwa provinsi di wilayah timur Indonesia umumnya tergolong dalam klaster dengan akses internet terendah, sedangkan wilayah barat dan perkotaan lebih banyak tergolong dalam klaster dengan akses tinggi. Visualisasi scatter plot dan box plot memperkuat temuan ini, serta memberikan gambaran yang jelas tentang sebaran kesenjangan infrastruktur digital secara nasional. Temuan ini diharapkan dapat mendukung perumusan kebijakan yang lebih tepat sasaran dalam mendorong pemerataan transformasi pendidikan digital di Indonesia.

Keywords—klasterisasi, K-Means, pendidikan, akses internet, infrastruktur digital

# PENDAHULUAN

Kemajuan teknologi informasi dan komunikasi (TIK) telah memberikan dampak besar dalam berbagai bidang, termasuk pendidikan. Perubahan ini telah menggeser metode pembelajaran konvensional yang semula bersifat pasif menjadi lebih dinamis, interaktif, dan mendorong kolaborasi, sehingga memungkinkan peserta didik untuk lebih aktif berpartisipasi dalam proses belajar [1]. Isu-isu global turut memengaruhi implementasi kurikulum di era digital. Perbedaan dalam akses terhadap teknologi [2], ketimpangan dalam penyelenggaraan pendidikan online [3], serta kendala infrastruktur di sejumlah daerah menjadi tantangan utama yang menghambat proses pendidikan secara merata [4]. Kurikulum perlu menyesuaikan diri dengan pesatnya kemajuan teknologi [5], begitu pula guru dan siswa yang dihadapkan pada tantangan pembelajaran digital yang kompleks dan menuntut adaptasi cepat [6]. Terlebih setelah pandemi COVID-19, kebutuhan terhadap infrastruktur internet di sekolah menjadi semakin mendesak dan fundamental [7].

Di Indonesia, pemerintah telah menginisiasi berbagai program digitalisasi pendidikan seperti *Merdeka Belajar*, *Asesmen Nasional Berbasis Komputer (ANBK)*, serta pengembangan platform daring seperti *Rumah Belajar* dan *Merdeka Mengajar* [8]. Namun demikian, kesenjangan infrastruktur digital antara wilayah perkotaan dan pedesaan, serta antara wilayah barat dan timur Indonesia masih sangat nyata [9]. Banyak sekolah di daerah tertinggal, terdepan, dan terluar (3T) yang masih mengalami kendala dalam akses internet, baik dari sisi ketersediaan jaringan, kecepatan koneksi, maupun stabilitas layanan [10].

Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi (Kemendikbudristek) bersama Kementerian Komunikasi dan Informatika (Kominfo) telah berupaya mengatasi permasalaha ini melalui program bantuan internet sekolah dan pembangunan infrastruktur BTS melalui Badan Aksesibilitas Telekomunikasi dan Informasi (BAKTI) [11]. Namun, efektivitas dari program-program tersebut sangat bergantung pada kemampuan pemerintah dalam mengidentifikasi wilayah prioritas secara objektif dan berbasis data.

Untuk itu, diperlukan pendekatan analitik berbasis data guna memberikan gambaran objektif tentang kondisi infrastruktur internet di sekolah-sekolah Indonesia. Salah satu pendekatan yang relevan adalah analisis klaster menggunakan metode K-Means, yang memungkinkan pengelompokan wilayah berdasarkan kemiripan atribut tertentu [12], seperti jumlah sekolah dengan akses internet, jenis koneksi, serta tingkat pendidikan. Dengan mengelompokkan provinsi atau wilayah ke dalam klaster tertentu, pemerintah dapat memetakan kesenjangan infrastruktur secara lebih sistematis dan efisien.

Melalui proses klasterisasi ini, hasil pengelompokan dapat menjadi acuan untuk penentuan prioritas pembangunan infrastruktur digital pendidikan. Wilayah yang tergolong ke dalam klaster dengan kondisi infrastruktur rendah dapat menjadi sasaran utama dalam program peningkatan akses internet dan dukungan teknologi. Selain itu, pendekatan ini juga berguna untuk mengevaluasi sejauh mana pemerataan akses digital telah dicapai di Indonesia, dan bagaimana dampaknya terhadap pelaksanaan kebijakan pendidikan berbasis teknologi.

Dengan memanfaatkan data terbuka dari Portal Satu Data Indonesia dan teknologi visualisasi data seperti Orange Data Mining, analisis ini diharapkan mampu menghasilkan informasi yang komprehensif, akurat, dan mendukung pengambilan keputusan berbasis bukti (evidence-based policy). Penelitian ini sekaligus mendorong pemanfaatan teknik data mining di bidang pendidikan sebagai alat bantu untuk menyusun strategi pembangunan yang lebih inklusif, adaptif, dan berkelanjutan dalam menghadapi tantangan pendidikan digital di Indonesia.

# Metodologi

Metode atau pendekatan ilmiah dalam sebuah penelitian berfungsi sebagai sarana untuk mencapai tujuan penelitian. Karena berperan sebagai alat, metode ini bersifat fleksibel dan tidak kaku, sehingga dapat digunakan dalam berbagai variasi sesuai dengan karakteristik objek dan bidang ilmu yang menjadi fokus penelitian [13].

Penelitian ini menggunakan pendekatan kuantitatif deskriptif eksploratif, dengan menerapkan teknik data mining untuk menganalisis pola kesenjangan infrastruktur internet di sekolah-sekolah Indonesia. Tujuan utama dari penelitian ini adalah untuk mengelompokkan sekolah berdasarkan kondisi infrastruktur internet yang dimilikinya. Metode ini dianggap sesuai karena memungkinkan pengelompokan berdasarkan kemiripan karakteristik tanpa perlu label kelas tertentu, yang secara umum dikenal sebagai unsupervised learning [14].

Proses analisis ini dilakukan dengan menggunakan perangkat lunak Orange Data Mining, yaitu platform analisis data open-source berbasis Python yang dirancang untuk mendukung analitik visual dan interaktif [15]. Orange menyediakan berbagai fitur dalam bentuk widget, seperti File Import, Data Preprocessing, Normalize, K-Means, Silhouette Plot, dan Scatter Plot yang digunakan dalam penelitian ini. Kelebihan Orange terletak pada kemudahan penggunaan melalui drag-and-drop, serta kemampuannya menampilkan proses data secara visual [16]. Orange juga banyak digunakan dalam penelitian pendidikan, biomedis, dan bisnis karena mendukung metode eksplorasi berbasis CRISP-DM.

Metode yang digunakan dalam penelitian ini adalah klasterisasi. Klasterisasi (clustering) merupakan salah satu teknik dalam machine learning yang bertujuan untuk mengelompokkan data ke dalam beberapa kelompok (klaster) berdasarkan kemiripan atau kedekatan nilai atributnya [17]. Teknik ini termasuk dalam kategori unsupervised learning karena proses pengelompokan dilakukan tanpa adanya label kelas sebelumnya [18].

Algoritma klasterisasi yang digunakan dalam penelitian ini adalah K-Means, yaitu metode yang bekerja dengan cara membagi data ke dalam sejumlah klaster (K) berdasarkan jarak terdekat ke pusat klaster (centroid) yang dihitung secara iteratif.Kelebihan dari algoritma K-Means adalah kesederhanaannya, efisiensinya dalam mengelola data berskala besar, serta hasil pengelompokan yang mudah diinterpretasikan secara visual [19]. Namun, metode ini juga memiliki beberapa kelemahan, seperti sensitivitas terhadap pemilihan nilai K dan centroid awal [20]. Penerapan teknik klasterisasi ini diharapkan dapat memberikan gambaran yang lebih terstruktur mengenai kondisi infrastruktur internet di sekolah-sekolah di Indonesia, yang nantinya dapat digunakan sebagai dasar dalam perumusan kebijakan pemerataan akses teknologi Pendidikan.

Untuk mendukung tahapan analisis yang sistematis, penelitian ini mengadopsi model proses data mining standar yaitu CRISP-DM (Cross Industry Standard Process for Data Mining). Model ini terdiri atas enam tahapan utama, yaitu pemahaman bisnis (business understanding), pemahaman data (data understanding), persiapan data (data preparation), pemodelan (modeling), evaluasi (evaluation), dan implementasi hasil (deployment) [21]. Model CRISP-DM telah terbukti efektif digunakan dalam berbagai penelitian eksploratif berbasis data di bidang pendidikan, kesehatan, dan sosial [22].

## Business Understanding

Pada tahap ini dilakukan pemahaman dan analisis terhadap latar belakang kesenjangan infrastruktur internet di sekolah-sekolah di Indonesia guna mengidentifikasi permasalahan yang terjadi. Hasil analisis tersebut kemudian dijadikan dasar untuk merumuskan tujuan penelitian serta merancang strategi penerapan proses pemodelan data mining menggunakan Orange Data Mining.

## Data Understanding

Pada tahap data understanding, data yang digunakan dalam penelitian ini diperoleh dari website Portal Satu Data Indonesia. Data tersebut mencakup informasi infrastruktur digital sekolah-sekolah di seluruh Indonesia, seperti kode kementrian, kode bps, provinsi, jumlah sekolah per wilayah, ketersediaan akses internet, Tingkat Pendidikan, kepemilikan (negeri atau swasta).

Data diperoleh dalam format spreadsheet (XLS/XLSX) dan mencerminkan kondisi infrastruktur internet pada kurun waktu 2023 hingga 2024. Setiap baris data mewakili unit wilayah administratif (provinsi/kabupaten) yang akan diklasifikasikan berdasarkan karakteristik infrastruktur digitalnya. Dalam tahap ini dilakukan identifikasi atribut penting dan hubungan antar atribut, serta eksplorasi awal melalui visualisasi dan analisis statistik deskriptif untuk mendapatkan pemahaman awal mengenai distribusi data.

## Data Preparation

Tahap data preparation merupakan proses penting dalam penelitian ini karena berfungsi menyiapkan data yang akan digunakan untuk membangun model klasterisasi [23]. Proses dimulai dengan melakukan cleaning, yaitu membersihkan data dari nilai-nilai kosong, data yang tidak konsisten, dan duplikat yang dapat mengganggu hasil analisis. Setelah itu dilakukan selection, yaitu pemilihan atribut-atribut yang dianggap relevan dan berpengaruh terhadap kualitas klasterisasi, seperti provinsi, jumlah sekolah yang memiliki akses internet, jumlah sekolah, Tingkat Pendidikan.

## Modeling

Pemodelan dilakukan menggunakan algoritma K-Means Clustering yang bertujuan untuk mengelompokkan wilayah atau provinsi berdasarkan kesamaan kondisi infrastruktur internet di sekolah-sekolahnya. Algoritma ini termasuk dalam kategori unsupervised learning karena tidak memerlukan label kelas sebelumnya.

## Evaluation

Tahap evaluasi bertujuan untuk menilai kualitas hasil klasterisasi yang telah dibentuk menggunakan algoritma K-Means [24]. Untuk mengukur kualitas klaster, digunakan metode evaluasi internal berupa Silhouette Score, yaitu metrik yang menunjukkan seberapa baik suatu data cocok dengan klasternya sendiri dibandingkan dengan klaster lain. Semakin tinggi nilai Silhouette Score (mendekati 1), maka semakin baik hasil pengelompokannya. Evaluasi dilakukan dengan membandingkan beberapa nilai k (jumlah klaster), lalu memilih nilai yang memberikan hasil skor terbaik secara statistik maupun visualisasi. Selain itu, dilakukan interpretasi hasil dengan membandingkan karakteristik antar klaster agar diketahui perbedaan signifikan antar kelompok wilayah.

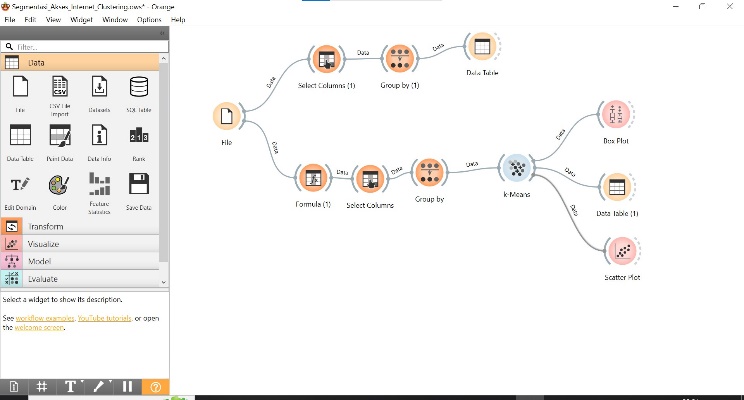
## Deployment

Tahap akhir dari proses CRISP-DM ini adalah deployment, yaitu penyajian hasil klasterisasi dalam bentuk laporan dan visualisasi yang mudah dipahami dan dapat dimanfaatkan oleh pihak-pihak terkait, seperti pemerintah, pengambil kebijakan, dan institusi Pendidikan [25]. Hasil pengelompokan disajikan melalui tabel distribusi provinsi dalam masing-masing klaster, grafik scatter plot, serta visualisasi silhouette plot yang mendukung interpretasi hasil. Hasil penelitian ini dapat dijadikan acuan untuk menyusun kebijakan pemerataan infrastruktur digital pendidikan, terutama bagi wilayah yang tergolong dalam klaster dengan tingkat infrastruktur rendah. Deployment juga memungkinkan pembaruan data secara berkala agar proses analisis dapat dilakukan secara berkelanjutan di masa depan.

# HASIL DAN PEMBAHASAN

Penelitian ini menggunakan dataset infrastruktur internet sekolah di Indonesia untuk periode tahun 2021 hingga 2024. Dataset diperoleh dari Portal Satu Data Indonesia dan mencakup tujuh atribut utama, yaitu Pendidikan, Kode Kemdagri, Kode BPS, Provinsi, Kepemilikan, Jumlah\_Internet, dan Jumlah\_Sekolah. Setiap baris dalam data mewakili informasi rekapitulasi berdasarkan kombinasi wilayah, jenjang pendidikan, dan kepemilikan sekolah.

Kolom Pendidikan menunjukkan jenjang satuan pendidikan (SD, SMP, SMA, SMK), sementara Kode Kemdagri dan Kode BPS merupakan kode identifikasi administratif resmi. Kolom Provinsi mencantumkan nama wilayah, Kepemilikan menunjukkan status sekolah (negeri atau swasta), Jumlah\_Internet berisi jumlah sekolah yang memiliki akses internet, dan Jumlah\_Sekolah berisi total sekolah yang terdaftar.



Gambar 1. Alur Orange

## Pra-pemrosesan Data

Sebelum dilakukan analisis, data mengalami tahap pra-pemrosesan. Langkah-langkah yang dilakukan meliputi pembersihan data dari nilai kosong (missing value), penghapusan duplikasi, serta penyelarasan format data. Dataset kemudian dimasukkan ke dalam Orange Data Mining sebagai platform utama pemrosesan visual.

## Perhitungan Rasio Akses Internet

Langkah awal analisis adalah membentuk variabel baru bernama Ratio\_Internet dengan rumus:

|  |
| --- |
| *Ratio\_Internet = Jumlah\_Internet / Jumlah\_Sekolah* |

Variabel ini dihitung menggunakan widget Formula di Orange dan menjadi indikator utama keterhubungan digital tiap provinsi. Rasio ini merepresentasikan proporsi sekolah yang memiliki akses internet dan digunakan sebagai basis dalam proses klasterisasi.

## Pemilihan Atribut dan Agregasi

Setelah variabel rasio terbentuk, proses dilanjutkan dengan menggunakan widget Select Columns untuk menyusun ulang peran atribut dalam analisis. Atribut Ratio\_Internet dipilih sebagai feature karena akan digunakan dalam perhitungan jarak antar data pada klasterisasi. Atribut Provinsi ditetapkan sebagai meta karena berguna dalam identifikasi hasil klaster berdasarkan wilayah administratif. Sementara itu, atribut seperti Kode BPS, Kode Kemdagri, Jumlah\_Internet, Jumlah\_Sekolah, Kepemilikan, dan Pendidikan dikesampingkan (ignored) karena tidak secara langsung berkontribusi terhadap pembentukan klaster.

Setelah menentukan atribut, dilakukan agregasi data menggunakan widget Group By. Data dikelompokkan berdasarkan Provinsi, dan nilai Ratio\_Internet dihitung rata-ratanya menggunakan fungsi Mean. Proses ini menghasilkan dataset terstruktur yang merepresentasikan nilai rerata rasio akses internet tiap provinsi, yang akan digunakan dalam proses klasterisasi.

## Klasterisasi Menggunakan K-Means

Tahap berikutnya adalah penerapan algoritma K-Means Clustering, yaitu salah satu metode klasterisasi paling banyak digunakan dalam bidang data mining dan machine learning. K-Means termasuk dalam kategori unsupervised learning, di mana pengelompokan data dilakukan tanpa label kelas. Tujuan utamanya adalah untuk membagi data ke dalam sejumlah kelompok (klaster) berdasarkan kedekatan nilai atribut.

Cara kerja K-Means dimulai dengan menentukan jumlah klaster (k) yang diinginkan. Kemudian, algoritma akan memilih secara acak k titik awal sebagai pusat klaster atau centroid. Selanjutnya, setiap data akan dihitung jaraknya dengan masing-masing centroid menggunakan rumus jarak umumnya jarak Euclidean dan kemudian data tersebut akan ditugaskan ke klaster dengan centroid terdekat. Setelah semua data dikelompokkan, posisi centroid akan diperbarui dengan menghitung rata-rata posisi seluruh titik data dalam setiap klaster. Proses ini terus diulang hingga terjadi konvergensi, yaitu ketika posisi centroid tidak lagi berubah secara signifikan atau perpindahan data antar klaster telah berhenti. Dengan demikian, algoritma ini berupaya meminimalkan Within-Cluster Sum of Squares (WCSS), yaitu total jarak kuadrat antar data dalam satu klaster terhadap centroid-nya.

Dalam penelitian ini, algoritma K-Means digunakan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan kondisi infrastruktur internet pada satuan pendidikan. Atribut yang digunakan sebagai dasar dalam proses klasterisasi adalah Ratio\_Internet, yaitu rasio antara jumlah sekolah yang memiliki akses internet terhadap total jumlah sekolah di tiap provinsi. Rasio ini dihitung terlebih dahulu menggunakan widget Formula, dan menjadi indikator utama yang merepresentasikan tingkat keterhubungan digital antarwilayah. Melalui pendekatan ini, provinsi-provinsi dikelompokkan ke dalam tiga klaster utama, yaitu wilayah dengan tingkat infrastruktur internet tinggi, sedang, dan rendah.

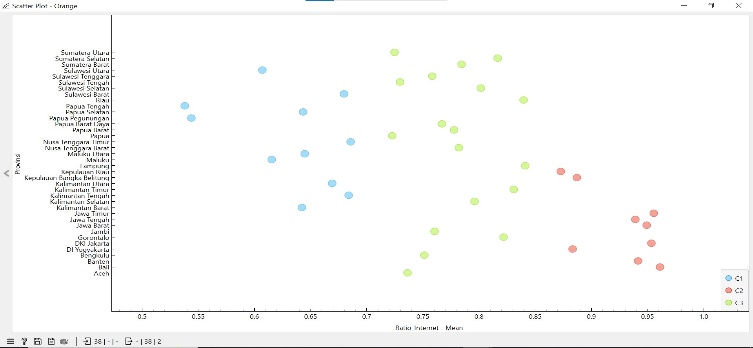
Dalam widget K-Means, digunakan konfigurasi Fixed = 3, yang berarti data akan dikelompokkan secara langsung ke dalam tiga klaster. Nilai ini dipilih berdasarkan hasil evaluasi menggunakan metode Elbow dan Silhouette Score, yang menunjukkan bahwa klasterisasi terbaik terjadi ketika jumlah klaster adalah tiga. Selain konfigurasi tetap, juga dilakukan eksplorasi dengan rentang nilai klaster dari From = 2 hingga To = 6, untuk melihat bagaimana variasi jumlah klaster memengaruhi hasil pemodelan.



Gambar 2. Hasil Klasterisasi

## Visualisasi dan Interpretasi

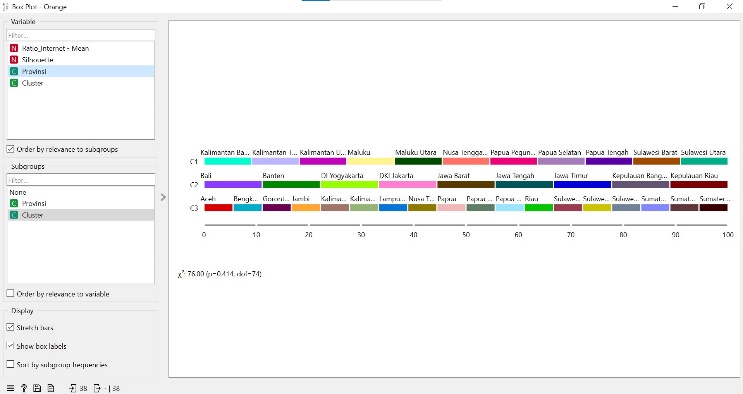
Pada tahap visualisasi data, digunakan dua jenis grafik utama, yaitu scatter plot dan box plot. Kedua metode visualisasi ini dipilih karena mampu menyajikan sebaran data serta membedakan karakteristik antar klaster secara visual dan intuitif. Scatter plot digunakan untuk menggambarkan distribusi masing-masing provinsi dalam ruang dua dimensi berdasarkan nilai Ratio\_Internet. Setiap titik pada scatter plot merepresentasikan satu provinsi, dengan posisinya menunjukkan proporsi sekolah yang memiliki akses internet. Melalui visualisasi ini, perbedaan tingkat keterhubungan digital antarprovinsi dapat diamati dengan jelas.



Gambar 3. Hasil Scatter Plot

Selain itu, visualisasi scatter plot juga memperlihatkan hasil dari proses klasterisasi menggunakan algoritma K-Means, di mana setiap klaster diberi simbol dan warna yang berbeda untuk membedakan masing-masing kelompok:

* Klaster 1 direpresentasikan dengan lingkaran berwarna biru, mencakup provinsi-provinsi yang memiliki rasio akses internet rendah.
* Klaster 2 digambarkan dengan lingkaran berwarna merah, berisi provinsi-provinsi dengan rasio akses internet yang tinggi, mencerminkan daerah yang telah relatif maju dalam infrastruktur digital pendidikan.
* klaster 3 ditunjukkan dengan lingkaran berwarna hijau, merepresentasikan wilayah dengan rasio akses internet pada tingkat sedang.



Gambar 4. Hasil Box Plot

Sementara itu, box plot digunakan untuk memberikan gambaran statistik yang lebih rinci terkait persebaran nilai *Ratio\_Internet* dalam setiap klaster. Visualisasi ini menunjukkan rentang nilai minimum, maksimum, kuartil pertama dan ketiga, serta median, sehingga memudahkan dalam mengidentifikasi struktur data di masing-masing klaster. Selain itu, box plot juga memungkinkan pendeteksian outlier atau nilai ekstrem yang dapat memengaruhi distribusi dan interpretasi data. Sebagai contoh, pada klaster 2 terlihat bahwa nilai median rasio akses internet berada pada posisi yang lebih tinggi dibandingkan dengan klaster lainnya, yang mengindikasikan bahwa kelompok ini terdiri dari provinsi-provinsi dengan tingkat keterhubungan digital yang paling merata dan optimal di lingkungan sekolah.

## Pembahasan

Berdasarkan Hasil Diatas, dapat disimpulkan bahwa penerapan algoritma K-Means berhasil mengelompokkan provinsi-provinsi di Indonesia ke dalam tiga kategori utama berdasarkan rasio akses internet pada satuan pendidikan. Penggunaan variabel Ratio\_Internet sebagai indikator utama terbukti efektif dalam menggambarkan tingkat keterhubungan digital di setiap wilayah secara objektif dan terukur. Hasil klasterisasi menunjukkan bahwa masih terdapat kesenjangan signifikan dalam infrastruktur internet antarprovinsi.

Provinsi-provinsi dalam Klaster 1 umumnya berasal dari wilayah timur Indonesia yang masih memiliki rasio akses internet rendah. Keterbatasan ini dapat disebabkan oleh faktor geografis, rendahnya penetrasi teknologi, serta terbatasnya infrastruktur pendukung. Sebaliknya, provinsi-provinsi dalam Klaster 2 seperti DKI Jakarta, Jawa Barat, dan DI Yogyakarta menunjukkan tingkat keterhubungan yang sangat baik. Hal ini mencerminkan adanya kemajuan signifikan dalam pemanfaatan teknologi informasi dan pembangunan infrastruktur pendidikan digital di wilayah-wilayah tersebut. Klaster 3 menampung provinsi-provinsi dengan tingkat rasio akses internet sedang, yang menunjukkan adanya potensi peningkatan, namun masih memerlukan dukungan kebijakan dan investasi infrastruktur lanjutan.

Secara keseluruhan, hasil ini menunjukkan bahwa pendekatan klasterisasi dengan K-Means dapat dijadikan sebagai alat bantu yang efektif dalam pemetaan kondisi infrastruktur pendidikan digital. Temuan ini dapat menjadi acuan bagi pengambil kebijakan dalam menentukan prioritas pembangunan infrastruktur internet di sektor pendidikan. Wilayah dengan rasio akses rendah perlu menjadi fokus utama intervensi pemerintah agar kesenjangan digital di Indonesia dapat diminimalisasi secara berkelanjutan.

# KESIMPULaN

Penelitian ini berhasil mengidentifikasi kesenjangan infrastruktur internet pada satuan pendidikan di Indonesia dengan menggunakan algoritma klasterisasi K-Means. Berdasarkan analisis terhadap rasio jumlah sekolah yang memiliki akses internet terhadap total sekolah di setiap provinsi, ditemukan bahwa provinsi-provinsi di Indonesia dapat dikelompokkan ke dalam tiga klaster utama, yaitu provinsi dengan akses internet tinggi, sedang, dan rendah.

Hasil klasterisasi menunjukkan bahwa wilayah dengan tingkat pembangunan dan akses teknologi yang lebih baik cenderung tergabung dalam klaster tinggi, sedangkan wilayah-wilayah di bagian timur Indonesia lebih banyak masuk ke dalam klaster rendah. Temuan ini menguatkan fakta bahwa kesenjangan digital masih menjadi tantangan serius, terutama dalam pemerataan kualitas pendidikan.

Penggunaan algoritma K-Means terbukti efektif dalam mengelompokkan wilayah berdasarkan kesamaan karakteristik digital pendidikan. Hasil pengelompokan ini dapat dimanfaatkan sebagai landasan bagi pengambil kebijakan untuk meranca[[1]](#footnote-1)ng strategi pembangunan infrastruktur pendidikan digital yang lebih merata dan tepat sasaran. Dengan demikian, pemerataan akses internet di sektor pendidikan tidak hanya menjadi target pembangunan, tetapi juga langkah nyata menuju transformasi digital yang inklusif di seluruh wilayah Indonesia.

##### References

[1] S. R. Ramadhan, R. R. Wirdani, H. Delpina, and S. Nelwati, “Pendidikan Di Era Teknologi Informasi DanKomunikasi,” *J. Media Akad.*, vol. 3, no. 1, pp. 3031–5220, 2025.

[2] E. J. Theobald *et al.*, “Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, no. 12, pp. 6476–6483, 2020, doi: 10.1073/pnas.1916903117.

[3] A. Pieterse, *Who benefits from online education? How the implementation of technology in higher education can result in increasing inequalities in higher education institutions*, Faculty of Commerce, School of Economics, 2020. [Online]. Available: <http://hdl.handle.net/11427/32966>

[4] I. G. N. Santika, N. K. Suarni, and I. W. Lasmawan, “Analisis Perubahan Kurikulum Ditinjau Dari Kurikulum Sebagai Suatu Ide,” *J. Educ. Dev.*, vol. 10, no. 3, pp. 694–700, 2022.

[5] C. Caroline, “Jurnal Ilmiah Edukatif Meningkatkan Aksesibilitas Pendidikan Melalui Teknologi: Tantangan Dan Solusi Di Negara Berkembang,” *J. Ilm. Edukatif*, vol. 11, no. 01, pp. 224–231, 2025.

[6] H. Hasanbasri, P. Algusyairi, N. Nurhayuni, and M. Mudasir, “Sumber Daya Teknologi Terhadap Pelaksanaan Kurikulum di Era Digital,” *AL-MIKRAJ J. Stud. Islam dan Hum. (E-ISSN 2745-4584)*, vol. 4, no. 1, pp. 874–888, 2023, doi: 10.37680/almikraj.v4i1.4181.

[7] A. Feldmann *et al.*, “The Lockdown Effect: Implications of the COVID-19 Pandemic on Internet Traffic,” *Proc. ACM SIGCOMM Internet Meas. Conf. IMC*, pp. 1–18, 2020, doi: 10.1145/3419394.3423658.

[8] I. G. R. Wayan Indra Praekanata, Ni Putu Erna Surim Virnayanthi, Eliska Juliangkary, *Menelusuri Arah Pendidikan: Dinamika dan Inovasi Kurikulum di Indonesia*. PT. Sonpedia Publishing Indonesia, 2024, 2024.

[9] Y. D. Hadiyat, “Kesenjangan Digital di Indonesia (Studi Kasus di Kabupaten Wakatobi),” *Pekommas*, vol. 17, no. 2, pp. 81–90, 2014, [Online]. Available: http://download.garuda.kemdikb

[10] K. Firdaus, I. Universitas Muhammadiyah Sumatera Barat, M. Ritonga, and I. Universitas Muhammadiyah Sumatera Barat, “Peran Teknologi Dalam Mengatasi Krisis Pendidikan,” *J. Kepemimp. Dan Pengur. Sekol.*, vol. 9, no. 1, pp. 43–57, 2024, [Online]. Available: https://ejurnal.stkip-pessel.ac.id/index.php/kp/article/view/303

[11] P. Purwitasari and M. A. Raihan, “Strategies for Providing Base Transceiver Stations in Remote Areas to Support Community Access to Information with a Collaborative Governance Approach in Bombana Regency,” vol. 18, no. December, pp. 1–26, 2025.

[12] M. Ahmed, R. Seraj, and S. M. S. Islam, “The k-means algorithm: A comprehensive survey and performance evaluation,” *Electron.*, vol. 9, no. 8, pp. 1–12, 2020, doi: 10.3390/electronics9081295.

[13] N. Qamar and F. Syah Rezah, *Metode Penelitian Hukum: Doktrinal dan Non-Doktrinal*. CV. Social Politic Genius (SIGn), 2020.

[14] Z. Dobesova, “Evaluation of Orange data mining software and examples for lecturing machine learning tasks in geoinformatics,” *Comput. Appl. Eng. Educ.*, vol. 32, no. 4, pp. 1–18, 2024, doi: 10.1002/cae.22735.

[15] E. Retnoningsih and R. Pramudita, “Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python,” *Bina Insa. Ict J.*, vol. 7, no. 2, p. 156, 2020, doi: 10.51211/biict.v7i2.1422.

[16] M. Salihoun, “State of Art of Data Mining and Learning Analytics Tools in Higher Education,” *Int. J. Emerg. Technol. Learn.*, vol. 15, no. 21, pp. 58–76, 2020, doi: 10.3991/ijet.v15i21.16435.

[17] M. Alshamrani, ✉ Sager Alharthi, M. Helmi, and T. Alwadei, “Journal of Business and Management Studies Determinants of Employee Retention in Pharmaceutical Companies: Case of Saudi Arabia,” *Determ. Empl. Retent. Pharm. Co. Case Saudi Arab.*, no. 2709–0876, pp. 8–22, 2023, doi: 10.32996/jbms.

[18] I. Budiman, T. Prahasto, and Y. Christyono, “Data Clustering Menggunakan Metodologi CRISP-DM Untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma,” *J. Sist. Inf. Bisnis*, vol. 1, no. 3, pp. 15–16, 2014, doi: 10.21456/vol1iss3pp129-134.

[19] S. Naeem, A. Ali, S. Anam, and M. M. Ahmed, “An Unsupervised Machine Learning Algorithms: Comprehensive Review,” *Int. J. Comput. Digit. Syst.*, vol. 13, no. 1, pp. 911–921, 2023, doi: 10.12785/ijcds/130172.

[20] Y. Durachman, “Clustering Student Behavioral Patterns: A Data Mining Approach Using K-Means for Analyzing Study Hours, Attendance, and Tutoring Sessions in Educational Achievement,” *Artif. Intell. Learn.*, vol. 1, no. 1, pp. 35–53, 2025, doi: 10.63913/ail.v1i1.5.

[21] B. Chong, “K-means clustering algorithm: a brief review,” *Acad. J. Comput. Inf. Sci.*, vol. 4, no. 5, pp. 37–40, 2021, doi: 10.25236/ajcis.2021.040506.

[22] C. Schröer, F. Kruse, and J. M. Gómez, “A systematic literature review on applying CRISP-DM process model,” *Procedia Comput. Sci.*, vol. 181, no. 2019, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.

[23] A. M. Shimaoka, R. C. Ferreira, and A. Goldman, “The evolution of CRISP-DM for Data Science: Methods, Processes and Frameworks,” *SBC Rev. Comput. Sci.*, vol. 4, no. 1, pp. 28–43, 2024, doi: 10.5753/reviews.2024.3757.

[24] U. Kannengiesser and J. S. Gero, “Modelling the Design of Models: an Example Using Crisp-Dm,” *Proc. Des. Soc.*, vol. 3, no. JULY, pp. 2705–2714, 2023, doi: 10.1017/pds.2023.271.

[25] M. Elkabalawy, A. Al-Sakkaf, E. Mohammed Abdelkader, and G. Alfalah, “CRISP-DM-Based Data-Driven Approach for Building Energy Prediction Utilizing Indoor and Environmental Factors,” *Sustain.*, vol. 16, no. 17, 2024, doi: 10.3390/su16177249.

1. <https://github.com/eruu250/Dataset-Orange> (URL Github) [↑](#footnote-ref-1)