

ANALISIS K-MEANS DENGAN RAPIDMINER UNTUK KLASIFIKASI KUALITAS PENDIDIKAN SEKOLAH DASAR DI INDONESIA

Muhammad Aziiz Irwansyah¹, Yelli Nur Alinda², Risma Nur'Aini³, Intan Aidita Alfitrah⁴,
Annisa Khairani⁵, Ken Ditha Tania⁶, Allsela Meiriza⁷, Ahmad Rifai⁸

^{1,2,3,4,5,6,7}Program Studi Sistem Informasi, Universitas Sriwijaya, Palembang

E-mail: ¹09031282227095@student.unsri.ac.id, ²09031182227119@student.unsri.ac.id,

³09031282227127@student.unsri.ac.id, ⁴09031282227101@student.unsri.ac.id,

⁵09031282227133@student.unsri.ac.id, ⁶kenya.tania@gmail.com, ⁷allsela_meiriza@gmail.com,

⁸rifai.bae@gmail.com

Abstrak

Penelitian ini menerapkan metode K-Means Clustering untuk mengelompokkan 38 provinsi di Indonesia berdasarkan kualitas pendidikan dasar. Tujuan penelitian ini adalah mengidentifikasi pola distribusi pendidikan dengan mempertimbangkan faktor tenaga pendidik, angka putus sekolah, kondisi infrastruktur sekolah, serta tingkat kesejahteraan guru. Dataset yang digunakan berasal dari Kaggle SD tahun 2023-2024 dan data Upah Minimum Provinsi (UMP) tahun 2024, kemudian dianalisis melalui tahapan Knowledge Discovery in Database (KDD) menggunakan RapidMiner. Hasil klasterisasi menghasilkan tiga kelompok provinsi dengan karakteristik berbeda: Klaster 0 dengan jumlah sekolah dan siswa tinggi serta angka putus sekolah sedang; Klaster 1 dengan tenaga pendidik dan ruang kelas terbanyak serta angka putus sekolah terendah; dan Klaster 2 dengan angka putus sekolah tertinggi meskipun UMP tertinggi. Evaluasi kualitas klasterisasi menggunakan Davies-Bouldin Index (DBI = 0,162) menunjukkan hasil yang baik. Berdasarkan analisis magnitudo vektor Euclidean, faktor dominan dalam pembentukan klaster adalah Kepala Sekolah dan Guru (1,376), Putus Sekolah (1,368), Ruang Kelas (baik) (1,324), Sekolah (1,312), Siswa (1,286), dan UMP (1,214). Penelitian ini menyimpulkan bahwa faktor tenaga pendidik dan kondisi infrastruktur memiliki dampak lebih besar terhadap kualitas pendidikan dasar dibandingkan faktor ekonomi.

Kata kunci: Pendidikan Dasar; Knowledge Discovery in Database; K-Means Clustering; Kualitas Pendidikan; Davies-Bouldin Index

Abstract

This study applied the K-Means Clustering method to group 38 provinces in Indonesia based on the quality of basic education. The aim of the study was to identify the distribution patterns of education by considering factors such as teaching staff, dropout rates, school infrastructure conditions, and teacher welfare levels. The dataset was sourced from the 2023–2024 Indonesian Elementary School data on Kaggle and the 2024 Indonesian Provincial Minimum Wage (UMP) data, then analyzed through the stages of Knowledge Discovery in Database (KDD) using RapidMiner. The clustering results produced three groups of provinces with distinct characteristics: Cluster 0 with a high number of schools and students and a moderate dropout rate; Cluster 1 with the highest number of teaching staff and classrooms as well as the lowest dropout rate; and Cluster 2 with the highest dropout rate despite having the highest UMP. The clustering quality evaluation using the Davies-Bouldin Index (DBI = 0.162) indicated a good clustering performance. Based on the Euclidean vector magnitude analysis, the dominant factors in cluster formation were Principals and Teachers (1.376), Dropout Rate (1.368), Good Classroom Conditions (1.324), Schools (1.312), Students (1.286), and UMP (1.214). The study concludes that teaching staff and school infrastructure have a greater impact on the quality of basic education than economic factors.

Keywords: Primary Education; Knowledge Discovery in Database; K-Means Clustering;

1. PENDAHULUAN

Pendidikan dasar memiliki peran yang sangat penting dalam membentuk sumber daya manusia yang berkualitas dan menjadi pondasi utama bagi kemajuan suatu negara [1]. Di Indonesia sendiri, pendidikan jenjang sekolah dasar di setiap provinsi menunjukkan variasi yang signifikan dalam hal jumlah sekolah, jumlah siswa, tingkat putus sekolah, kualitas guru, serta kondisi infrastruktur pendidikan [2]. Perbedaan ini dipengaruhi oleh berbagai faktor, termasuk kebijakan pemerintah daerah, alokasi anggaran [3], serta kondisi sosial dan ekonomi masyarakat setempat [4]. Oleh karena itu, diperlukan analisis berbasis data untuk memahami pola distribusi kualitas pendidikan di seluruh provinsi guna merumuskan kebijakan yang lebih tepat sasaran.

Clustering adalah metode agregasi data yang digunakan untuk menemukan pola dalam data pendidikan [5]. Dalam penelitian ini, metode K-Means Clustering diterapkan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan indikator pendidikan jenjang sekolah dasar, seperti jumlah siswa, tingkat putus sekolah, jumlah tenaga pendidik berdasarkan tingkat pendidikan, serta kondisi ruang kelas. Dengan teknik ini, provinsi dengan karakteristik serupa dapat dikelompokkan sehingga memberikan wawasan yang lebih mendalam bagi pengambil kebijakan dalam merancang strategi pemerataan dan peningkatan kualitas pendidikan.

Beberapa penelitian sebelumnya, seperti yang dikemukakan oleh [5], [6], [7], dan [8], telah menerapkan metode clustering dalam bidang pendidikan, tetapi dengan fokus, atribut, serta jenjang yang berbeda. Studi ini mencoba untuk memperluas cakupan dengan memasukkan lebih banyak variabel yang mencerminkan kondisi pendidikan dasar secara holistik. Dengan menerapkan metode *K-Means Clustering* pada data pendidikan dari seluruh provinsi di Indonesia, penelitian ini bertujuan untuk: Clustering pada data pendidikan dari seluruh provinsi di Indonesia, penelitian ini bertujuan untuk:

1. Mengidentifikasi kelompok provinsi berdasarkan karakteristik pendidikan dasar.
2. Menganalisis faktor-faktor utama yang memengaruhi perbedaan kualitas pendidikan antar kelompok provinsi, termasuk faktor sosial-ekonomi seperti Upah Minimum Provinsi (UMP).
3. Memberikan rekomendasi berbasis data untuk peningkatan kualitas pendidikan dasar yang lebih merata di Indonesia.

Hasil penelitian ini diharapkan dapat membantu pemerintah dalam merancang kebijakan pendidikan yang lebih tepat sasaran dengan memahami pola distribusi kualitas pendidikan dasar di setiap provinsi. Analisis ini juga dapat menjadi dasar bagi penelitian lanjutan untuk mengeksplorasi faktor-faktor yang memengaruhi kualitas pendidikan dan strategi peningkatannya.

2. METODE PENELITIAN

Penelitian ini menggunakan kerangka metodologi dengan tujuh tahapan utama untuk mengklasifikasikan kualitas pendidikan Sekolah Dasar (SD) per provinsi di Indonesia.



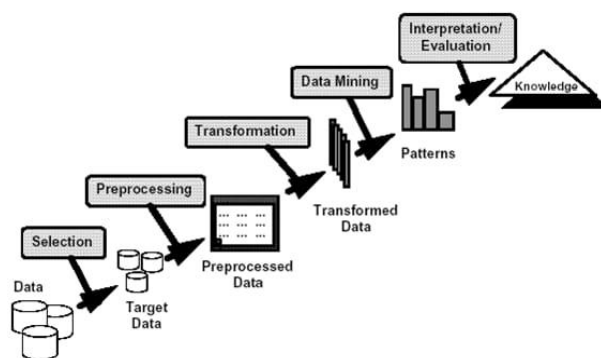
Gambar 1. Kerangka Metodologis

Deskripsi :

1. Identifikasi Masalah: Menentukan masalah utama untuk memastikan penelitian memiliki fokus yang jelas dan signifikan.
2. Menentukan Tujuan: Menetapkan tujuan penelitian yang terperinci dan terukur sebagai acuan dalam mencapai hasil yang diharapkan.
3. Studi Literatur: Mengumpulkan referensi dari buku, jurnal ilmiah, dan penelitian terdahulu untuk memperkuat landasan teori.
4. Pengumpulan Data: Mengambil data dari riset public di Kaggle dan data dari Kementerian Ketenagakerjaan untuk memastikan data yang valid dan signifikan.
5. Penerapan KDD: Menggunakan metodologi *Knowledge Discovery in Database* (KDD) melalui tahapan seleksi data, pra proses, transformasi, data mining dan evaluasi interpretasi untuk menemukan informasi baru.
6. Analisis Hasil: Membandingkan hasil pengolahan data dengan penelitian terdahulu untuk mendapatkan kesimpulan yang akurat.
7. Kesimpulan dan Saran: Menyusun laporan akhir, memberikan saran pengembangan, dan menawarkan rekomendasi praktis berdasarkan temuan penelitian.

2.1 Metodologi Knowledge Discovery in Database (KDD)

Tujuan penelitian ini adalah untuk mendapatkan pengetahuan dalam pengambilan keputusan terkait klasifikasi kualitas pendidikan sekolah dasar (SD) di Indonesia berdasarkan provinsi. *Knowledge Discovery in Database* (KDD) adalah proses pengumpulan dan pemanfaatan data historis untuk mengidentifikasi pola dan mengekstraksi pengetahuan yang membantu pengambilan keputusan [14]. Dalam penelitian ini, tahapan KDD berikut digunakan:



Gambar 2. Tahapan dalam Knowledge Discovery in Database (KDD)

1. *Selection* (Seleksi Data)

Seleksi data dilakukan untuk memastikan hanya data yang relevan dan sesuai kebutuhan penelitian yang dipilih. Proses ini melibatkan pemilihan data pendidikan SD per provinsi di Indonesia yang mencakup variabel-variabel penting untuk proses clustering menggunakan algoritma *K-Means*.

2. *Preprocessing* (Praproses Data)

Tahap *preprocessing* bertujuan untuk membersihkan dan mempersiapkan data sebelum diolah lebih lanjut. Tahap *Preprocessing/Data Cleaning* adalah proses yang meliputi penghapusan variabel yang tidak digunakan dan memperbaiki kesalahan data seperti kesalahan cetak [15]. Dalam penelitian ini, dilakukan proses data reduction untuk menghilangkan data yang tidak konsisten, memperbaiki kesalahan data, serta memastikan data dalam kondisi siap untuk proses transformasi dan data mining.

3. *Transformations* (Transformasi Data)

Transformasi data dilakukan untuk mengubah data mentah menjadi format yang sesuai dengan algoritma *clustering K-Means*, yang membutuhkan data dalam bentuk numerik atau integer. Oleh karena itu, tahap ini melibatkan proses normalisasi data untuk membuat proses clustering lebih mudah.

4. *Data Mining* (Penambangan Data)

Pada tahap ini, algoritma *K-Means* digunakan untuk melakukan *clustering* terhadap data pendidikan. Proses data mining ini bertujuan untuk menemukan pola atau hubungan antar data sehingga dapat mengklasifikasikan kualitas pendidikan di setiap provinsi berdasarkan kriteria tertentu. Pemilihan algoritma *K-Means* didasarkan pada kemampuannya dalam mengelompokkan data dengan efisien sesuai dengan karakteristiknya.

5. *Interpretation/Evaluation* (Interpretasi dan Evaluasi)

Tahap ini merupakan proses untuk menginterpretasikan hasil clustering dari algoritma *K-Means*. Evaluasi dilakukan dengan menganalisis pola-pola yang muncul dan menilai performa algoritma dalam mengelompokkan atribut yang diberikan. Hasil evaluasi kemudian disajikan dalam bentuk visualisasi data dan laporan yang mudah dipahami oleh pembaca.

2.2 Clustering K-Means

Clustering *K-Means* adalah pendekatan pembelajaran tanpa pengawasan yang digunakan untuk mengelompokkan data menjadi beberapa klaster berdasarkan kemiripan atribut [16]. Metode *K-Means* digunakan dalam penelitian ini untuk mengklasifikasikan kualitas pendidikan sekolah dasar (SD) di setiap provinsi Indonesia.

Metode *K-Means* bekerja dengan meminimalkan jarak antara data dengan centroid cluster. Proses *K-Means* melibatkan perhitungan ulang centroid hingga cluster stabil atau mencapai iterasi maksimum. Berikut ini adalah tahapan penerapan algoritma *K-Means* di RapidMiner:

1. Persiapan Data
 - a Impor dataset.
 - b Lakukan pembersihan data (*Remove Duplicates, Handle Missing Values*).
2. Preprocessing Data
 - a Gunakan operator *Normalize* untuk memastikan data dalam skala yang sama.
 - b Gunakan *Select Attributes* untuk memilih fitur-fitur relevan, seperti jumlah siswa, tingkat putus sekolah, dan tingkat pendidikan guru.
3. Mengatur *K-Means Clustering*
 - a Tambahkan operator K-Means dari *Modeling > Clustering*.
 - b Tentukan jumlah kluster (K) sesuai kebutuhan.
 - c Pilih *Distance Measure* sebagai *Cosine Similarity*.
4. Menjalankan Proses *Clustering*
 - a Jalankan model untuk menghasilkan kluster dan RapidMiner akan mengelompokkan data dan menampilkan hasil klusterisasi dalam bentuk *table* dan grafik.
5. Evaluasi Hasil
 - a Gunakan operator *Performance* untuk menilai kualitas klusterisasi.
 - b Lihat dan analisis nilai Davies Bouldin Index (DBI) untuk setiap nilai clusternya.

2.3 Davies-Bouldin Index

Davies-Bouldin Index (DBI) adalah metrik evaluasi yang digunakan untuk menghitung kualitas klusterisasi pada algoritma *K-Means* [17]. Matriks ini menghitung seberapa baik *cluster* dibentuk dengan mengukur jarak antar cluster (*inter-cluster distance*) dan kerapatan dalam *cluster* (*intra-cluster distance*). Semakin rendah nilai DBI, semakin optimal hasil pengelompokan yang dihasilkan.

2.4 Magnitudo Vektor (*Norma Eucliden*)

Magnitudo vektor adalah teknik untuk mengukur kekuatan pengaruh atribut dalam ruang multidimensi dengan menghitung norma Euclidean dari koordinat vektor atribut [18]. Dalam konteks clusterisasi, magnitudo ini menggambarkan seberapa dominan suatu atribut dalam membentuk perbedaan antar-cluster. Semakin tinggi magnitudo, semakin besar kontribusi atribut tersebut terhadap *cluster*. Rumus Magnitudo :

$$\text{Magnitudo} = \sqrt{(X_i)^2 + (Y_i)^2 + (Z_i)^2}$$

Di mana:

X_i, Y_i, Z_i = Nilai koordinat atribut ke-n.

3. HASIL DAN PEMBAHASAN

Tahap ini menyajikan temuan berdasarkan data yang telah diolah serta memberikan interpretasi terhadap hasil tersebut. Pada penelitian ini, proses analisis dilakukan dengan menggunakan metodologi *Knowledge Discovery in Database (KDD)* yang diterapkan melalui aplikasi RapidMiner untuk melakukan klusterisasi kualitas pendidikan jenjang Sekolah Dasar (SD) per provinsi di Indonesia menggunakan algoritma *K-Means*.

3.1 Dataset

Dataset diperoleh dari Dataset Kaggle Pendidikan SD 2023-2024 dan Data UMP Indonesia 2024 dari Kemnaker. Data awal yang didapat dari kaggle dan Kemnaker memiliki 15 atribut yaitu Provinsi, Sekolah, Siswa, Mengulang, Putus Sekolah, Kepala Sekolah dan Guru(<S1), Kepala Sekolah dan Guru(≥S1), Tenaga Kependidikan (SM), Tenaga Kependidikan (>SM), Rombongan Belajar, Ruang kelas (baik), Ruang kelas (rusak ringan), Ruang kelas (rusak sedang), Ruang kelas (rusak berat), dan UMP. Serta dataset memiliki *record* sebanyak 38 buah.

3.2 Selection

Dataset yang diperoleh kemudian dilakukan Selection sehingga sesuai dengan kebutuhan penelitian dan penerapan KDD yang direncanakan pada metodologi penelitian. Atribut yang diperoleh dari Kaggle Pendidikan SD 2023 - 2024 dan Data UMP Indonesia 2024 dari Kemnaker telah dipertimbangkan untuk tidak digunakan semua dalam proses pengolahan data. Sehingga diperlukan proses selection atribut. Proses *Selection* dilakukan dengan aplikasi excel.

Nomor	Provinsi	Sekolah	Siswa	Putus Sekolah	Kepala Sekolah dan Guru	Ruang kelas(baik)	UMP	Cluster
1	Prov. D.K.I. Jakarta	2.224,00	756.455,00	869,00	39.394,00	27.206,00	5.067.381,00	2
2	Prov. Jawa Barat	19.553,00	4.487.262,00	5.675,00	214.259,00	105.598,00	2.057.495,00	1
3	Prov. Jawa Tengah	18.618,00	2.595.050,00	2.399,00	169.872,00	93.406,00	2.036.947,00	1
4	Prov. D.I. Yogyakarta	1.846,00	270.610,00	75,00	20.097,00	12.095,00	2.125.898,00	0
5	Prov. Jawa Timur	18.980,00	2.530.849,00	3.327,00	183.415,00	101.575,00	2.165.244,00	1
6	Prov. Aceh	3.532,00	489.863,00	905,00	49.552,00	17.744,00	3.460.672,00	2
38	Prov. Papua Selatan	606,00	94.976,00	936,00	4.470,00	2.462,00	4.024.270,00	2
39	Prov. Papua Pegunungan	665,00	140.165,00	928,00	4.094,00	2.543,00	4.024.270,00	2
40	Prov. Papua Barat Daya	547,00	60.227,00	334,00	5.193,00	2.490,00	3.393.500,00	2

Gambar 3. Tampilan Selection pada Aplikasi Excel

Gambar 3 menunjukkan cara untuk memilih atribut untuk digunakan dalam pengolahan data. Atribut yang dipilih dalam pengolahan data terdiri dari 7 atribut yaitu Provinsi, Sekolah, Siswa, Putus Sekolah, Kepala Sekolah dan Guru, Ruang Kelas (baik) Ruang Kelas (baik) yang digabung dari *field* Ruang kelas (baik) dan Ruang kelas (rusak ringan), serta UMP.

3.3 Preprocessing

Preprocessing dilakukan untuk mengecek data rusak dan data tidak konsisten menjadi data yang dapat olah. Pada *Preprocessing* juga melakukan data reduction untuk data yang tidak perlu digunakan.

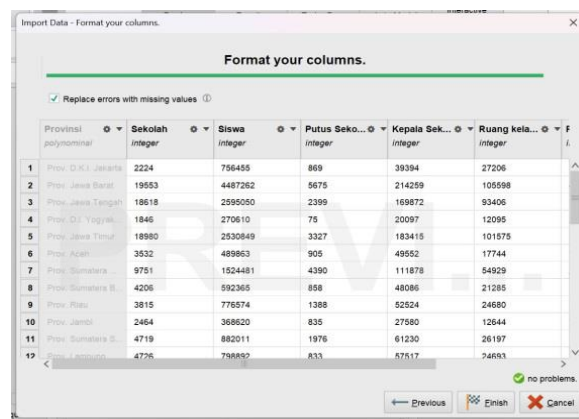
Gambar 4. Data Reduction

Gambar 5. Preprocessing

Pada Gambar 5 *preprocessing* data dilakukan dengan menggunakan operator *read excel* serta mengklik *replace error with missing values*. Proses ini juga melakukan *exclude* melakukan perbaikan data yang tidak konsisten.

3.4 Transformations

Tahap *transformations* data melibatkan konversi nilai dari polinomial ke integer dan normalisasi data untuk mencapai skala yang konsisten, yang menandai akhir dari proses penyiapan data. Penyiapan data dimulai dengan identifikasi kumpulan data yang salah, diikuti dengan pemilihan karakteristik dan manipulasi data berikutnya [18]. *Transformations* data yang dilakukan dalam penelitian ini mengubah data dari polynominal menjadi data numeric yaitu pada atribut Provinsi, Sekolah, Siswa, Putus Sekolah, Kepala Sekolah dan Guru, Ruang kelas (baik), Ruang kelas (rusak), dan UMP. Kemudian dilakukan normalisasi pada atribut-atribut Provinsi, Sekolah, Siswa, Putus Sekolah, Kepala Sekolah dan Guru, Ruang kelas(baik), Ruang kelas (rusak), dan UMP. Kemudian dilakukan normalisasi pada atribut-atribut tersebut.



Gambar 6. Tahapan Transformations Data

Gambar 6 menunjukkan desain proses transformations data dengan hasil proses diketahui setiap atribut memiliki value numeric integer.

id	Sekolah	Siswa	Putus Sekolah	Kepala Sekolah dan Guru	Ruang kelas(baik)	UMP
1	-0.346	0.142	-0.268	-0.066	0.247	3.045
2	3.198	4.426	3.839	3.469	3.266	-1.647
3	3.007	2.253	1.039	2.571	2.797	-1.679
4	-0.423	-0.416	-0.947	-0.456	-0.335	-1.540
5	3.081	2.179	1.832	2.845	3.111	-1.479
6	-0.078	-0.164	-0.238	0.140	-0.118	0.540
7	1.194	1.024	2.741	1.399	1.315	-0.474
8	0.060	-0.046	-0.278	0.110	0.019	-0.472
9	-0.020	0.165	0.175	0.200	0.149	0.281
10	-0.297	-0.303	-0.298	-0.304	-0.314	-0.120
11	0.164	0.286	0.678	0.376	0.208	0.534
12	0.166	0.191	-0.299	0.301	0.150	-0.620
13	0.109	-0.080	0.120	-0.051	-0.053	-0.641
14	-0.259	-0.409	-0.434	-0.339	-0.339	0.230
15	-0.203	-0.321	-0.556	-0.253	-0.220	0.263

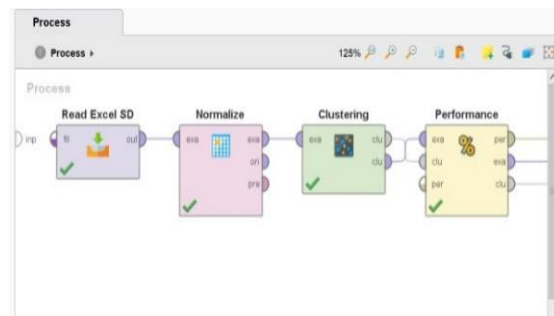
ExampleSet (38 examples, 2 special attributes, 6 regular attributes)

Gambar 7. Output Normalisasi

Gambar 7 menunjukkan output normalisasi untuk menghindari bias atribut dengan skala besar terhadap atribut dengan skala kecil. Data yang telah ditransformasi dan dinormalisasi siap di mining menggunakan metode Clustering dengan algoritma K-Means.

3.5 Data Mining

Dataset yang telah melalui tahapan *preprocessing*, *transformations*, dan *normalization* maka siap untuk tahapan berikutnya yaitu *mining process*. Jumlah dataset yang valid saat ini sebanyak 38 record dengan 7 attributes yang kemudian dijalankan pada RapidMiner Clustering K-Means seperti gambar



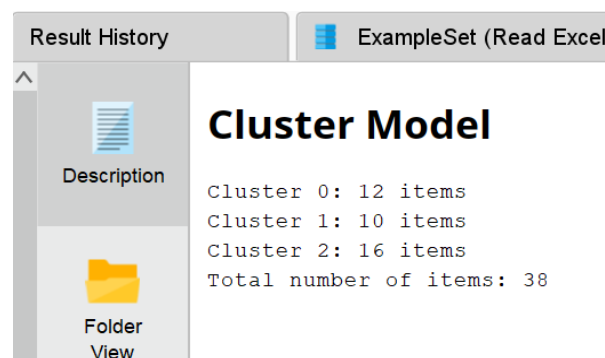
Gambar 8. Mining Process

Pengujian performa algoritma dalam mengolah data dilakukan sebanyak enam kali dengan menentukan nilai k yang berbeda.

Tabel 1. Cluster DBI

Nilai K	DBI
2	0,172
3	0,162
4	0,177
5	0,167

Pada pengujian pertama dengan $k=2$, diperoleh nilai Davies-Bouldin Index (DBI) sebesar 0,172. Pengujian dilanjutkan dengan $k=3$ hingga $k=5$, dengan hasil masing-masing 0,162 untuk $k=3$, 0,177 untuk $k=4$, dan 0,167 untuk $k=5$. Untuk menentukan performa terbaik dalam pengolahan data, dipilih nilai *Davies-Bouldin Index* terkecil, karena semakin kecil nilai ini, semakin baik kualitas *cluster* yang terbentuk. Berdasarkan hasil pada tabel, nilai *Davies-Bouldin Index* terkecil diperoleh pada $k=3$ dengan nilai 0,162. Dengan demikian, dapat disimpulkan bahwa membagi data menjadi tiga cluster dan memberikan hasil seperti pada gambar 9.



Gambar 9. Cluster Model Pengelompokan

Gambar 9 menampilkan keanggotaan dari masing-masing *Cluster*. Pada *Cluster 0* keanggotaan terdiri dari 12 items artinya dalam *Cluster 0* memiliki 12 Provinsi yang dikategorikan dalam satu kelompok. *Cluster 1* memiliki 10 items artinya dalam *Cluster 1* terdapat 10 Provinsi serta *Cluster 2* memiliki 16 items artinya dalam *Cluster 2* terdapat 16 Provinsi.

Attribute	cluster_0	cluster_1	cluster_2
Sekolah	-0.298	1.172	-0.509
Siswa	-0.349	1.150	-0.457
Putus Sekolah	-0.387	1.224	-0.474
Kepala Sekolah dan Guru	-0.324	1.230	-0.526
Ruang kelas(baik)	-0.347	1.184	-0.480
UMP	-0.570	-0.663	0.842

Gambar 10. Tabel Centroid

Pada Gambar 10 juga menunjukkan tabel centroid hasil dari proses K-Means Clustering yang merupakan tabel dalam menunjukkan nilai rata-rata atau titik pusat (centroid) dari setiap atribut dalam masing-masing *cluster* yang terbentuk melalui proses K-Means Clustering.

3.6 Interpretation/Evaluation

Pada tahap ini, dilakukan proses analisis *cluster* menggunakan metode clusterisasi *K-Means* untuk mengelompokkan provinsi di Indonesia berdasarkan sejumlah indikator pendidikan dan ekonomi, yaitu jumlah sekolah, jumlah siswa, angka putus sekolah, jumlah kepala sekolah dan guru, jumlah ruang kelas dalam kondisi baik, serta Upah Minimum Provinsi (UMP).

Menggunakan metode evaluasi Davies-Bouldin Index (DBI) yang mendekati 0 menunjukkan kualitas cluster yang lebih baik karena mengindikasikan bahwa *cluster* memiliki penyebaran internal yang kecil dan terpisah jelas dari cluster lainnya [19]. Pada analisis ini, diperoleh nilai Davies-Bouldin Index sebesar 0,162, yang menunjukkan bahwa *cluster* yang dihasilkan memiliki kualitas yang baik dengan $k = 3$.

Criterion	Davies Bouldin
Avg. within centroid dis...	
Avg. within centroid dis...	
Avg. within centroid dis...	
Avg. within centroid dis...	
Davies Bouldin	Davies Bouldin: 0.162

Gambar 11. Nilai DBI

Hasil klasterisasi kemudian membagi provinsi menjadi tiga kalster sebagai berikut:

Provinsi	Sekolah	Siswa	Putus Sekolah	Kepala Sekolah dan Guru	Ruang kelas(baik)	UMP	Cluster
Prov. D.K.I. Jakarta	2224	756455	869	39394	27206	5067381	2
Prov. Jawa Barat	19553	4487262	5675	214259	105596	2057495	1
Prov. Jawa Tengah	18618	2595050	2399	169872	93406	2036947	1
Prov. D.I. Yogyakarta	1846	270610	75	20097	12095	2125888	0
Prov. Jawa Timur	18980	2530849	3327	183415	101575	2165244	1
Prov. Aceh	3532	489863	905	49552	17744	3460672	2
Prov. Sumatera Utara	9751	1524481	4390	111878	54929	2809915	1
Prov. Sumatera Barat	4206	592365	858	48086	21285	2811449	0
Prov. Riau	3815	776574	1388	52524	24680	3294626	1
Prov. Jambi	2464	368620	835	27580	13644	3037122	0
Prov. Sumatera Selatan	4719	882011	1976	61230	26197	3458874	1
Prov. Lampung	4726	798892	833	57517	24693	2716497	1
Prov. Kalimantan Barat	4447	563178	1324	40115	19418	2702616	0
Prov. Kalimantan Tengah	2650	276421	675	25883	11988	3261616	2
Prov. Kalimantan Selatan	2920	353174	532	30139	15093	3282812	2
Prov. Kalimantan Timur	1929	419472	669	26416	14109	3360858	2
Prov. Sulawesi Utara	2209	218823	435	19725	9962	3545009	2
Prov. Sulawesi Tengah	2944	312352	1033	28430	12977	2736998	0
Prov. Sulawesi Selatan	6434	883153	1613	70646	28057	3434298	1
Prov. Sulawesi Tenggara	2347	294170	784	27562	10237	2885964	0
Prov. Maluku	1815	206984	1076	18371	6999	2949953	0
Prov. Bali	2409	383616	179	25443	13481	2813672	0
Prov. Nusa Tenggara Barat	3330	512620	1235	43933	15613	2444067	0
Prov. Nusa Tenggara Timur	5225	663307	2771	55377	24912	2188826	1
Prov. Papua	831	116836	651	7643	3368	4024270	2
Prov. Bengkulu	1407	194870	390	10862	7603	2507079	0
Prov. Maluku Utara	1317	141138	630	12087	4711	3200000	2
Prov. Banten	4610	1203432	1779	58359	31338	2727812	1
Prov. Kepulauan Bangka Belitung	833	161703	332	9775	5088	3640000	2
Prov. Gorontalo	931	108058	458	8614	4523	3025100	0
Prov. Kepulauan Riau	982	227249	385	14432	7950	3402492	2
Prov. Papua Barat	569	71737	474	5428	2839	3393500	2
Prov. Sulawesi Barat	1229	143300	511	14113	4635	2941958	0
Prov. Kalimantan Utara	485	79877	188	6217	2863	3361653	2
Prov. Papua Tengah	553	143052	1106	5128	3144	4024270	2
Prov. Papua Selatan	606	94976	936	4470	2462	4024270	2
Prov. Papua Pegunungan	665	140165	928	4094	2543	4024270	2
Prov. Papua Barat Daya	547	60227	334	5193	2490	3393500	2

Gambar 12. Hasil Klasterisasi

- Cluster 0: Provinsi D.I. Yogyakarta, Provinsi Sumatera Barat, Provinsi Jambi, Provinsi Kalimantan Barat, Provinsi Sulawesi Tengah, Provinsi Sulawesi Tenggara, Provinsi Maluku, Provinsi Bali, Provinsi Nusa Tenggara Barat, Provinsi Bengkulu, Provinsi Gorontalo, Provinsi Sulawesi Barat.
- Cluster 1: Provinsi Jawa Barat, Provinsi Jawa Tengah, Provinsi Jawa Timur, Provinsi Sumatera Utara, Provinsi Riau, Provinsi Sumatera Selatan, Provinsi Lampung, Provinsi Sulawesi Selatan, Provinsi Nusa Tenggara Timur, Provinsi Banten.
- Cluster 2: Provinsi D.K.I. Jakarta, Provinsi Aceh, Provinsi Kalimantan Tengah, Provinsi Kalimantan Selatan, Provinsi Kalimantan Timur, Provinsi Sulawesi Utara, Provinsi Papua, Provinsi Maluku Utara, Provinsi Kepulauan Bangka Belitung, Provinsi Kepulauan Riau, Provinsi Papua Barat, Provinsi Kalimantan Utara, Provinsi Papua Tengah, Provinsi Papua Selatan, Provinsi Papua Pegunungan, Provinsi Papua Barat Daya.

Dengan karakteristik total *clusternya* sebagai berikut:

Cluster	Sekolah	Siswa	Putus Sekolah	Persentase putus sekolah	Kepsek & guru	Ruang kelas (baik)	rata-rata UMP
0	29475	3950743	8758	0,222	319206	141510	Rp2.748.464,67
1	96431	16345011	26151	0,16	1035077	515385	Rp2.688.653,40
2	22852	3751168	10049	0,268	266178	133570	Rp3.654.160,25

Gambar 13. Karakteristik Total Atribut *Cluster*

Sehingga dapat disimpulkan bahwa:

Tabel 2. *Clusters*

Cluster	Ciri-ciri Utama
Cluster 0	Memiliki jumlah sekolah dan siswa yang cukup tinggi, dengan angka putus sekolah yang median. Namun, rata-rata UMP di cluster ini tidak terlalu tinggi dibandingkan dengan cluster lainnya.
Cluster 1	Menonjol dengan jumlah sekolah, siswa, tenaga pengajar, dan ruang kelas yang paling besar. Meskipun demikian rata-rata UMP merupakan yang terendah di antara semua cluster.
Cluster 2	Memiliki angka putus sekolah tertinggi, dengan persentase putus sekolah terbesar di antara semua cluster. Selain itu, jumlah sekolah, siswa, tenaga pengajar, dan ruang kelas yang tersedia lebih sedikit dibandingkan cluster lainnya. Namun, rata-rata UMP di cluster ini merupakan yang tertinggi.
Cluster 2	Memiliki angka putus sekolah tertinggi, dengan persentase putus sekolah terbesar di antara semua cluster. Selain itu, jumlah sekolah, siswa, tenaga pengajar, dan ruang kelas yang tersedia lebih sedikit dibandingkan cluster lainnya. Namun, rata-rata UMP di cluster ini merupakan yang tertinggi.

3.7 Pembahasan

Berdasarkan analisis di atas yang telah dilakukan, dapat disimpulkan bahwa Upah Minimum Provinsi (UMP) tidak secara langsung memengaruhi kualitas pendidikan di setiap *cluster*. Meskipun *cluster 2* memiliki nilai UMP tertinggi, hal ini tidak diikuti oleh kualitas pendidikan yang lebih baik. Hal ini terbukti dari rendahnya jumlah sekolah, ruang kelas, dan tingginya angka putus sekolah terutama di wilayah Papua. Sebaliknya, cluster 1 yang memiliki UMP yang lebih kecil justru menunjukkan performa yang lebih baik daripada ketiga cluster

Hal ini selaras dengan gambar 4.8 yang menunjukkan tabel centroid hasil dari proses K-Means Clustering sebelumnya dengan mencari atribut yang paling mempengaruhi dalam pembentukan cluster, dapat menggunakan *magnitudo vektor* (norma Euclidean) dari setiap atribut *centroid* tersebut hingga

menghasilkan tabel seperti dibawah.

Atribut	Atribut Clusters	Magnitudo
Kepala Sekolah dan Guru	$\sqrt{(-0,324)^2 + (1,230)^2 + (-0,526)^2}$	1,376
Putus Sekolah	$\sqrt{(-0,387)^2 + (1,224)^2 + (-0,474)^2}$	1,368
Ruang kelas (baik)	$\sqrt{(-0,347)^2 + (1,184)^2 + (-0,480)^2}$	1,324
Sekolah	$\sqrt{(-0,298)^2 + (1,172)^2 + (-0,509)^2}$	1,312
Siswa	$\sqrt{(-0,349)^2 + (1,150)^2 + (-0,457)^2}$	1,286
UMP	$\sqrt{(-0,570)^2 + (-0,663)^2 + (0,842)^2}$	1,214

Gambar 14. Atribut Clusters

Berdasarkan analisis magnitudo vektor euclidean dapat disimpulkan bahwa atribut atau faktor “Kepala Sekolah dan Guru” adalah yang paling dominan dalam penyusunan klaster kualitas pendidikan sekolah dasar (SD) di Indonesia, dengan magnitudo tertinggi (1,376). Diikuti oleh “Putus Sekolah” (1,368) dan “Ruang kelas (baik)” (1,324), ketiga elemen ini secara bersama menunjukkan betapa pentingnya sumber daya manusia pendidik, serta kualitas ruang belajar untuk mengelompokkan provinsi berdasarkan kualitas pendidikan. Namun, atribut seperti “UMP” atau Upah Minimum Provinsi, memiliki pengaruh paling rendah (1,214), menunjukkan bahwa dalam situasi ini, faktor ekonomi tidak secara langsung mendominasi segmentasi klaster.

Berdasarkan hasil analisis, terdapat beberapa rekomendasi penting untuk meningkatkan kualitas pendidikan dasar secara lebih merata di Indonesia:

- **Pembangunan Infrastruktur Pendidikan**
Pemerintah perlu memprioritaskan pembangunan infrastruktur pendidikan, terutama di daerah dengan kondisi pendidikan yang masih tertinggal, seperti di Klaster 1. Fokus utama adalah peningkatan jumlah sekolah dan perbaikan ruang kelas untuk menciptakan lingkungan belajar yang lebih baik.
- **Peningkatan Kualitas dan Kesejahteraan Tenaga Pendidik**
Kualitas dan kesejahteraan tenaga pendidik harus menjadi perhatian utama. Program pelatihan berkala dan peningkatan insentif bagi tenaga pendidik sangat diperlukan untuk meningkatkan kualitas pengajaran. Kebijakan insentif tambahan juga dapat diberikan terutama kepada guru di daerah dengan angka putus sekolah tinggi.
- **Intervensi Khusus untuk Wilayah dengan Tingkat Putus Sekolah Tinggi**
Di wilayah seperti Papua dan sekitarnya, diperlukan program khusus untuk menekan angka putus sekolah, seperti beasiswa, kampanye pendidikan, dan bantuan biaya sekolah. Intervensi ini bertujuan untuk memastikan bahwa semua anak memiliki akses yang sama terhadap pendidikan dasar berkualitas.

4. KESIMPULAN

Penelitian ini bertujuan mengelompokkan 38 provinsi di Indonesia berdasarkan karakteristik pendidikan dasar menggunakan metode K-Means Clustering dengan 7 atribut utama. Hasil analisis menunjukkan pembentukan tiga klaster, yaitu: Klaster 0 dengan 12 provinsi yang memiliki jumlah sekolah dan siswa cukup tinggi serta angka putus sekolah sedang (0,222%) dan UMP rata-rata Rp2.748.464,67; Klaster 1 dengan 10 provinsi yang memiliki jumlah sekolah, siswa, tenaga pengajar, dan ruang kelas terbesar, angka putus sekolah terendah (0,160%), namun UMP terendah (Rp2.688.653,40); dan Klaster 2 dengan 16 provinsi yang memiliki angka putus sekolah tertinggi (0,268%) dan UMP tertinggi (Rp3.654.160,25). Evaluasi menggunakan Davies-Bouldin Index (DBI) dengan nilai 0,162 menunjukkan klasterisasi memiliki kualitas baik. Perbedaan kualitas pendidikan antar provinsi dipengaruhi oleh tenaga pendidik, angka putus sekolah, kondisi ruang kelas, dan tingkat

kesejahteraan guru.

Berdasarkan analisis magnitudo vektor Euclidean, faktor-faktor yang berpengaruh dalam pembentukan klaster dari yang terbesar sampai terkecil adalah:

- Kepala Sekolah dan Guru
- Putus Sekolah
- Ruang Kelas
- Sekolah
- Siswa
- UMP

Penelitian selanjutnya disarankan menambahkan variabel lain seperti rasio guru terhadap siswa, tingkat kelulusan, dan kualitas fasilitas pembelajaran, serta menggunakan metode klasterisasi alternatif atau teknik predictive modeling untuk analisis yang lebih komprehensif.

Daftar Pustaka

- [1] Muliastri, N. K. E. (2020). New Literacy sebagai upaya peningkatan mutu pendidikan sekolah dasar di abad 21. *PENDASI: Jurnal Pendidikan Dasar Indonesia*, 4(1), 115-124.
- [2] P. B. Pastika, *Dataset pendidikan SD Indonesia 2023–2024*, Kaggle, 2023. [Online]. Available: <https://www.kaggle.com/datasets/puanbeningpastika/dataset-pendidikan-sd-indonesia-2023-2024/data>
- [3] Toyamah, N., dan Usman, S. (2004). Alokasi Anggaran Pendidikan di Era Otonomi Daerah: Implikasinya terhadap Pengelolaan Pelayanan Pendidikan Dasar. Jakarta: Laporan Lapangan SMERU. Lembaga Penelitian SMERU
- [4] Basrowi, & Juariyah, S. (2010). Analisis Kondisi Sosial Ekonomi dan Tingkat Pendidikan Masyarakat Desa Srigading, Kecamatan Labuhan Maringgai, Kabupaten Lampung Timur. *Jurnal Ekonomi & Pendidikan*, 7(1), 58-79.
- [5] Robbani, K. R., & Fatah, Z. (2024). Clusterisasi Pendidikan SD untuk Mengetahui Daerah dengan Pendidikan Terendah Menggunakan Algoritma K-Means. *Jurnal Ilmiah Multidisiplin Ilmu*, 1(5), 74-80.
- [6] Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika dan Ilmu Komputer (JIMA-ILKOM)*, 3(1), 46-56.
- [7] Hidayat, R. (2022). Analisis Klasifikasi Sekolah Berdasarkan Faktor Akademik Menggunakan Algoritma Clustering. *Jurnal Teknologi dan Sistem Informasi*, 5(2), 112-121.
- [8] Alodia, D. A., et al. (2021). Implementasi Metode K-Medoids Clustering untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan. *SEPREN: Journal of Mathematics Education and Applied*, 2(2), 1-13.
- [9] Idris, M., Adam, R. I., Brianorman, Y., Munir, R., & Mahayana, D. (2022). Kebenaran dalam perspektif filsafat ilmu pengetahuan dan implementasi dalam Data Science dan Machine Learning. *Jurnal Filsafat Indonesia*, 5(2).
- [10] Amna, W., et al. (2023). *Data mining: Teknik eksplorasi data untuk menemukan pola dan hubungan dalam kumpulan data besar*. PT Global Eksekutif Teknologi.
- [11] Jalil, A., et al. (2024). Implementasi algoritma Support Vector Machine untuk klasifikasi status stunting pada balita. *G-Tech: Jurnal Teknologi Terapan*, 8(3), 2070-2079 .
- [12] Pattipeilohy, R. L., & Pakereng, M. A. I. (2023). Penerapan K-Means Clustering pada Data Mahasiswa Fakultas Interdisiplin Program Studi D4 Destinasi Pariwisata untuk Menentukan Strategi Promosi. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 7(1), 320-331.
- [13] Siregar, A. H., et al. (2024). Implementasi Algoritma K-Means Menggunakan RapidMiner untuk Clusterisasi Data Obat pada Rumah Sakit Royal Prima. *Jurnal Teknologi dan Ilmu Komputer Prima (JUTIKOMP)*, 7(2), 200-210.
- [14] Nurahman, Purwanto, A., & Mulyanto, S. (2022). Klasterisasi sekolah menggunakan algoritma K-Means berdasarkan fasilitas, pendidik, dan tenaga pendidik. *Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, 21(2), 337-350.

<https://doi.org/10.30812/matrik.v21i2.1411>

- [15] Lenda, Julita Veronika., Magdalena A. Ineke Pakereng. (2024). Analisis Tingkat Pengangguran Di Kota Palopo Menggunakan Metode K-Means. *ZONAsi Jurnal Sistem Informasi*, 6(2), 400-409.
- [16] Wibowo, A., et al. (2021). Penentuan Cluster Koridor TransJakarta dengan Metode Majority Voting pada Algoritma Data Mining. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(3), 565-575.
- [17] Ikotun, A. M., et al. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178-210.
- [18] Sa'diah, H., et al. (2023). Penerapan Algoritme K-Means Dalam Segmentasi Daerah Rawan Kekerasan Anak Di Jawa Barat. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(2), 1351-1357.
- [19] Carudin, C., Marisa, M., Murnawan, M., Reba, F., Koibur, M. E., Thantawi, A. M., & Wattimena, F. Y. (2024). *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia.
- [20] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224-227, 1979, doi: 10.1109/TPAMI.1979.4766



ZONAsi: Jurnal Sistem Informasi

Is licensed under a [Creative Commons Attribution International \(CC BY-SA 4.0\)](https://creativecommons.org/licenses/by-sa/4.0/)