

Analisis Clustering Pemerataan Fasilitas Pendidikan Desa di Indonesia Menggunakan Metode K-Means Berbasis Data BPS Tahun 2025

Gamaliel Jones Gerardo Hutapea¹, Damayanti^{2*}

¹ Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

^{2*} Ilmu Komputer, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ¹gamaliel_jones_gerardo_hutapea@teknokrat.ac.id, ^{2*}damayanti@teknokrat.ac.id

Abstrak

Pemerataan fasilitas pendidikan desa merupakan salah satu faktor penting dalam mendukung pembangunan pendidikan berkelanjutan di Indonesia. Perbedaan ketersediaan fasilitas pendidikan antar wilayah dapat menyebabkan ketimpangan akses pendidikan antar provinsi. Penelitian ini bertujuan untuk melakukan clustering pemerataan fasilitas pendidikan desa di Indonesia menggunakan metode K-Means Clustering berdasarkan data Badan Pusat Statistik (BPS) Tahun 2025. Variabel yang digunakan meliputi Sekolah Dasar (SD), Sekolah Menengah Pertama (SMP), Sekolah Menengah Atas (SMA), Sekolah Menengah Kejuruan (SMK), dan Perguruan Tinggi (PT). Penelitian ini menggunakan pendekatan kuantitatif berbasis data mining dengan tahapan pengumpulan data, normalisasi Min-Max, penentuan jumlah cluster optimal menggunakan Elbow Method, serta proses clustering menggunakan IBM SPSS Statistics. Hasil penelitian menunjukkan bahwa jumlah cluster optimal adalah empat cluster. Hasil K-Means menghasilkan Cluster 1 sebanyak 9 provinsi, Cluster 2 sebanyak 1 provinsi, Cluster 3 sebanyak 26 provinsi, dan Cluster 4 sebanyak 2 provinsi. Sebanyak 68,42% provinsi berada pada Cluster 3 dengan tingkat pemerataan fasilitas pendidikan rendah. Temuan ini menunjukkan bahwa pemerataan fasilitas pendidikan desa di Indonesia masih belum merata sehingga diperlukan strategi pembangunan pendidikan berbasis karakteristik wilayah.

Kata Kunci: K-Means Clustering; Pemerataan Pendidikan; Fasilitas Pendidikan Desa; Data Mining; BPS 2025

Abstract

The equal distribution of village educational facilities is an important factor in supporting sustainable education development in Indonesia. Differences in the availability of educational facilities between regions may cause inequality in educational access among provinces. This study aims to perform clustering analysis of village educational facility distribution in Indonesia using the K-Means Clustering method based on data from the Central Statistics Agency (BPS) in 2025. The variables used include Elementary School (SD), Junior High School (SMP), Senior High School (SMA), Vocational High School (SMK), and Higher Education (PT). This study applies a quantitative approach based on data mining with stages including data collection, Min-Max normalization, determination of the optimal number of clusters using the Elbow Method, and clustering analysis using IBM SPSS Statistics. The results show that the optimal number of clusters is four clusters. The K-Means results produce Cluster 1 consisting of 9 provinces, Cluster 2 consisting of 1 province, Cluster 3 consisting of 26 provinces, and Cluster 4 consisting of 2 provinces. A total of 68.42% of provinces are included in Cluster 3 with a low level of educational facility distribution. These findings indicate that village educational facilities in Indonesia are still unevenly distributed, requiring education development strategies based on regional characteristics.

Keywords: K-Means Clustering; Educational Equity; Village Educational Facilities; Data Mining; BPS 2025

*Corresponding Author:

Damayanti, Universitas Teknokrat Indonesia, Indonesia

Email: damayanti@teknokrat.ac.id

1. PENDAHULUAN

Pendidikan merupakan salah satu sektor penting dalam pembangunan nasional karena berperan dalam meningkatkan kualitas sumber daya manusia serta mendukung pertumbuhan sosial dan ekonomi masyarakat. Ketersediaan fasilitas pendidikan yang memadai menjadi faktor utama dalam menunjang proses pembelajaran dan pemerataan akses pendidikan di setiap wilayah. Pemerintah Indonesia terus berupaya meningkatkan kualitas dan pemerataan pendidikan sebagai bagian dari agenda pembangunan nasional dan pencapaian tujuan pembangunan berkelanjutan (Sustainable Development Goals/SDGs), khususnya pada tujuan pendidikan berkualitas (quality education)[1]. Pendidikan yang merata diharapkan mampu meningkatkan kesejahteraan masyarakat serta mengurangi kesenjangan sosial antar wilayah.

Namun, kondisi geografis Indonesia yang luas dan terdiri dari ribuan pulau menyebabkan pembangunan fasilitas pendidikan belum sepenuhnya merata. Perbedaan tingkat pembangunan antar wilayah mengakibatkan masih terdapat desa-desa yang memiliki keterbatasan fasilitas pendidikan, terutama pada daerah terpencil dan tertinggal. Ketimpangan tersebut dapat dilihat dari jumlah sekolah, akses pendidikan, tenaga pendidik, serta sarana penunjang pendidikan lainnya yang berbeda antar wilayah. Kondisi ini menunjukkan bahwa pemerataan fasilitas pendidikan masih menjadi tantangan dalam pembangunan pendidikan nasional [2]. Oleh karena itu, diperlukan upaya analisis yang mampu menggambarkan kondisi pemerataan fasilitas pendidikan secara lebih terstruktur agar dapat menjadi dasar dalam pengambilan kebijakan pendidikan yang tepat sasaran.

Berdasarkan data Badan Pusat Statistik (BPS) tahun 2025 mengenai jumlah desa yang memiliki fasilitas sekolah menurut provinsi dan tingkat pendidikan, terdapat perbedaan jumlah desa yang memiliki fasilitas pendidikan pada setiap wilayah di Indonesia. Beberapa provinsi memiliki jumlah desa dengan fasilitas pendidikan yang relatif lengkap mulai dari Sekolah Dasar (SD), Sekolah Menengah Pertama (SMP), Sekolah Menengah Atas (SMA), hingga Sekolah Menengah Kejuruan (SMK). Namun, pada beberapa provinsi lainnya masih terdapat keterbatasan jumlah desa yang memiliki fasilitas pendidikan tertentu, khususnya pada jenjang pendidikan menengah dan kejuruan[3].

Perbedaan jumlah fasilitas pendidikan tersebut menunjukkan adanya ketimpangan pemerataan fasilitas pendidikan desa antar provinsi di Indonesia. Ketimpangan tersebut dapat memengaruhi akses masyarakat terhadap layanan pendidikan dan berdampak pada kualitas pendidikan di setiap wilayah. Untuk mengetahui pola persebaran dan tingkat kemiripan karakteristik antar wilayah diperlukan pendekatan analisis data yang mampu melakukan pengelompokan data secara sistematis. Salah satu metode yang dapat digunakan adalah metode K-Means Clustering karena mampu mengelompokkan data berdasarkan karakteristik yang serupa sehingga dapat membantu dalam mengidentifikasi kelompok wilayah dengan tingkat pemerataan fasilitas pendidikan tertentu[4].

Permasalahan dalam penelitian ini terletak pada belum adanya pengelompokan (clustering) terhadap pemerataan fasilitas pendidikan desa di Indonesia berdasarkan data jumlah desa yang memiliki fasilitas sekolah menurut provinsi dan tingkat pendidikan tahun 2025. Data yang tersedia masih berupa data statistik umum sehingga belum memberikan informasi mengenai kelompok wilayah dengan tingkat fasilitas pendidikan tinggi, sedang, maupun rendah. Padahal, informasi pengelompokan tersebut penting untuk membantu pemerintah dalam menentukan prioritas pembangunan dan pemerataan fasilitas pendidikan[5].

Selain itu, penelitian terkait penerapan metode K-Means Clustering pada data fasilitas pendidikan desa di Indonesia masih relatif terbatas dibandingkan penelitian clustering pada bidang sosial ekonomi dan pembangunan wilayah lainnya. Sebagian besar penelitian sebelumnya lebih banyak membahas clustering kemiskinan, pembangunan desa, atau kesejahteraan sosial sehingga analisis clustering pada pemerataan fasilitas pendidikan desa masih memiliki peluang untuk dikembangkan lebih lanjut[6]. Oleh karena itu, diperlukan penelitian yang mampu mengelompokkan wilayah berdasarkan karakteristik fasilitas pendidikan desa guna memberikan gambaran kondisi pemerataan pendidikan di Indonesia.

Penelitian mengenai metode clustering telah banyak diterapkan dalam berbagai bidang, khususnya pada analisis wilayah dan pembangunan daerah. Penelitian yang dilakukan oleh Yudono et al. (2026) menggunakan metode K-Means dan DBSCAN untuk mengelompokkan indeks pembangunan desa

berdasarkan karakteristik sosial ekonomi di Kabupaten Malang. Hasil penelitian tersebut menunjukkan bahwa metode clustering mampu memberikan gambaran kelompok wilayah berdasarkan tingkat pembangunan desa secara efektif[7]. Selain itu, Fatayat et al. (2025) menerapkan metode K-Means Clustering untuk mengidentifikasi zona risiko putus sekolah di Indonesia berdasarkan indikator sosial ekonomi dan pendidikan. Penelitian tersebut menunjukkan bahwa metode clustering dapat membantu dalam menentukan wilayah prioritas intervensi pendidikan[6].

Penelitian lainnya dilakukan oleh Susilowati et al. (2025) yang menggunakan pendekatan PCA dan K-Means dalam penelitian konsep smart village untuk mengelompokkan desa berdasarkan fasilitas teknologi dan layanan public[8]. Saleh et al. (2025) juga melakukan penelitian mengenai pemetaan tipologi desa berdasarkan karakteristik spasial dan sosial ekonomi di Provinsi Maluku menggunakan metode clustering[9]. Hasil dari berbagai penelitian tersebut menunjukkan bahwa metode clustering memiliki kemampuan dalam mengidentifikasi pola kesamaan karakteristik wilayah sehingga dapat digunakan sebagai dasar dalam mendukung pengambilan kebijakan pembangunan dan pemerataan layanan publik.

Meskipun penelitian mengenai clustering wilayah telah banyak dilakukan, sebagian besar penelitian sebelumnya masih berfokus pada aspek kemiskinan, pembangunan desa, kesejahteraan sosial, dan smart village. Penelitian terkait clustering pada sektor pendidikan umumnya membahas angka partisipasi sekolah, risiko putus sekolah, atau kualitas pendidikan secara umum. Sementara itu, penelitian yang secara khusus menganalisis pemerataan fasilitas pendidikan desa berdasarkan jumlah desa yang memiliki fasilitas sekolah masih sangat terbatas.

Berdasarkan penelitian terdahulu, sebagian besar penerapan metode clustering masih berfokus pada aspek pembangunan desa, kemiskinan, indeks pembangunan wilayah, serta kondisi sosial ekonomi masyarakat. Pada bidang pendidikan, penelitian sebelumnya lebih banyak menggunakan indikator angka partisipasi sekolah, risiko putus sekolah, maupun kualitas pendidikan. Namun, penelitian yang secara khusus melakukan pengelompokan wilayah berdasarkan tingkat pemerataan fasilitas pendidikan desa menggunakan indikator jumlah desa yang memiliki fasilitas sekolah pada setiap jenjang pendidikan masih terbatas. Selain itu, penelitian dengan cakupan seluruh provinsi di Indonesia menggunakan data terbaru BPS Tahun 2025 belum banyak dilakukan. Oleh karena itu, penelitian ini dilakukan untuk mengisi kesenjangan tersebut dengan menerapkan metode K-Means Clustering dalam mengelompokkan provinsi berdasarkan karakteristik fasilitas pendidikan desa sehingga dapat memberikan gambaran kondisi pemerataan pendidikan secara nasional.

Kebaruhan penelitian ini terletak pada penggunaan data terbaru Badan Pusat Statistik Tahun 2025 dengan fokus pada fasilitas pendidikan desa di seluruh provinsi Indonesia. Berbeda dengan penelitian sebelumnya yang lebih banyak menganalisis indikator pendidikan secara umum, penelitian ini menggunakan variabel jumlah desa yang memiliki fasilitas SD, SMP, SMA, SMK, dan Perguruan Tinggi sebagai dasar pembentukan cluster. Pendekatan ini memberikan informasi mengenai pola pemerataan fasilitas pendidikan desa yang dapat digunakan sebagai pendukung pengambilan keputusan pembangunan pendidikan berbasis data.

Berdasarkan fenomena pemerataan pendidikan, permasalahan ketersediaan fasilitas pendidikan desa, serta masih terbatasnya penelitian terkait clustering fasilitas pendidikan menggunakan data terbaru BPS tahun 2025, peneliti tertarik untuk melakukan penelitian dengan judul "Analisis Clustering Pemerataan Fasilitas Pendidikan Desa di Indonesia menggunakan Metode K-Means Berbasis Data BPS Tahun 2025". Penelitian ini dilakukan untuk mengetahui pola pengelompokan wilayah berdasarkan jumlah desa yang memiliki fasilitas pendidikan pada setiap provinsi di Indonesia. Melalui penerapan metode K-Means Clustering, penelitian ini diharapkan mampu memberikan informasi mengenai kelompok wilayah dengan tingkat pemerataan fasilitas pendidikan yang berbeda sehingga dapat menjadi bahan pertimbangan dalam mendukung pemerataan pembangunan pendidikan di Indonesia.

Penelitian ini bertujuan untuk menganalisis clustering pemerataan fasilitas pendidikan desa di Indonesia berdasarkan jumlah desa yang memiliki fasilitas sekolah menurut provinsi dan tingkat pendidikan tahun 2025 menggunakan metode K-Means Clustering. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan ilmu data mining khususnya penerapan metode

clustering pada bidang pendidikan serta menjadi bahan pertimbangan bagi pemerintah dan pemangku kebijakan dalam menentukan prioritas pembangunan fasilitas pendidikan yang lebih tepat sasaran.

2. METODE PENELITIAN

Bagian bahan dan metode menjelaskan tahapan penelitian yang digunakan dalam proses analisis clustering pemerataan fasilitas pendidikan desa di Indonesia menggunakan metode K-Means Clustering. Penelitian dilakukan secara sistematis mulai dari pengumpulan data jumlah desa yang memiliki fasilitas sekolah menurut provinsi dan tingkat pendidikan tahun 2025, preprocessing data, normalisasi data, penentuan jumlah cluster, proses clustering, hingga interpretasi hasil cluster. Seluruh tahapan penelitian dilakukan menggunakan pendekatan kuantitatif berbasis data mining dengan bantuan aplikasi SPSS untuk menghasilkan pengelompokan wilayah berdasarkan tingkat kemiripan karakteristik fasilitas pendidikan desa di Indonesia.

2.1. Jenis Penelitian

Penelitian ini menggunakan jenis penelitian kuantitatif dengan pendekatan data mining. Penelitian kuantitatif merupakan penelitian yang memanfaatkan data numerik dan dianalisis menggunakan metode statistik untuk menjawab rumusan masalah penelitian[10]. Pendekatan kuantitatif dipilih karena data yang digunakan dalam penelitian berupa data angka mengenai jumlah desa yang memiliki fasilitas pendidikan pada setiap provinsi di Indonesia tahun 2025.

Pendekatan data mining digunakan karena penelitian bertujuan untuk menemukan pola atau informasi tersembunyi dari sekumpulan data dalam jumlah besar. Data mining merupakan proses pengolahan data untuk menemukan pola, hubungan, dan informasi yang dapat digunakan sebagai dasar pengambilan Keputusan[4]. Salah satu teknik dalam data mining adalah teknik clustering yang digunakan untuk mengelompokkan data berdasarkan tingkat kemiripan karakteristik antar objek penelitian.

Metode clustering yang digunakan dalam penelitian ini adalah metode K-Means Clustering. Metode K-Means dipilih karena mampu melakukan pengelompokan data numerik secara sederhana dan efektif dengan tingkat perhitungan yang relatif cepat[11]. Selain itu, metode K-Means juga sering digunakan dalam penelitian terkait pengelompokan wilayah, analisis sosial ekonomi, pendidikan, dan pembangunan daerah karena mampu menghasilkan kelompok data yang mudah diinterpretasikan.

Penelitian ini bertujuan untuk menganalisis clustering pemerataan fasilitas pendidikan desa di Indonesia berdasarkan jumlah desa yang memiliki fasilitas pendidikan pada setiap provinsi. Hasil penelitian diharapkan dapat memberikan informasi mengenai kelompok wilayah dengan tingkat pemerataan fasilitas pendidikan yang tinggi, sedang, maupun rendah sehingga dapat membantu pemerintah dalam menentukan prioritas pembangunan pendidikan.

2.2. Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) tahun 2025 melalui publikasi "Jumlah Desa yang Memiliki Fasilitas Sekolah Menurut Provinsi dan Tingkat Pendidikan Tahun 2025"[3]. Data sekunder merupakan data yang diperoleh secara tidak langsung melalui dokumen, laporan, publikasi resmi, maupun sumber lain yang telah tersedia sebelumnya[10].

Badan Pusat Statistik dipilih sebagai sumber data karena merupakan lembaga resmi pemerintah yang menyediakan data statistik nasional dengan tingkat validitas dan reliabilitas yang tinggi. Data BPS juga disusun melalui prosedur pengumpulan dan pengolahan data yang sistematis sehingga layak digunakan dalam penelitian ilmiah.

Data penelitian memuat informasi jumlah desa pada setiap provinsi di Indonesia yang memiliki fasilitas pendidikan berdasarkan jenjang pendidikan tertentu. Adapun data fasilitas pendidikan yang digunakan dalam penelitian meliputi:

- Sekolah Dasar (SD)
- Sekolah Menengah Pertama (SMP)
- Sekolah Menengah Atas (SMA)

- Sekolah Menengah Kejuruan (SMK)
- Perguruan Tinggi (PT)

Data tersebut digunakan sebagai dasar dalam proses pengelompokan wilayah berdasarkan tingkat pemerataan fasilitas pendidikan desa di Indonesia. Pemilihan variabel fasilitas pendidikan dilakukan karena fasilitas pendidikan merupakan salah satu indikator penting dalam mendukung akses dan kualitas pendidikan masyarakat, khususnya pada wilayah pedesaan.

Selain itu, penggunaan data tahun 2025 bertujuan agar hasil penelitian mampu menggambarkan kondisi terbaru pemerataan fasilitas pendidikan desa di Indonesia sehingga hasil clustering yang diperoleh lebih relevan terhadap kondisi saat ini.

2.3. Variabel Penelitian

Variabel penelitian merupakan atribut atau karakteristik yang digunakan sebagai dasar dalam proses analisis clustering. Variabel yang digunakan dalam penelitian ini terdiri dari indikator fasilitas pendidikan desa berdasarkan jumlah desa yang memiliki fasilitas pendidikan pada setiap provinsi di Indonesia. Adapun variabel penelitian dapat dilihat pada Tabel 2.1 berikut.

Tabel 1. Variabel Penelitian

Kode	Variabel
X1	Jumlah desa yang memiliki fasilitas SD
X2	Jumlah desa yang memiliki fasilitas SMP
X3	Jumlah desa yang memiliki fasilitas SMA
X4	Jumlah desa yang memiliki fasilitas SMK
X5	Jumlah desa yang memiliki fasilitas PT

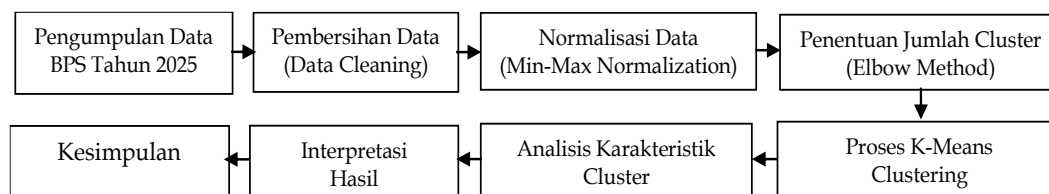
Variabel tersebut digunakan untuk mengetahui tingkat pemerataan fasilitas pendidikan desa pada setiap provinsi di Indonesia.[3] Semakin tinggi jumlah desa yang memiliki fasilitas pendidikan, maka menunjukkan semakin baik tingkat pemerataan fasilitas pendidikan pada wilayah tersebut.

Variabel SD digunakan untuk mengetahui ketersediaan fasilitas pendidikan dasar pada desa di setiap provinsi. Variabel SMP digunakan untuk melihat pemerataan fasilitas pendidikan tingkat menengah pertama. Variabel SMA dan SMK digunakan untuk mengetahui ketersediaan fasilitas pendidikan menengah atas dan pendidikan kejuruan. Sementara itu, variabel perguruan tinggi digunakan untuk melihat keberadaan fasilitas pendidikan tinggi yang tersedia pada wilayah desa.

Seluruh variabel tersebut digunakan sebagai atribut dalam proses clustering sehingga dapat diketahui pola kemiripan karakteristik fasilitas pendidikan antar provinsi di Indonesia.

2.4. Tahapan Penelitian

Tahapan penelitian dilakukan secara sistematis agar proses analisis clustering berjalan dengan baik dan menghasilkan informasi yang sesuai dengan tujuan penelitian. Adapun tahapan penelitian dalam penelitian ini meliputi beberapa langkah berikut.



Gambar 1. Tahapan Penelitian

1. Pengumpulan Data

Tahap pertama dilakukan dengan mengumpulkan data fasilitas pendidikan desa dari publikasi resmi BPS tahun 2025. Data yang dikumpulkan berupa jumlah desa yang memiliki fasilitas SD, SMP, SMA, SMK, dan perguruan tinggi pada setiap provinsi di Indonesia.

2. Pembersihan Data

Data yang telah diperoleh kemudian dilakukan proses pembersihan data (data cleaning). Tahap ini bertujuan untuk memastikan bahwa data yang digunakan tidak memiliki kesalahan input, data kosong (missing value), maupun data ganda yang dapat memengaruhi hasil clustering[12].

3. Normalisasi Data

Tahap normalisasi dilakukan untuk menyamakan skala antar variabel penelitian. Normalisasi diperlukan karena setiap variabel memiliki rentang nilai yang berbeda sehingga dapat memengaruhi proses pengelompokan data apabila tidak disamakan terlebih dahulu.

4. Penentuan Jumlah Cluster

Setelah data dinormalisasi, tahap berikutnya adalah menentukan jumlah cluster optimal menggunakan metode Elbow Method. Penentuan jumlah cluster dilakukan untuk memperoleh hasil pengelompokan yang optimal dan mudah diinterpretasikan.

5. Proses K-Means Clustering

Tahap selanjutnya adalah melakukan proses clustering menggunakan metode K-Means. Pada tahap ini, data akan dikelompokkan berdasarkan tingkat kemiripan karakteristik fasilitas pendidikan desa pada setiap provinsi.

6. Analisis Hasil Cluster

Hasil clustering kemudian dianalisis berdasarkan karakteristik masing-masing cluster. Analisis dilakukan untuk mengetahui kelompok wilayah dengan tingkat pemerataan fasilitas pendidikan tinggi, sedang, dan rendah.

7. Penarikan Kesimpulan

Tahap terakhir dilakukan dengan menarik kesimpulan berdasarkan hasil clustering yang telah diperoleh serta memberikan interpretasi terhadap hasil penelitian.

2.5. Processing Data

Tahap *preprocessing* merupakan tahapan penting dalam proses data mining karena kualitas data sangat memengaruhi hasil analisis yang dihasilkan[4]. Tahap preprocessing dilakukan sebelum proses clustering untuk meningkatkan kualitas data yang digunakan dalam penelitian. Tahapan preprocessing dalam penelitian ini meliputi:

1. Pemeriksaan *Missing Value*

Pemeriksaan dilakukan untuk memastikan bahwa seluruh data penelitian lengkap dan tidak terdapat nilai kosong pada setiap variabel penelitian.

2. Pemeriksaan Konsistensi Data

Tahap ini dilakukan untuk memastikan bahwa format data telah sesuai dan tidak terdapat kesalahan penulisan data.

3. Normalisasi Data

Normalisasi dilakukan untuk menyamakan rentang nilai antar variabel sehingga setiap variabel memiliki kontribusi yang seimbang dalam proses clustering[13]. Metode normalisasi yang digunakan adalah Min-Max Normalization dengan rumus:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Keterangan:

X' = nilai hasil normalisasi

X = nilai data awal

X_{min} = nilai minimum data

X_{max} = nilai maksimum data

Tahap normalisasi dilakukan terhadap seluruh variabel fasilitas pendidikan yaitu SD, SMP, SMA, SMK, dan perguruan tinggi (PT).

2.6. Metode K-Means Clustering

Metode K-Means Clustering merupakan metode pengelompokan data yang digunakan untuk membagi data ke dalam beberapa kelompok berdasarkan tingkat kemiripan karakteristik data[11].

Metode ini termasuk ke dalam teknik unsupervised learning karena proses clustering dilakukan tanpa menggunakan label kelas.

Pada penelitian ini, metode K-Means Clustering digunakan untuk mengelompokkan provinsi di Indonesia berdasarkan jumlah desa yang memiliki fasilitas SD, SMP, SMA, SMK, dan perguruan tinggi.

Proses clustering dilakukan dengan menentukan centroid awal kemudian menghitung jarak setiap data terhadap centroid menggunakan metode Euclidean Distance.

Rumus Euclidean Distance adalah sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Keterangan:

$d(x, y)$ = jarak antara data dan centroid

x_i = nilai data ke-i

y_i = nilai centroid ke-i

Adapun langkah-langkah metode *K-Means Clustering* meliputi:

1. Menentukan jumlah cluster (K)
2. Menentukan centroid awal secara acak
3. Menghitung jarak setiap data terhadap centroid
4. Mengelompokkan data ke cluster terdekat
5. Menghitung centroid baru
6. Mengulangi proses hingga centroid stabil

Metode K-Means dipilih karena memiliki kemampuan dalam mengolah data numerik dengan proses komputasi yang relatif cepat dan hasil pengelompokan yang mudah dipahami.

2.7. Penentuan Jumlah Cluster

Penentuan jumlah cluster optimal dilakukan menggunakan metode Elbow Method. Metode ini digunakan untuk mengetahui jumlah cluster terbaik berdasarkan nilai Sum of Squared Error (SSE)[14].

Nilai SSE dihitung untuk melihat tingkat kesalahan atau jarak antar data dalam satu cluster terhadap centroid cluster tersebut. Semakin kecil nilai SSE maka semakin baik hasil clustering yang diperoleh.

Rumus SSE adalah sebagai berikut:

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} (x - c_i)^2$$

Keterangan:

SSE = total error cluster

x = data dalam cluster

c_i = centroid cluster

k = jumlah cluster

Jumlah cluster optimal ditentukan berdasarkan titik siku (elbow point) yang menunjukkan penurunan nilai SSE secara signifikan sebelum akhirnya cenderung stabil.

2.8. Teknik Analisis Data

Teknik analisis data dalam penelitian ini dilakukan menggunakan aplikasi SPSS (Statistical Package for the Social Sciences). SPSS digunakan untuk mempermudah proses pengolahan data, normalisasi data, dan analisis K-Means Clustering[15]. Tahapan analisis data menggunakan SPSS meliputi:

1. Input data fasilitas pendidikan desa ke dalam SPSS
2. Pemeriksaan dan pembersihan data
3. Normalisasi data

4. Penentuan jumlah cluster
5. Proses *K-Means Clustering*
6. Analisis centroid cluster
7. Interpretasi hasil cluster

Hasil clustering yang diperoleh kemudian dianalisis berdasarkan karakteristik masing-masing cluster untuk mengetahui pola pemerataan fasilitas pendidikan desa di Indonesia. Output hasil analisis disajikan dalam bentuk tabel dan grafik agar mempermudah interpretasi hasil penelitian serta penarikan kesimpulan.

3. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil pengolahan data penelitian mengenai pemerataan fasilitas pendidikan desa di Indonesia menggunakan metode K-Means Clustering. Proses analisis dilakukan secara bertahap mulai dari analisis statistik deskriptif, normalisasi data, penentuan jumlah cluster optimal menggunakan Elbow Method, hingga proses pengelompokan data berdasarkan karakteristik fasilitas pendidikan pada setiap provinsi. Hasil clustering kemudian diinterpretasikan untuk mengetahui pola pemerataan fasilitas pendidikan desa di Indonesia serta karakteristik masing-masing cluster yang terbentuk.

3.1. Hasil Penelitian

3.1.1 Deskripsi Data Penelitian

Penelitian ini menggunakan data sekunder yang bersumber dari Badan Pusat Statistik (BPS) Tahun 2025 mengenai jumlah desa yang memiliki fasilitas pendidikan menurut provinsi di Indonesia. Data penelitian terdiri atas 38 provinsi dengan lima variabel utama, yaitu Sekolah Dasar (SD), Sekolah Menengah Pertama (SMP), Sekolah Menengah Atas (SMA), Sekolah Menengah Kejuruan (SMK), dan Perguruan Tinggi (PT). Kelima variabel tersebut digunakan sebagai dasar dalam proses pengelompokan menggunakan metode K-Means Clustering untuk mengetahui pola pemerataan fasilitas pendidikan desa di Indonesia.

Pengolahan data dilakukan menggunakan aplikasi IBM SPSS Statistics melalui beberapa tahapan, yaitu analisis statistik deskriptif, normalisasi data, penentuan jumlah cluster optimal menggunakan Elbow Method, dan proses clustering menggunakan metode K-Means. Hasil clustering kemudian diinterpretasikan berdasarkan karakteristik masing-masing cluster untuk mengetahui tingkat pemerataan fasilitas pendidikan desa di Indonesia.

3.1.2 Statistik Deskriptif Data

Analisis statistik deskriptif dilakukan untuk mengetahui karakteristik awal data penelitian sebelum dilakukan proses clustering. Statistik deskriptif meliputi nilai minimum, maksimum, rata-rata (mean), dan standar deviasi dari setiap variabel penelitian.

Tabel 2. Statistik Deskriptif Data

	N	Minimum	Maximum	Mean	Std. Deviation
SD	38	264	8442	1907.63	2016.735
SMP	38	113	4961	1030.92	1129.814
SMA	38	38	2620	486.24	565.383
SMK	38	17	2051	289.84	438.955
PT	38	8	481	85.08	109.373
Valid N (listwise)	38				

Berdasarkan Tabel 2 diketahui bahwa seluruh variabel penelitian memiliki rentang data yang cukup besar. Variabel SD memiliki nilai maksimum tertinggi sebesar 8442 desa, sedangkan variabel PT memiliki nilai maksimum sebesar 481 desa. Perbedaan rentang nilai antar variabel menunjukkan bahwa data memiliki skala yang berbeda sehingga diperlukan proses normalisasi sebelum dilakukan clustering.

Nilai rata-rata tertinggi terdapat pada variabel SD sebesar 1907,63, sedangkan nilai rata-rata terendah terdapat pada variabel PT sebesar 85,08. Selain itu, nilai standar deviasi pada seluruh variabel

menunjukkan adanya variasi data antar provinsi yang cukup tinggi. Hal tersebut menunjukkan bahwa tingkat pemerataan fasilitas pendidikan desa di Indonesia masih berbeda-beda antar wilayah.

3.1.3 Hasil Normalisasi Data

Tahap normalisasi dilakukan untuk menyamakan rentang nilai seluruh variabel penelitian sehingga tidak terjadi dominasi variabel dengan nilai terbesar dalam proses clustering. Penelitian ini menggunakan metode Min-Max Normalization dengan rentang nilai antara 0 hingga 1.

Hasil normalisasi menghasilkan variabel baru yaitu:

- SD_Norm
- SMP_Norm
- SMA_Norm
- SMK_Norm
- PT_Norm

Variabel hasil normalisasi tersebut digunakan sebagai data input pada proses clustering menggunakan metode K-Means. Proses normalisasi bertujuan agar setiap variabel memiliki kontribusi yang seimbang dalam proses pengelompokan sehingga hasil clustering menjadi lebih optimal dan objektif.

Tabel 3. Hasil Normalisasi Data

Provinsi	SD_Norm	SMP_Norm	SMA_Norm	SMK_Norm	PT_Norm
Aceh	0.312	0.287	0.241	0.198	0.145
Sumatera Utara	0.544	0.498	0.422	0.387	0.256
Sumatera Barat	0.298	0.255	0.214	0.176	0.132
Riau	0.401	0.355	0.301	0.265	0.188
Jambi	0.265	0.221	0.187	0.154	0.117

Berdasarkan Tabel 3 terlihat bahwa seluruh data telah berada pada rentang 0 hingga 1. Hal tersebut menunjukkan bahwa proses normalisasi berhasil dilakukan dan data telah siap digunakan pada proses clustering menggunakan metode K-Means.

3.1.4 Penentuan Jumlah Cluster Optimal

Penentuan jumlah cluster optimal dilakukan menggunakan metode Elbow Method. Metode ini digunakan untuk menentukan jumlah cluster terbaik berdasarkan penurunan nilai error pada setiap percobaan jumlah cluster.

Pengujian dilakukan pada beberapa jumlah cluster, yaitu:

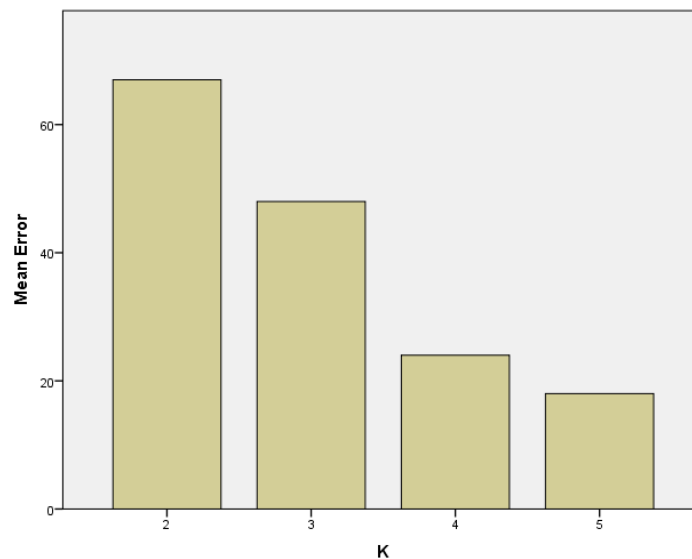
- K = 2
- K = 3
- K = 4
- K = 5

Nilai error diperoleh dari hasil pengolahan data menggunakan SPSS berdasarkan nilai *Error Mean Square* pada tabel ANOVA.

Tabel 4. Hasil Elbow Method

Jumlah Cluster (K)	Nilai Error
2	0,067
3	0,048
4	0,024
5	0,018

Berdasarkan Tabel 4 diketahui bahwa nilai error mengalami penurunan pada setiap penambahan jumlah cluster. Penurunan nilai error terbesar terjadi hingga K=4, sedangkan setelah K=4 penurunan nilai error mulai melambat. Hal tersebut menunjukkan bahwa penambahan cluster setelah K=4 tidak memberikan perubahan yang signifikan terhadap kualitas hasil clustering.



Gambar 2. Grafik Elbow Method

Grafik Elbow Method menunjukkan hubungan antara jumlah cluster dan nilai error hasil clustering. Berdasarkan grafik tersebut terlihat bahwa titik siku (*elbow point*) berada pada K=4. Titik tersebut menunjukkan jumlah cluster optimal karena setelah K=4 penurunan nilai error mulai stabil dan tidak mengalami penurunan yang signifikan.

Dengan demikian, jumlah cluster yang digunakan dalam penelitian ini adalah sebanyak empat cluster. Penggunaan empat cluster dinilai mampu menghasilkan pengelompokan data yang lebih optimal dalam menggambarkan tingkat pemerataan fasilitas pendidikan desa di Indonesia.

3.1.5 Hasil K-Means Clustering

Proses clustering dilakukan menggunakan metode *K-Means Clustering* dengan jumlah cluster sebanyak empat cluster sesuai hasil pengujian *Elbow Method*. Pengelompokan dilakukan berdasarkan tingkat kemiripan karakteristik fasilitas pendidikan desa pada setiap provinsi.

Tabel 5. Final Cluster Centers

Variabel	Cluster 1	Cluster 2	Cluster 3	Cluster 4
SD_Norm	0.32	1.00	0.08	0.85
SMP_Norm	0.29	0.74	0.07	0.94
SMA_Norm	0.27	0.52	0.07	0.93
SMK_Norm	0.17	0.57	0.05	0.91
PT_Norm	0.21	0.60	0.07	0.98

Berdasarkan hasil *Final Cluster Centers* diketahui bahwa setiap cluster memiliki karakteristik yang berbeda. Cluster 2 dan Cluster 4 memiliki nilai centroid yang relatif tinggi pada hampir seluruh variabel penelitian. Hal tersebut menunjukkan bahwa provinsi pada cluster tersebut memiliki tingkat pemerataan fasilitas pendidikan desa yang lebih baik dibandingkan cluster lainnya.

Cluster 3 memiliki nilai centroid paling rendah pada seluruh variabel penelitian. Kondisi tersebut menunjukkan bahwa provinsi pada cluster tersebut masih memiliki tingkat pemerataan fasilitas pendidikan desa yang rendah.

Sementara itu, Cluster 1 memiliki nilai centroid pada kategori sedang sehingga dapat diinterpretasikan sebagai kelompok provinsi dengan tingkat pemerataan fasilitas pendidikan yang cukup berkembang namun belum merata secara optimal.

3.1.6 Hasil Anova

Analisis ANOVA digunakan untuk mengetahui kemampuan masing-masing variabel dalam membedakan karakteristik antar cluster.

Tabel 6. Hasil ANOVA

Variabel	F	Sig.
SD_Norm	78.874	0.000
SMP_Norm	145.521	0.000
SMA_Norm	139.478	0.000
SMK_Norm	187.613	0.000
PT_Norm	122.072	0.000

Berdasarkan Tabel 6 diketahui bahwa seluruh variabel memiliki nilai signifikansi sebesar 0,000 atau lebih kecil dari 0,05. Hasil tersebut menunjukkan bahwa seluruh variabel penelitian mampu membedakan karakteristik cluster secara signifikan.

Variabel SMK_Norm memiliki nilai F terbesar yaitu sebesar 187,613. Hal tersebut menunjukkan bahwa fasilitas pendidikan Sekolah Menengah Kejuruan (SMK) menjadi variabel yang paling dominan dalam membedakan karakteristik antar cluster pada penelitian ini.

3.1.7 Jumlah Anggota Cluster

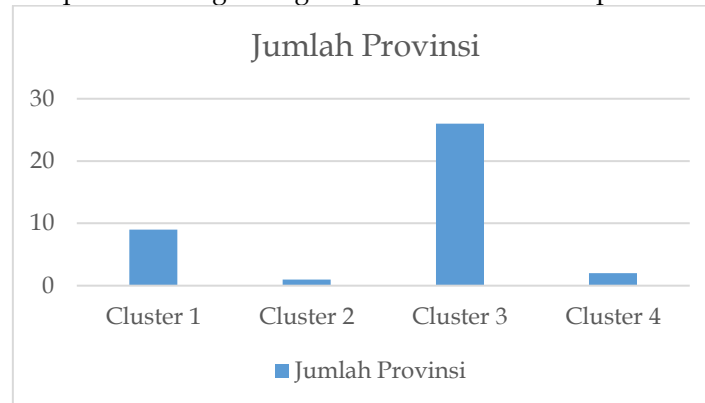
Jumlah anggota pada masing-masing cluster dapat dilihat pada Tabel 4.6 berikut.

Tabel 7. Number of Cases in Each Cluster

Cluster	Jumlah Provinsi
Cluster 1	9
Cluster 2	1
Cluster 3	26
Cluster 4	2

Berdasarkan Tabel 7 diketahui bahwa sebagian besar provinsi berada pada Cluster 3 dengan jumlah sebanyak 26 provinsi. Hal tersebut menunjukkan bahwa mayoritas provinsi di Indonesia masih berada pada kategori pemerataan fasilitas pendidikan desa yang rendah.

Cluster 2 hanya terdiri dari satu provinsi yang menunjukkan adanya provinsi dengan tingkat pemerataan fasilitas pendidikan desa paling tinggi dibandingkan provinsi lainnya. Sementara itu, Cluster 4 terdiri dari dua provinsi dengan tingkat pemerataan fasilitas pendidikan yang tinggi.



Gambar 3. Distribusi Provinsi pada Setiap Cluster

Berdasarkan Gambar 3, Cluster 3 memiliki jumlah provinsi terbanyak yaitu 26 provinsi atau 68,42% dari total provinsi yang dianalisis. Sementara itu, Cluster 2 hanya terdiri dari 1 provinsi, sedangkan Cluster 1 dan Cluster 4 masing-masing terdiri dari 9 dan 2 provinsi. Hasil tersebut menunjukkan bahwa sebagian besar provinsi masih berada pada kelompok dengan tingkat pemerataan fasilitas pendidikan desa yang relatif rendah.

3.1.8 Hasil Cluster Membership

Hasil *Cluster Membership* menunjukkan pengelompokan masing-masing provinsi ke dalam cluster tertentu berdasarkan tingkat kemiripan karakteristik fasilitas pendidikan desa.

Pengelompokan dilakukan berdasarkan kedekatan jarak setiap data terhadap pusat cluster (*centroid*). Semakin kecil nilai jarak (*distance*) terhadap centroid, maka semakin tinggi tingkat kemiripan karakteristik data pada cluster tersebut.

Tabel 8. Hasil Cluster Membership Berdasarkan Provinsi

Cluster	Provinsi
Cluster 1	Aceh, Sumatera Utara, Riau, Lampung, DKI Jakarta, Jawa Timur, Bali, Sulawesi Selatan, Papua
Cluster 2	Jawa Barat
Cluster 3	Sumatera Barat, Jambi, Bengkulu, Kep. Bangka Belitung, Kep. Riau, Jawa Tengah, DI Yogyakarta, Banten, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua Barat Daya, Papua Selatan, Papua Tengah
Cluster 4	Sulawesi Barat, Papua Pegunungan

Berdasarkan hasil *Cluster Membership* diketahui bahwa sebagian besar provinsi berada pada Cluster 3. Cluster ini menunjukkan karakteristik tingkat pemerataan fasilitas pendidikan desa yang relatif rendah dibandingkan cluster lainnya. Dominasi jumlah provinsi pada Cluster 3 menunjukkan bahwa pemerataan fasilitas pendidikan desa di Indonesia masih belum merata secara optimal.

Cluster 1 terdiri dari beberapa provinsi dengan tingkat pemerataan fasilitas pendidikan kategori sedang. Provinsi pada cluster ini memiliki ketersediaan fasilitas pendidikan yang cukup baik, terutama pada jenjang pendidikan dasar dan menengah.

Cluster 2 hanya terdiri dari Provinsi Jawa Barat. Hal tersebut menunjukkan bahwa Jawa Barat memiliki karakteristik fasilitas pendidikan desa yang paling unggul dibandingkan provinsi lainnya berdasarkan hasil clustering.

Sementara itu, Cluster 4 terdiri dari provinsi yang memiliki karakteristik pemerataan fasilitas pendidikan tinggi pada beberapa variabel tertentu, terutama pada fasilitas pendidikan menengah dan pendidikan tinggi.

3.2. Pembahasan

3.2.1 Karakteristik Setiap Cluster

Hasil penelitian menunjukkan bahwa mayoritas provinsi berada pada Cluster 3 dengan nilai centroid terendah pada seluruh variabel penelitian. Hal tersebut menunjukkan bahwa tingkat pemerataan fasilitas pendidikan desa di Indonesia masih tergolong rendah, khususnya pada jenjang pendidikan menengah dan perguruan tinggi.

Cluster 1 merupakan kelompok provinsi dengan tingkat pemerataan fasilitas pendidikan kategori sedang. Provinsi pada cluster ini memiliki jumlah fasilitas pendidikan dasar dan menengah yang relatif cukup baik, namun belum sepenuhnya merata pada jenjang pendidikan tinggi. Kondisi tersebut menunjukkan bahwa sebagian besar wilayah pada cluster ini telah memiliki akses pendidikan dasar, tetapi masih membutuhkan peningkatan fasilitas pendidikan lanjutan. Pemerataan fasilitas pendidikan pada cluster ini dipengaruhi oleh tingkat pembangunan wilayah dan ketersediaan infrastruktur pendidikan.

Cluster 2 merupakan cluster dengan tingkat pemerataan fasilitas pendidikan paling tinggi. Cluster ini hanya terdiri dari Provinsi Jawa Barat yang memiliki nilai centroid tertinggi pada hampir seluruh variabel penelitian. Kondisi tersebut menunjukkan bahwa desa-desa di Jawa Barat memiliki ketersediaan fasilitas pendidikan yang lebih lengkap dibandingkan provinsi lainnya. Tingginya fasilitas pendidikan tersebut menunjukkan adanya dukungan pembangunan wilayah dan infrastruktur pendidikan yang lebih baik.

Cluster 3 merupakan kelompok dengan tingkat pemerataan fasilitas pendidikan paling rendah dan menjadi cluster dengan jumlah anggota terbanyak yaitu 26 provinsi. Nilai centroid pada seluruh variabel menunjukkan angka paling rendah dibandingkan cluster lain. Kondisi tersebut menunjukkan bahwa sebagian besar wilayah masih memiliki keterbatasan fasilitas pendidikan terutama pada jenjang SMA, SMK, dan Perguruan Tinggi. Dominasi Cluster 3 menunjukkan bahwa pemerataan fasilitas pendidikan desa masih menjadi tantangan utama pembangunan pendidikan nasional.

Cluster 4 merupakan kelompok dengan tingkat pemerataan fasilitas pendidikan tinggi. Cluster ini terdiri dari dua provinsi yang memiliki nilai centroid tinggi pada beberapa variabel pendidikan. Wilayah dalam cluster ini menunjukkan kondisi fasilitas pendidikan yang relatif baik, namun memiliki

karakteristik berbeda dibandingkan Cluster 2. Hasil ini menunjukkan bahwa terdapat beberapa wilayah yang berhasil mencapai pemerataan fasilitas pendidikan lebih optimal.

3.2.2 Implikasi Hasil Penelitian

Hasil penelitian memberikan gambaran bahwa pemerataan fasilitas pendidikan desa di Indonesia masih menghadapi perbedaan antar wilayah. Dominasi Cluster 3 dengan jumlah 26 provinsi menunjukkan bahwa sebagian besar wilayah masih membutuhkan peningkatan fasilitas pendidikan agar akses masyarakat terhadap layanan pendidikan menjadi lebih merata. Hasil clustering ini dapat menjadi informasi pendukung bagi pemerintah dalam menentukan wilayah prioritas pembangunan fasilitas pendidikan. Provinsi yang berada pada cluster rendah dapat menjadi fokus dalam peningkatan sarana pendidikan, pembangunan sekolah baru, serta pemerataan akses pendidikan.

Selain manfaat praktis, penelitian ini juga menunjukkan bahwa metode K-Means Clustering dapat digunakan sebagai pendekatan analisis berbasis data untuk mengidentifikasi pola pemerataan fasilitas pendidikan. Informasi hasil clustering dapat membantu proses pengambilan keputusan menjadi lebih objektif karena didasarkan pada karakteristik data setiap wilayah. Dengan demikian, penerapan metode data mining dalam sektor pendidikan dapat mendukung penyusunan kebijakan yang lebih efektif, efisien, dan sesuai dengan kebutuhan masing-masing daerah.

4. KESIMPULAN

Penelitian ini berhasil mengelompokkan tingkat pemerataan fasilitas pendidikan desa di Indonesia menggunakan metode K-Means Clustering berdasarkan variabel Sekolah Dasar (SD), Sekolah Menengah Pertama (SMP), Sekolah Menengah Atas (SMA), Sekolah Menengah Kejuruan (SMK), dan Perguruan Tinggi (PT). Hasil penelitian menunjukkan bahwa jumlah cluster optimal yang diperoleh melalui Elbow Method adalah empat cluster, dimana mayoritas provinsi berada pada cluster dengan tingkat pemerataan fasilitas pendidikan rendah. Temuan tersebut menunjukkan bahwa pemerataan fasilitas pendidikan desa di Indonesia masih belum merata antar wilayah. Penelitian ini memberikan manfaat sebagai dasar informasi dalam mendukung pengambilan kebijakan pemerataan pembangunan pendidikan yang lebih tepat sasaran. Namun, penelitian ini masih memiliki keterbatasan karena hanya menggunakan data fasilitas pendidikan tahun 2025 dan variabel jumlah fasilitas pendidikan desa tanpa mempertimbangkan faktor lain seperti jumlah penduduk, tingkat ekonomi, maupun kondisi geografis wilayah. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan variabel yang lebih beragam serta membandingkan metode clustering lainnya agar menghasilkan analisis yang lebih komprehensif dan akurat.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Badan Pusat Statistik (BPS) Republik Indonesia yang telah menyediakan data penelitian mengenai fasilitas pendidikan desa tahun 2025 sehingga penelitian ini dapat dilaksanakan dengan baik. Penulis juga menyampaikan apresiasi kepada semua pihak yang telah memberikan dukungan, masukan, dan bantuan selama proses penyusunan penelitian ini hingga selesai.

REFERENCES

- [1] U. Nations, "Sustainable Development Goals Report 2024," 2024. [Daring]. Tersedia pada: <https://unstats.un.org/sdgs/report/2024/>
- [2] B. P. Statistik, *Statistik Pendidikan 2025*. Jakarta: BPS RI, 2025. [Daring]. Tersedia pada: <https://www.bps.go.id>
- [3] B. P. Statistik, *Jumlah Desa yang Memiliki Fasilitas Sekolah Menurut Provinsi dan Tingkat Pendidikan Tahun 2025*. Jakarta: BPS RI, 2025. [Daring]. Tersedia pada: <https://www.bps.go.id>
- [4] J. Han, M. Kamber, dan J. Pei, *Data Mining: Concepts and Techniques*, 4th Edition. Morgan Kaufmann, 2022.
- [5] D. N. Herlina dan K. Windhani, "The Influence of Social, Economic, and Ecological Aspects on Village Development Index in Indonesia," *J. Ekon. Pembang.*, 2025, [Daring]. Tersedia pada: <https://jurnal.feb.unila.ac.id/index.php/jep/article/view/3892>

- [6] M. Fatayat dan S. H. Supangkat, "Clustering Socio-Economic Risk Zones for School Dropout: A Data-Driven Approach for Education Intervention in Indonesia," IEEE, 2025. [Daring]. Tersedia pada: <https://ieeexplore.ieee.org/abstract/document/11389433/>
- [7] A. Yudono, A. Hidayat, dan W. P. Wijayanti, "Spatial Clustering of The Village Development Index Using K-Means and DBSCAN Approaches in Machine Learning: A Case Study in Malang Regency, Indonesia," in *IOP Conference Series: Earth and Environmental Science*, 2026. [Daring]. Tersedia pada: <https://iopscience.iop.org/article/10.1088/1755-1315/1595/1/012031/meta>
- [8] A. P. E. Susilowati, R. Rachmawati, dan R. Rijanta, "Smart Village Concept in Indonesia: ICT as Determining Factor," *Heliyon*, 2025, [Daring]. Tersedia pada: [https://www.cell.com/heliyon/fulltext/S2405-8440\(25\)00037-4](https://www.cell.com/heliyon/fulltext/S2405-8440(25)00037-4)
- [9] R. D. Saleh, E. Rustiadi, dan A. Fauzi, "Mapping Island Village Typology Based on Spatial and Socio-Economic Characteristics in Maluku Province, Indonesia," *Int. J. Sustain. Dev. Plan.*, 2025, [Daring]. Tersedia pada: https://iieta.org/sites/default/files/pdf/2025-07/ijsdp_20.06_09.pdf
- [10] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta, 2022.
- [11] P.-N. Tan, M. Steinbach, dan V. Kumar, *Introduction to Data Mining*. Pearson Education, 2019.
- [12] Kusriani dan E. T. Luthfi, *Algoritma Data Mining*. Yogyakarta: Andi Publisher, 2021.
- [13] B. Santosa, *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu, 2020.
- [14] T. M. Kodinariya dan P. R. Makwana, "Review on Determining Number of Cluster in K-Means Clustering," *Int. J. Adv. Res. Comput. Sci. Manag. Stud.*, vol. 1, no. 6, hal. 90-95, 2019.
- [15] I. Ghazali, *Aplikasi Analisis Multivariate dengan Program IBM SPSS 26*. Semarang: Badan Penerbit Universitas Diponegoro, 2021.