

PERBANDINGAN ALGORITMA K-MEANS DAN K-MEDOIDS DALAM PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN INDIKATOR KEADAAN SEKOLAH DASAR

I Putu Arya Vidyananta^{1*}, Kadek Teguh Dermawan²

^{1,2}Program Studi Sistem Informasi, Universitas Pendidikan Ganesha; Jl. Udayana No. 11 Singaraja-Bali 81116, Indonesia

Keywords:

Clustering;
K-Means;
K-Medoids;
Silhouette Score.

Correspondent Email:

arya.vidyananta@student.undiksha.ac.id

Abstrak. Penelitian ini dilatarbelakangi oleh ketidakmerataan kualitas pendidikan dasar di Indonesia yang terlihat dari perbedaan jumlah sekolah, guru, siswa, dan sarana prasarana antarprovinsi. Analisis berbasis data diperlukan untuk memetakan kondisi tersebut agar pemerintah dapat merumuskan kebijakan yang tepat sasaran. Penelitian ini menerapkan metode clustering dengan membandingkan algoritma K-Means dan K-Medoids menggunakan data sekolah dasar dari Portal Data Kementerian Pendidikan Dasar dan Menengah. Proses penelitian mengikuti kerangka CRISP-DM meliputi pemahaman masalah, persiapan data, pemodelan, dan evaluasi. Jumlah kluster optimal ditentukan melalui metode Elbow dan Silhouette Score. Hasil menunjukkan algoritma K-Means dengan dua kluster memiliki performa terbaik dengan nilai Silhouette Score 0,7069, lebih tinggi dibanding K-Medoids sebesar 0,6702. Kluster pertama mencakup sebagian besar provinsi dengan skala pendidikan kecil, sedangkan kluster kedua berisi provinsi besar dengan jumlah sekolah, siswa, dan guru jauh lebih tinggi. Temuan ini menunjukkan K-Means lebih sesuai digunakan dalam pemetaan kondisi pendidikan dasar di Indonesia, serta diharapkan dapat mendukung kebijakan pemerataan pendidikan berbasis bukti.



Copyright © [JITET](http://www.jitet.org) (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

Abstract. This study is motivated by the inequality of primary education quality in Indonesia, reflected in disparities in the number of schools, teachers, students, and facilities across provinces. Data-driven analysis is needed to map these conditions so the government can design more targeted policies. This research applies clustering by comparing K-Means and K-Medoids algorithms using primary school data from the Ministry of Primary and Secondary Education portal. The study follows the CRISP-DM framework, including problem understanding, data preparation, modeling, and evaluation. The optimal cluster number was determined using the Elbow method and Silhouette Score. Results show that K-Means with two clusters achieved the best performance with a Silhouette Score of 0.7069, higher than K-Medoids at 0.6702. The first cluster represents most provinces with smaller education scales, while the second cluster includes larger provinces with significantly more schools, students, and teachers. These findings suggest that K-Means is more suitable for mapping primary education conditions in Indonesia and may support evidence-based policies for educational equity.

1. PENDAHULUAN

Pendidikan merupakan kebutuhan dasar setiap individu dalam menjalani kehidupan, sekaligus menjadi fondasi utama dalam pembentukan dan peningkatan kualitas sumber

daya manusia suatu bangsa [1]. Jenjang Sekolah Dasar (SD) menjadi fondasi pertama yang menentukan perkembangan pengetahuan, keterampilan, dan karakter peserta didik di jenjang berikutnya. Keberhasilan

penyelenggaraan pendidikan dasar sangat dipengaruhi oleh kualitas dan ketersediaan sarana-prasarana sekolah, jumlah dan kompetensi tenaga pendidik, serta rasio guru terhadap siswa yang ideal.

Kualitas pendidikan di berbagai wilayah tidak dapat disamaratakan, sehingga diperlukan perhatian khusus terhadap daerah yang pendidikannya masih tergolong rendah. Jumlah penduduk yang besar saat ini menjadi tantangan bagi pemerintah dalam menentukan prioritas wilayah yang perlu difokuskan untuk pengembangan pendidikan Jenjang Sekolah Dasar (SD) [2]. Dengan mengelompokkan wilayah berdasarkan tingkat pendidikan, pemerintah dapat lebih mudah mengidentifikasi daerah yang memerlukan perhatian khusus dalam penanganan pendidikannya. Ketimpangan ini memerlukan analisis yang komprehensif agar dapat memberikan gambaran yang jelas mengenai pola distribusi kondisi pendidikan dasar di Indonesia.

Berdasarkan permasalahan tersebut, diperlukan pengolahan data untuk mengetahui Keadaan Sekolah Dasar (SD) di Indonesia berdasarkan kriteria tertentu guna memetakan perbedaan antar provinsi. Untuk mendapatkan hasil yang sesuai, penelitian ini menggunakan data keadaan sekolah dasar yang diperoleh dari Portal Data Kementerian Pendidikan Dasar dan Menengah. Data tersebut mencakup indikator jumlah guru, jumlah murid, jumlah sekolah, jumlah kelas, jumlah rombongan untuk setiap provinsi di Indonesia. Pengolahan data pada penelitian akan dilakukan menggunakan teknik *clustering*. *Clustering* adalah proses membagi sekumpulan objek data ke dalam beberapa kelompok yang disebut kluster. Objek-objek dalam satu kluster memiliki karakteristik yang mirip satu sama lain, namun berbeda dengan objek yang berada di kluster lain [3]. Penelitian ini secara khusus membandingkan dua algoritma *clustering* yang banyak digunakan, yaitu K-Means dan K-Medoid. K-Means dikenal efektif untuk data berukuran besar dengan prinsip centroid sebagai pusat kluster, sementara K-Medoid lebih robust terhadap keberadaan *outlier* karena menggunakan representasi objek aktual sebagai pusat kluster [4]. Dengan membandingkan kedua algoritma tersebut, penelitian ini bertujuan mengidentifikasi metode yang lebih tepat dalam

menghasilkan pengelompokan provinsi berdasarkan kondisi sekolah dasar di Indonesia.

Penelitian sebelumnya pernah dilakukan oleh [5] yang menerapkan algoritma K-Means untuk klusterisasi hasil Ujian Nasional SMA di Indonesia menggunakan data rata-rata nilai ujian per provinsi. Hasil evaluasi menunjukkan bahwa algoritma K-Means mampu menghasilkan partisi kluster yang cukup baik dengan nilai *Partition Coefficient Index (PCI)* sebesar 0,81. Selanjutnya, penelitian yang dilakukan oleh [6] menggunakan algoritma K-Medoids untuk mengelompokkan tingkat kemiskinan di 34 provinsi di Indonesia berdasarkan data BPS tahun 2015–2022. Hasil penelitian tersebut membentuk tiga kluster utama dengan nilai evaluasi *Silhouette Score* sebesar 0,4. Adapun penelitian lain oleh [7], membandingkan efektivitas algoritma K-Means dan K-Medoids dalam proses pengelompokan sekolah di Provinsi Riau berdasarkan indikator ketersediaan sarana dan prasarana pendidikan. Hasil pengujian dengan Davies-Bouldin Index (DBI) pada jumlah kluster $K=3$ menunjukkan bahwa algoritma K-Medoids menghasilkan kualitas kluster yang lebih optimal (DBI=0,61) dibandingkan K-Means (DBI=0,80).

Berbeda dengan penelitian-penelitian sebelumnya yang cenderung hanya menggunakan salah satu algoritma klusterisasi, penelitian ini secara khusus difokuskan pada perbandingan kinerja antara algoritma K-Means dan K-Medoids dalam menganalisis kondisi Sekolah Dasar di Indonesia. Melalui perbandingan kedua algoritma tersebut pada indikator jumlah kelas, jumlah murid, jumlah guru, jumlah sekolah, dan jumlah rombongan belajar per provinsi, penelitian ini bertujuan mengidentifikasi metode yang lebih sesuai dalam menghasilkan kluster provinsi dengan kualitas terbaik. Dengan demikian, hasil penelitian diharapkan mampu memberikan rekomendasi algoritma yang lebih tepat digunakan dalam pemetaan kondisi pendidikan dasar di Indonesia serta menjadi acuan dalam penentuan kebijakan pemerataan pendidikan.

2. TINJAUAN PUSTAKA

2.1 Data Mining

Data mining merupakan proses untuk mengekstraksi informasi yang bernilai dari kumpulan data berukuran besar dan kompleks dengan memanfaatkan berbagai teknik serta

algoritma analisis yang canggih. Terdapat berbagai metode yang dapat diterapkan dalam teknologi ini untuk menyelesaikan masalah, melakukan analisis, perencanaan, diagnosis, deteksi, integrasi, pencegahan, pembelajaran, serta mendorong inovasi. Dorongan utama dalam pengembangan data mining adalah kebutuhan akan metode yang lebih canggih untuk menganalisis, memahami, dan memvisualisasikan sejumlah besar data yang diperoleh dari berbagai aplikasi ilmiah maupun bisnis [8]. Tujuan utama data mining adalah mengidentifikasi pola, keterkaitan, serta informasi tersembunyi dalam data yang dapat memberikan wawasan berharga guna mendukung proses pengambilan keputusan [9]. Berbagai metode digunakan dalam data mining, diantaranya clustering, klasifikasi, *association rule mining*, dan deteksi anomali [10].

2.2 Clustering

Clustering Adalah salah satu metode analisis data dalam bidang *unsupervised learning* yang bertujuan mengelompokkan objek-objek berdasarkan tingkat kemiripannya, tanpa memerlukan label kelas yang telah ditentukan sebelumnya. Tujuan utamanya adalah menemukan struktur alami dalam data tanpa harus mengetahui label sebelumnya. Pendekatan ini banyak dimanfaatkan dalam sejumlah sektor seperti pemasaran, bioinformatika, pengenalan pola, analisis teks, serta pengambilan keputusan berbasis data. Pendekatan clustering dapat membantu dalam memahami perilaku konsumen, mengoptimalkan strategi bisnis, hingga meningkatkan efisiensi dalam pengelolaan sumber daya [11]. Clustering memiliki beragam metode yang dapat diterapkan untuk melakukan pengelompokan data, di antaranya yaitu *K-Means*, *K-Medoids*, *Self-Organizing Map (SOM)*, dan *Fuzzy C-Means* [12].

2.3 Algoritma K-Means

K-Means adalah salah satu algoritma *clustering* yang paling populer dan sederhana dalam data mining, yang bertujuan untuk membagi sekumpulan data ke dalam sejumlah *cluster* berdasarkan tingkat kemiripan antar data [13]. Proses ini dilakukan secara iteratif dengan memperbarui posisi *centroid* berdasarkan rata-rata data dalam setiap kluster hingga mencapai konvergensi, yaitu ketika pergerakan *centroid* sudah tidak signifikan. Meskipun efektif untuk dataset berukuran

besar, *K-Means* memiliki kelemahan seperti sensitivitas terhadap pemilihan *centroid* awal, asumsi bentuk kluster yang bulat, serta pengaruh besar dari keberadaan *outlier* [14].

2.4 Algoritma K-Medoids

K-Medoids adalah algoritma clustering yang mirip dengan *K-Means* tetapi pusat klusternya (*medoid*) harus merupakan salah satu objek data aktual, bukan rata-rata atau centroid yang bisa saja bukan data nyata. Keunggulan metode ini ialah daya tahan terhadap *outlier* dan data yang memiliki distribusi tidak merata, karena *medoid* sebagai representatif cluster merupakan anggota nyata dari kelompok. *K-Medoids clustering* berfungsi untuk meminimalkan jarak antara sekumpulan data dalam suatu kluster dengan titik data yang mewakili kluster tersebut. Titik data yang menjadi perwakilan atau representatif dari setiap kluster ini disebut sebagai *medoid* [14].

2.5 Silhouette Score

Silhouette Score atau *Silhouette Coefficient* merupakan metrik evaluasi internal pada algoritma klusterisasi yang digunakan untuk menilai sejauh mana suatu data berada pada kluster yang tepat, dengan mempertimbangkan tingkat kedekatan (*cohesion*) dan keterpisahan (*separation*) antar kluster. Perhitungan *Silhouette Score* dapat dilihat pada persamaan 1 berikut:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (1)$$

di mana $a(i)$ adalah rata-rata jarak objek ke anggota klusternya sendiri, dan $b(i)$ adalah rata-rata jarak objek ke objek dalam kluster lain terdekat. Nilai *Silhouette* berkisar antara -1 hingga $+1$: nilai mendekati $+1$ menyiratkan bahwa objek sangat cocok berada dalam klusternya, nilai mendekati 0 menandakan objek berada di batas antara dua kluster, dan nilai negatif menunjukkan bahwa objek mungkin mendapat kluster yang salah [15].

3. METODE PENELITIAN

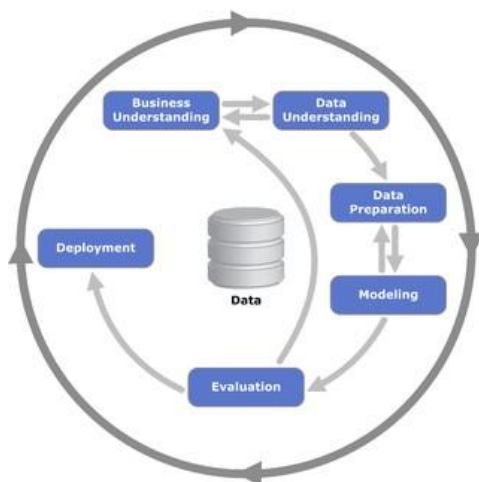
Penelitian ini menerapkan pendekatan analisis data untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan kondisi sekolah dasar. Proses penelitian dilakukan dengan mengikuti tahapan dalam kerangka kerja CRISP-DM (Cross-Industry Standard Process for Data Mining) yang terdiri atas enam tahap utama, di mana setiap tahap memiliki

peran penting dalam membangun model yang efektif untuk menjawab permasalahan penelitian [16]. Tahapan-tahapan dalam CRISP-DM meliputi *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment* [17]. Setiap tahap memiliki penjelasan dan fungsi tersendiri yang dijabarkan secara rinci berikut ini.

Gambar 1 Framework CRISP-DM

3.1 Business Understanding

Business Understanding adalah tahap untuk



memahami proses bisnis organisasi, termasuk tujuan, kondisi, permasalahan, dan kebutuhan yang ada [18]. Tahap awal penelitian ini berfokus pada pemahaman tujuan penelitian dan permasalahan yang ingin diselesaikan, yaitu mengelompokkan provinsi di Indonesia berdasarkan kondisi sekolah dasar. Pada tahap ini, peneliti merumuskan masalah secara jelas, yakni bagaimana mengidentifikasi kelompok provinsi yang memiliki karakteristik serupa dalam hal indikator pendidikan dasar, seperti rasio guru-murid, fasilitas sekolah, dan capaian pembelajaran. Tujuan penelitian ini adalah menyediakan informasi yang dapat membantu pembuat kebijakan dalam merancang strategi pendidikan yang lebih tepat sasaran.

3.2 Data Understanding

Tahap awal penelitian ini berfokus pada pemahaman tujuan penelitian dan permasalahan yang ingin diselesaikan, yaitu mengelompokkan provinsi di Indonesia berdasarkan kondisi sekolah dasar. Pada tahap

ini, peneliti merumuskan masalah secara jelas, yakni bagaimana mengidentifikasi kelompok provinsi yang memiliki karakteristik serupa dalam hal indikator pendidikan dasar, seperti rasio guru-murid, fasilitas sekolah, dan capaian pembelajaran. Tujuan penelitian ini adalah menyediakan informasi yang dapat membantu pembuat kebijakan dalam merancang strategi pendidikan yang lebih tepat sasaran. Rencana penelitian disusun untuk memastikan seluruh langkah analisis data mendukung pencapaian tujuan tersebut.

3.3 Data Preparation

Data mentah yang telah dikumpulkan kemudian dipersiapkan agar siap digunakan dalam proses pemodelan. Langkah-langkah yang dilakukan meliputi pembersihan data dari nilai hilang atau duplikat, normalisasi pada variabel numerik, serta pengkodean variabel kategorikal jika diperlukan. Selain itu, variabel yang tidak relevan dihapus, sementara fitur-fitur penting seperti rasio guru-murid, jumlah fasilitas sekolah, dan capaian pembelajaran dipilih untuk membentuk dataset akhir yang optimal. Tahap ini sangat penting karena kualitas hasil pengelompokan provinsi di Indonesia berdasarkan indikator keadaan sekolah dasar sangat bergantung pada kebersihan, keseragaman, dan kelengkapan data yang digunakan dalam penerapan algoritma K-Means dan K-Medoids.

3.4 Modeling

Pada tahap *modeling*, algoritma K-Means dan K-Medoids diterapkan pada dataset yang telah dipersiapkan untuk mengelompokkan provinsi di Indonesia berdasarkan kesamaan karakteristik pada indikator keadaan sekolah dasar. Kedua algoritma ini dipilih untuk dibandingkan karena memiliki pendekatan yang berbeda dalam menentukan pusat *cluster*—K-Means menggunakan nilai rata-rata (*centroid*), sedangkan K-Medoids menggunakan data aktual (*medoid*) sebagai representatif kluster. Proses pemodelan dilakukan dengan menentukan jumlah kluster optimal menggunakan metode *Elbow* dan *Silhouette Score* untuk memperoleh hasil pengelompokan yang paling representatif terhadap kondisi nyata di lapangan.

3.5 Evaluation

Tahap evaluasi dilakukan untuk menilai kualitas hasil pengelompokan yang dihasilkan oleh kedua algoritma. Evaluasi mencakup

analisis terhadap jumlah dan distribusi *cluster*, jarak antar-klaster, serta keseragaman data dalam setiap kelompok. Nilai *Silhouette Score* digunakan sebagai metrik utama untuk mengukur seberapa baik setiap provinsi ditempatkan dalam klaster yang sesuai. Penilaian ini bertujuan memastikan bahwa hasil pengelompokan tidak hanya akurat secara matematis, tetapi juga relevan secara substantif terhadap kondisi pendidikan dasar di Indonesia.

3.6 Deployment

Tahap terakhir penelitian adalah penerapan hasil pengelompokan untuk memberikan manfaat praktis bagi pengambil kebijakan di sektor pendidikan. Hasil dari analisis K-Means dan K-Medoids digunakan sebagai dasar dalam menyusun strategi dan kebijakan yang lebih tepat sasaran, seperti pemerataan fasilitas sekolah, peningkatan kualitas tenaga pendidik, dan intervensi pada provinsi dengan indikator pendidikan dasar yang rendah.

4. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil analisis data yang telah dilakukan serta pembahasan terkait pengelompokan provinsi di Indonesia berdasarkan indikator keadaan sekolah dasar. Hasil penelitian dijelaskan secara rinci, mencakup pengelompokan provinsi menggunakan algoritma K-Means dan K-Medoid, karakteristik masing-masing cluster, serta interpretasi temuan yang relevan. Pembahasan difokuskan pada perbandingan performa kedua algoritma, kualitas cluster yang dihasilkan, serta implikasinya terhadap pemahaman kondisi sekolah dasar di setiap provinsi. Dengan demikian, hasil clustering tidak hanya bersifat deskriptif, tetapi juga memberikan insight yang dapat mendukung pengambilan keputusan dan perencanaan kebijakan di bidang pendidikan.

4.1 Business Understanding

Penelitian ini dilatarbelakangi oleh ketidakmerataan kondisi pendidikan dasar di Indonesia, yang tercermin dari perbedaan fasilitas sekolah, kompetensi guru, dan capaian pembelajaran antarprovinsi. Tujuan penelitian adalah melakukan pengelompokan 34 provinsi berdasarkan kesamaan karakteristik sekolah dasar menggunakan metode K-Means dan K-Medoid, sehingga setiap cluster merepresentasikan provinsi dengan kondisi pendidikan yang serupa. Hasil pengelompokan diharapkan dapat menjadi acuan bagi

Kementerian Pendidikan dan pemerintah daerah dalam menyusun kebijakan, merancang program peningkatan kualitas pendidikan, serta menyalurkan bantuan secara tepat sasaran. Penelitian ini menekankan pada pembangunan model clustering yang mampu menghasilkan kelompok provinsi yang jelas, interpretable, dan dapat digunakan sebagai dasar rekomendasi kebijakan pendidikan yang berbasis bukti (*evidence-based*).

4.2 Data Understanding

Tahap *Data Understanding* merupakan proses pengenalan terhadap data yang digunakan untuk memperoleh pemahaman awal mengenai karakteristik dan kualitas data. Data yang digunakan dalam penelitian ini bersumber dari Portal Data Kementerian Pendidikan Dasar dan Menengah. Dataset terdiri dari 38 observasi seluruh provinsi di Indonesia dan 9 variabel indikator.

Tabel 1 Variabel Dataset

Variabel	Deskripsi	Skala
Sekolah	Jumlah sekolah dasar di provinsi	Numerik
Siswa	Jumlah siswa sekolah dasar di provinsi	Numerik
Mengulang	Jumlah siswa yang mengulang kelas	Numerik
Putus Sekolah	Jumlah siswa yang putus sekolah	Numerik
Kepala Sekolah dan Guru	Jumlah guru dan kepala sekolah	Numerik
Tenaga Kependidikan	Jumlah tenaga non-guru	Numerik
Rombel	Jumlah Rombongan belajar	Numerik
Ruang Kelas	Jumlah ruang kelas yang tersedia	Numerik

4.3 Data Preparation

Pada tahap persiapan data, dilakukan serangkaian langkah untuk memastikan kualitas dan kelayakan data sebelum dianalisis lebih lanjut. Langkah awal yang dilakukan dalam persiapan data adalah pembersihan data

(Cleaning Data) yang meliputi penanganan *missing value* dan penghapusan duplikasi. Langkah ini penting agar dataset yang digunakan akurat dan tidak memengaruhi hasil clustering. Gambar 2 dan Gambar 3 menyajikan ilustrasi proses penanganan *missing values* dan penghapusan duplikasi.

```
print("Cek Missing Value:\n", df.isnull().sum())
```

Gambar 2 Cek Missing Value

```
print("Cek Dupikat:\n", df.duplicated().sum())

Cek Dupikat:
0
```

Gambar 3 Proses Cek Duplikasi

Tahap persiapan data selanjutnya yaitu normalisasi/standarisasi data untruk mencegah atribut dengan skala besar mendominasi algoritma clustering, sehingga setiap indicator Pendidikan memiliki kontribusi yang seimbang dalam pembentukan cluster. Pada penelitian ini menggunakan Z-score normalization untuk standarisasi. Gambar 4 merupakan ilustrasi dari proses standarisasi.

```
# Standarisasi data
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
```

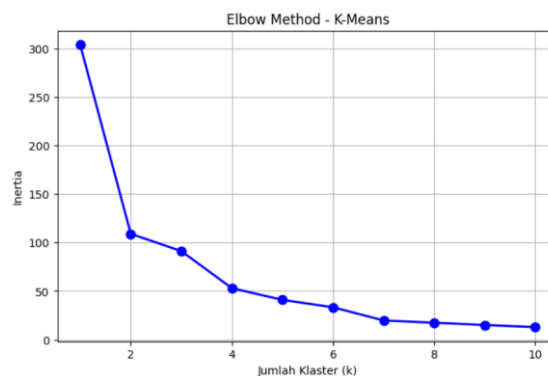
Gambar 4 Proses Standarisasi Data

4.4 Modeling

Pada tahap pemodelan, dilakukan penerapan dua algoritma klasterisasi, yaitu K-Means dan K-Medoid, untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan indikator keadaan sekolah dasar.

1. K-Means

Proses pada K-Means dimulai dengan menentukan jumlah cluster optimal menggunakan metode Elbow, yang memberikan gambaran titik tekuk (*elbow point*) pada grafik *within-cluster sum of squares (WCSS)*. Gambar 5 menunjukan hasil Elbow Point dari K-Means.



Gambar 5 Elbow Point K-Means

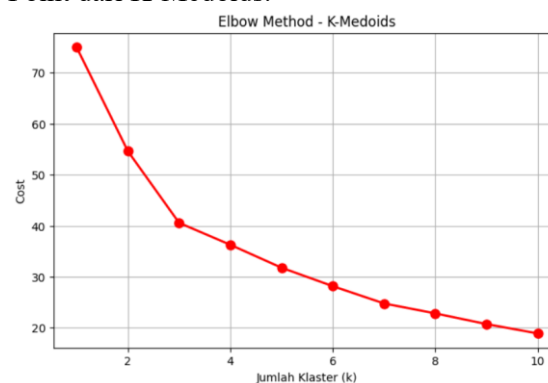
Berdasarkan hasil tersebut, dilakukan percobaan clustering dengan variasi jumlah cluster sebanyak 2, 3, dan 4. Tabel 2 menunjukkan distribusi provinsi pada masing-masing percobaan klaster.

Tabel 2 Distrubusi data di setiap cluster

Ekperimen	Klaster			
	1	2	3	4
K= 2	34	4	-	-
K=3	34	3	1	-
K=4	10	3	1	24

2. K-Medoids

Proses di K-Medoids dimulai dengan menetapkan jumlah cluster yang paling tepat melalui pendekatan metode Elbow, yang menunjukkan gambaran titik tekuk (*elbow point*) dalam grafik *within-cluster sum of squares (WCSS)*. Gambar 6 merupakan hasil Elbow Point dari K-Medoids.



Gambar 6 Elbow Point K-Medoids

Berdasarkan hasil tersebut, dilakukan percobaan clustering dengan variasi jumlah cluster sebanyak 2, 3, dan 4. Tabel 3 menunjukkan distribusi provinsi pada masing-masing percobaan.

Tabel 3 Distribusi data setiap cluster

Eksperimen	Klaster			
	1	2	3	4
K=2	34	4	-	-
K=3	3	12	23	-
K=4	3	14	16	5

4.5 Evaluation

Pada tahap evaluasi, penelitian ini menekankan pada proses penilaian kualitas hasil pengelompokan yang diperoleh dari algoritma K-Means dan K-Medoid. Evaluasi dilakukan dengan menggunakan metrik evaluasi internal, yaitu *Silhouette Coefficient* yang secara umum digunakan untuk menilai kualitas clustering.

1. *Silhouette Score*

Indeks ini mengukur seberapa baik objek berada dalam cluster yang tepat dibandingkan dengan cluster terdekat lainnya. Nilai mendekati +1 menunjukkan bahwa data berada dalam cluster yang tepat, nilai mendekati 0 menunjukkan data berada di batas antar cluster. Tabel 4 menunjukkan hasil *Silhouette Score* K-Means dan K-Medoids.

Tabel 4 Perbandingan *Silhouette Score* K-Means dan K-Medoids

Klaster	<i>Silhouette Score</i> K-Means	<i>Silhouette Score</i> K-Medoids
2	0.7069	0.6702
3	0.6526	0.4782
4	0.4535	0.3370

Berdasarkan hasil perhitungan *Silhouette Score* pada tabel di atas, dapat dilihat bahwa nilai tertinggi diperoleh pada jumlah klaster $k = 2$, baik untuk algoritma K-Means maupun K-Medoids, dengan nilai masing-masing sebesar 0.7069 dan 0.6702. Nilai ini menunjukkan bahwa pengelompokan dengan dua klaster memberikan hasil yang paling baik dan representatif, karena jarak antar data dalam satu klaster relatif kecil dan jarak antar klaster cukup besar.

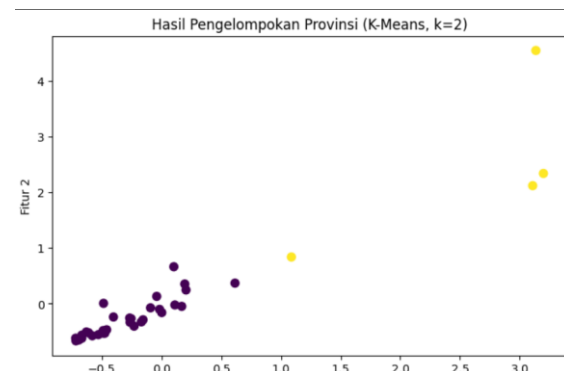
Sementara itu, ketika jumlah klaster meningkat menjadi $k = 3$ dan $k = 4$, nilai *Silhouette Score* cenderung menurun pada kedua algoritma. Hal ini menandakan

bahwa pembagian data menjadi lebih banyak klaster justru mengurangi kualitas pengelompokan, karena beberapa data berada di batas antar klaster atau kurang sesuai dengan klaster yang seharusnya.

Dari hal tersebut menunjukan K-Means memiliki performa yang lebih baik dibandingkan K-Medoids dalam konteks penelitian ini, karena menghasilkan nilai *Silhouette Score* yang lebih tinggi pada setiap jumlah klaster.

4.6 Deployment

Berdasarkan hasil evaluasi tersebut, tahap Deployment dilakukan dengan menerapkan hasil pengelompokan terbaik, yaitu menggunakan algoritma K-Means dengan jumlah klaster sebanyak dua. Untuk memvisualisasikan hasil pengelompokan, dapat dilihat pada gambar 7 yang menampilkan distribusi provinsi ke dalam dua klaster.



Gambar 7 Hasil Clustering Provinsi dengan K-Means cluster 2

Hasil pengelompokan ditampilkan pada Tabel daftar klaster, yang memuat pembagian provinsi ke dalam klaster 1 dan klaster 2. Tabel ini berfungsi untuk menunjukkan perbedaan wilayah secara lebih eksplisit, sekaligus menjadi dasar dalam memahami karakteristik umum yang dimiliki oleh setiap klaster.

Tabel 5 Distribusi Provinsi di Setiap Klaster

Klaster 1	Klaster 2
DKI Jakarta, D.I. Yogyakarta, Aceh, Sumatra Barat, Riau, Jambi, Sumatera Selatan, Lampung, Kalimantan Barat,	Jawa Barat, Jawa Tengah, Jawa Timur, dan Sumatera Utara

Kalimantan Tengah,
 Kalimantan Selatan,
 Kalimantan Timur,
 Sulawesi Utara,
 Sulawesi Tengah,
 Sulawesi Selatan,
 Sulawesi Tenggara,
 Maluku, Bali, NTB,
 NTT, Banten,
 Kepulauan Bangka
 Belitung, Gorontalo,
 Kepulauan Riau,
 Papua Barat,
 Sulawesi Barat,
 Kalimantan Utara,
 Papua Tengah,
 Papua Selatan, Papua
 Pegunungan, dan
 Papua Barat Daya.

Kemudian, untuk memperdalam pemahaman mengenai perbedaan antar klaster, Dihitung juga rata-rata variabel tiap klaster. Tabel 6 dan 7 menunjukkan nilai rata-rata dari indikator yang digunakan dalam penelitian, sehingga dapat menjadi dasar untuk menginterpretasikan kondisi khas yang membedakan antara klaster 1 dan klaster 2.

Tabel 6 Rata -Rata Nilai Variabel Cluster 1

Rata – Rata Variabel Klaster 1	
Sekolah	2.060
Siswa	315.720
Mengulang	1.583
Putus Sekolah	618
Kepala Sekolah/	22.502
Tenaga Kependidikan	2.455
Rombel	15.882
Ruang Kelas	16.130

Tabel 7 Rata-Rata Nilai Variabel Klaster 2

Rata – Rata Variabel Klaster 2	
Sekolah	14.803
Siswa	2.349.661
Mengulang	3.357
Putus Sekolah	2.628
Kepala Sekolah/	13.1384
Tenaga Kependidikan	9.691
Rombel	108.775
Ruang Kelas	108.590

Berdasarkan tabel rata-rata, terlihat bahwa Klaster 1 umumnya terdiri dari provinsi dengan jumlah sekolah, siswa, kepala sekolah,

rombongan belajar, dan ruang kelas yang relatif lebih sedikit. Hal ini menunjukkan bahwa provinsi-provinsi dalam klaster ini memiliki skala penyelenggaraan pendidikan yang lebih kecil, meskipun tersebar di banyak wilayah. Sementara itu, Klaster 2 mencakup provinsi-provinsi besar seperti Jawa Barat, Jawa Tengah, Jawa Timur, dan Sumatera Utara yang memiliki jumlah sekolah, siswa, dan tenaga kependidikan jauh lebih tinggi dibandingkan klaster pertama

5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, 38 provinsi di Indonesia berdasarkan indikator keadaan sekolah dasar dengan menggunakan algoritma K-Means dan K-Medoid dapat disimpulkan.

1. Hasil evaluasi menunjukkan bahwa metode K-Means dengan jumlah klaster dua menghasilkan performa terbaik dengan nilai Silhouette Score sebesar 0.7069, lebih tinggi dibandingkan metode K-Medoids sebesar 0.6702.
2. Klaster pertama berisi sebagian besar provinsi dengan jumlah sekolah, siswa, guru, rombongan belajar, dan ruang kelas yang relatif lebih sedikit, sedangkan klaster kedua mencakup provinsi dengan skala pendidikan yang lebih besar, seperti Jawa Barat, Jawa Tengah, Jawa Timur, dan Sumatera Utara.
3. Kelebihan penelitian ini adalah penggunaan perbandingan dua algoritma sekaligus sehingga memungkinkan perbandingan performa, serta pemanfaatan data resmi nasional yang relevan untuk mendukung perencanaan kebijakan berbasis bukti.
4. Penelitian ini masih memiliki beberapa keterbatasan dimana jumlah variabel yang digunakan hanya sembilan indikator sehingga belum mampu menggambarkan kondisi pendidikan dasar secara menyeluruh, selanjutnya kualitas data yang tersedia belum sepenuhnya merata, sehingga distribusi hasil clustering cenderung timpang yang menunjukkan adanya ketidakseimbangan jumlah anggota antar klaster.
5. Penelitian selanjutnya disarankan untuk menambah variabel yang lebih beragam, meningkatkan kualitas dan kelengkapan data, serta mencoba algoritma clustering lain

agar hasil yang diperoleh semakin mendalam

UCAPAN TERIMA KASIH

Penulis menyampaikan apresiasi dan rasa terima kasih yang sebesar-besarnya kepada seluruh pihak yang telah memberikan dukungan, bimbingan, serta kontribusi dalam pelaksanaan dan penyelesaian penelitian ini

DAFTAR PUSTAKA

- [1] R. Kurniawan, M. M. M. Mukarrobin, and M. Mahradianur, "Klasterisasi Tingkat Pendidikan Di Dki Jakarta Pada Tingkat Kecamatan Menggunakan Algoritma K-Means," *Technologia: Jurnal Ilmiah*, vol. 12, no. 4, pp. 234–239, 2021.
- [2] N. Nurahman and D. D. Aulia, "klasterisasi pendidikan masyarakat untuk mengetahui daerah dengan pendidikan terendah menggunakan algoritma K-Means," *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 5, no. 1, pp. 38–44, 2023.
- [3] N. K. Zuhail, "Study Comparison K-Means Clustering Dengan Algoritma Hierarchical Clustering," in *Seminar Nasional Teknologi & Sains*, 2022, pp. 200–205.
- [4] J. Heidari, N. Daneshpour, and A. Zangeneh, "A novel K-means and K-medoids algorithms for clustering non-spherical-shape clusters non-sensitive to outliers," *Pattern Recognit*, vol. 155, p. 110639, 2024, doi: <https://doi.org/10.1016/j.patcog.2024.110639>.
- [5] W. A. Suputra, I. Candiasa, and I. Suryawan, "Klasterisasi hasil ujian nasional SMA/MA dengan algoritma k-means. Wahana Matematika dan Sains: Jurnal Matematika, Sains, dan Pembelajarannya, 15 (1), 22-30," 2021.
- [6] F. Zahra, A. Khalif, and B. N. Sari, "Pengelompokan Tingkat Kemiskinan Di Setiap Provinsi Di Indonesia Menggunakan Algoritma K-Medoids," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, Apr. 2024, [Online]. Available: <https://journal.eng.unila.ac.id/index.php/jitet/article/view/4199>
- [7] M. D. Salman *et al.*, "Comparison of K-Means and K-Medoids Clustering Algorithm Performance in Grouping Schools in Riau Province Based on Availability of Facilities and Infrastructure," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 3, pp. 797–806, Jun. 2025, doi: [10.57152/malcom.v5i3.1950](https://doi.org/10.57152/malcom.v5i3.1950).
- [8] I. G. S. D. Putra and I. N. T. A. Putra, "Implementasi Metode Naïve Bayes Pada Analisis Sentimen Pengguna Aplikasi Mobile Kita Bisa," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 2, 2025.
- [9] Z. Setiawan *et al.*, *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia, 2023.
- [10] N. W. Wardani, P. G. S. C. Nugraha, and G. S. Mahendra, "Implementasi Naïve Bayes Pada Data Mining Untuk Mengklasifikasikan Penjualan Barang Terlaris Pada Perusahaan Ritel," *JST (Jurnal Sains dan Teknologi)*, vol. 12, no. 3, pp. 656–668, 2023.
- [11] C. Ergenç and R. Aktaş, "Clustering S&P 500 companies by machine learning for sustainable decision-making," *Economics and Business Review*, vol. 11, no. 3, pp. 91–117, 2025.
- [12] D. D. Hariyanti, G. A. Pradnyana, and I. G. M. Darmawiguna, "Kombinasi metode Naive Bayes dan K-Medoid dalam memprediksi penjurusan siswa di sekolah menengah atas," *J. Ilmu Komput*, vol. 14, no. 2, p. 88, 2021.
- [13] N. Wijaya, "Performa Algoritma K-Means Dan Fuzzy C-Means Dalam Analisis Klaster Pendidikan Di Tingkat Kecamatan Jakarta," *Jurnal Komputer dan Informatika*, vol. 18, no. 2, pp. 80–86, 2023.
- [14] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf Sci (N Y)*, vol. 622, pp. 178–210, 2023.
- [15] Y. Hasan, "Pengukuran Silhouette Score dan Davies-Bouldin Index pada Hasil Cluster K-Means dan DBSCAN," *KAKIFIKOM (Kumpulan Artikel Karya Ilmiah Fakultas Ilmu Komputer)*, vol. 6, no. 1, pp. 60–74, Apr. 2024, [Online]. Available: <https://ejournal.ust.ac.id/index.php/KAKIFIKOM/article/view/3938>
- [16] G. S. Mahendra, T. Santhi, K. D. A. Sutrisna, P. P. Cahayani, I. G. Hendrayana, and P. G. S. C. Nugraha, "Sistem Pendukung Keputusan untuk Merekomendasikan Wisata di Kabupaten Klungkung Menggunakan Metode MOORA," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 1, pp. 567–575, 2025.
- [17] G. S. Mahendra and I. N. I. Wiradika, "Sistem Pendukung Keputusan Pemilihan Daya Tarik Wisata Favorit Menggunakan PIPRECIA-CoCoSo dengan Implementasi

- Python,” *Teknomatika*, vol. 14, no. 01, pp. 1–12, 2024.
- [18] K. A. Pratama, G. A. Pradnyana, and I. K. R. Arthana, “Pengembangan Sistem Cerdas Untuk Prediksi Daftar Kembali Mahasiswa Baru Dengan Metode Naive Bayes (Studi Kasus: Universitas Pendidikan Ganesha),” *SINTECH (Science and Information Technology) Journal*, vol. 3, no. 1, pp. 22–34, 2020.