AnoaTIK: Jurnal Teknologi Informasi dan Komputer e-ISSN: **2987-7652**
Vol. 4, No. 1, Juni 2026 Hal. 21-28

IMPLEMENTASI K-MEANS DAN DAVIES BOULDIN INDEX UNTUK PENGELOMPOKAN DAERAH DI SULAWESI TENGGARA BERDASARKAN INDIKATOR PENDIDIKAN DENGAN VISUALISASI SISTEM INFORMASI GEOGRAFIS

Sahrul[*1], Natalis Ransi[2], Muhammad Arfan[3]

[1,2,3]Universitas Halu Oleo
Email: [1]shahrul0522@gmail.com, [2]natalis.ransi@uho.ac.id, [3]muhammadarfanbandi@uho.ac.id
***** *Penulis Korespondensi*

Abstrak

Penelitian ini menerapkan algoritma K-Means yang diintegrasikan dengan Davies-Bouldin Index (DBI) untuk mengelompokkan wilayah berdasarkan indikator pendidikan. Metode DBI berfungsi menentukan jumlah klaster optimal secara objektif sebelum proses klasterisasi dengan K-Means dilakukan. Hasil penelitian memperoleh jumlah klaster terbaik k=2 (nilai DBI 0,6035) yang mengelompokkan 17 wilayah menjadi dua klaster yang kohesif. Hasil klastering divalidasi secara visual menggunakan *Principal Component Analysis* (PCA) dan berhasil divisualisasikan dalam peta tematik berbasis Sistem Informasi Geografis (SIG) melalui sistem web yang dibangun dengan *framework Streamlit*. Simpulan penelitian ini adalah kombinasi K-Means dan DBI menghasilkan klasterisasi wilayah yang objektif dan terukur, serta sistem web memudahkan visualisasi hasilnya secara interaktif.

Kata kunci: K-Means, DBI, SIG, Klasterisasi

Abstract

This research implemented the K-Means algorithm integrated with the Davies-Bouldin Index (DBI) to cluster regions based on educational indicators. The DBI method served to determine the optimal number of clusters objectively before the clustering process with K-Means was conducted. The results obtained the best number of clusters at k=2 (DBI value 0,6035), which grouped 17 regions into two cohesive clusters. The clustering results were visually validated using Principal Component Analysis (PCA) and successfully visualized in thematic maps based on a Geographic Information System (GIS) through a web system built with the Streamlit framework. In conclusion, the combination of K-Means and DBI produced an objective and measurable regional clustering, and the web system facilitates interactive visualization of the results.

***Keywords**:* K-Means, DBI, GIS, *Clustering*

1. PENDAHULUAN

Perkembangan teknologi dalam bidang *data mining* memberikan kemampuan untuk menggali pola-pola tersembunyi dari kumpulan data yang besar dan kompleks. Salah satu metode yang populer digunakan dalam analisis eksploratif adalah algoritma K-Means, yang termasuk ke dalam pendekatan klasterisasi tak terawasi (*unsupervised clustering*). Algoritma ini dikenal luas karena kemudahannya dalam implementasi dan efisiensinya dalam menangani data berskala besar [1].

Efektivitas klasterisasi tidak hanya ditentukan oleh algoritma, tetapi juga oleh metode evaluasi yang digunakan untuk mengukur kualitas pemisahan dan kekompakan klaster. Pemilihan metode evaluasi yang tepat sangat penting untuk memperoleh hasil klasterisasi yang optimal, karena dapat membantu menentukan jumlah klaster terbaik secara objektif dan akurat [2].

Penelitian sebelumnya lebih banyak mengandalkan metode visual seperti Elbow Method dan Silhouette Coefficient [3], [4]. Pendekatan visual ini bersifat interpretatif dan subjektif, dengan keputusan akhir yang bergantung pada persepsi pengguna terhadap grafik yang ditampilkan.

Penggunaan Davies-Bouldin Index (DBI) ditawarkan sebagai alternatif evaluasi yang lebih kuantitatif dan objektif. DBI memungkinkan pengukuran yang lebih sistematis terhadap tingkat pemisahan dan kekompakan antar klaster tanpa mengandalkan visualisasi, sehingga hasil evaluasi lebih dapat diandalkan, terutama ketika data yang digunakan kompleks atau memiliki dimensi yang tinggi. DBI memberikan hasil evaluasi berbasis formula matematis, sehingga tidak memerlukan interpretasi visual dan mengurangi subjektivitas [5].

Algoritma K-Means dikombinasikan dengan evaluasi DBI untuk mengelompokkan wilayah di Provinsi Sulawesi Tenggara berdasarkan indikator pendidikan, yaitu Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), dan Angka Melek Huruf (AMH), Angka Partisipasi Sekolah (APS), dan Angka Partisipasi Kasar (APK). Beberapa indikator ini juga digunakan oleh Badan Pusat Statistik (BPS) dalam penghitungan Indeks Pembangunan Manusia (IPM), sehingga menjamin kesesuaian dan keterandalan data yang digunakan

Penelitian ini bertujuan untuk membangun model klasterisasi wilayah berbasis indikator pendidikan menggunakan pendekatan K-Means dan DBI, serta memvisualisasikannya dalam bentuk peta tematik berbasis Sistem Informasi Geografis (SIG). Pendekatan ini diharapkan mampu menyajikan pola distribusi wilayah secara spasial dan informatif. Berdasarkan penelusuran literatur, belum ditemukan penelitian yang menggabungkan metode ini dalam konteks Provinsi Sulawesi Tenggara, sehingga penelitian ini menawarkan kontribusi kebaruan dari sisi penerapan metode pada konteks wilayah yang belum banyak dieksplorasi.

2. TINJAUAN PUSTAKA

2.1 Klastering

Metode klastering adalah sebuah metode untuk mengelompokkan atau mengklaster setiap data yang ada pada sebuah set data [6]. Dalam *machine learning* dan analisis data, normalisasi biasanya dilakukan untuk memastikan bahwa setiap fitur atau variabel memiliki skala yang sama, sehingga tidak ada fitur yang mendominasi proses analisis karena perbedaan satuan atau rentang nilai [7]. Metode ini dapat menggunakan rumus yang dapat dilihat pada Persamaan 1.

$$x_i = \frac{x - x\_\min}{x\_\max - x\_\min} \quad (1)$$

Keterangan:
x_i : nilai hasil normalisasi.
x : nilai asli dari suatu variabel.
x_min : nilai minimum dari suatu variabel.
x_max : nilai maksimum dari suatu variabel.

2.2 Evaluasi DBI

Davies Bouldin Index (DBI) merupakan salah satu metode yang digunakan untuk mengukur validitas klaster dalam suatu metode klasterisasi. Pengukuran dengan DBI memaksimalkan jarak antar klaster dan sekaligus berusaha meminimalkan jarak antar titik dalam suatu klaster [8]. Langkah-langkah DBI sebagai berikut:

$$S_i = \frac{1}{m_i}\sum_{j=1}^{n} d(x_j, c_i) \quad (2)$$

Keterangan:
S_i : menyatakan dispersi (*scatteer*)dari klaster ke-i, yaitu rata-rata jarak antara setiap titik dalam klaster dan *centroidnya*
m_i : jumlah anggota (data point) dalam klaster ke-*i*.
x_j : data ke-*j* yang berada pada klaster ke-*i*.
c_i : *centroid* (titik pusat) dari klaster ke-*i*.
$d(x_j, c_i)$: fungsi jarak antara titik data x_j dan *centroid* c_i, biasanya menggunakan *Euclidean distance* atau metrik jarak lain sesuai kebutuhan.

Mengukur seberapa rapat anggota-anggota klaster terhadap pusatnya. Semakin kecil nilainya, semakin kompak klasternya.

$$D_{ij} = d(c_i, c_j) \quad (3)$$

Keterangan:
D_{ij} : menyatakan jarak (*distance*) antara *centroid* klaster ke-i dan klaster ke-j.
c_i : *centroid* klaster ke-*i.*
c_j : *centroid* dari klaster ke-*j*.
$d(c_i, c_j)$: fungsi jarak antara *centroid* c_i dan c_j, biasanya *Euclidean distance*.

Mengukur seberapa jauh antar klaster. Semakin besar nilainya, semakin terpisah klaster satu dengan lainnya.

$$R_{ij} = \frac{S_i + S_j}{D_{ij}} \quad (4)$$

Keterangan:
R_{ij} : Rasio antara jumlah dispersi *internal* dua klaster (*i* dan *j*) terhadap jarak antar pusat kedua klaster tersebut
S_i : rata-rata jarak titik-titik dalam klaster *i* ke *centroid* c_i (dispersi dalam klaster).
S_j : rata-rata jarak titik-titik dalam klaster *j* ke centroid c_j (dispersi dalam klaster).
D_{ij} : jarak antara *centroid* c_i dan c_j.

Semakin kecil nilai rasio berarti kedua klaster semakin baik terpisah dan masing-masing lebih rapat, semakin besar nilainya berarti klaster makin tumpang tindih.

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max\, j \neq i(R_{ij}) \quad (5)$$

Keterangan:
DBI : nilai indeks Davies-Bouldin, semakin kecil nilainya, semakin baik kualitas klaster.
k : jumlah klaster yang dihasilkan.
R_{ij} : nilai rasio antara klaster I dan klaster j, dihitunga dari rumus Rasio.
$\max_{i \neq j}(R_{ij})$: untuk setiap klaster i, ambil nilai rasio terbesar terhadap klaster lain (*j*) karena itu dianggap sebagai "kemiripan terburuk" klaster tersebut.

Nilai DBI yang lebih rendah menunjukkan kualitas klastering yang lebih baik, karena menunjukkan bahwa klaster-klasternya lebih terpisah dan lebih dekat dengan pusat klasternya [9].

2.3 K-Means

Algoritma K-Means adalah metode *clustering non hierarchical* berbasis jarak yang membagi data dalam klaster dan algoritma ini bekerja pada atribut numerik [10]. Algoritma K-Means termasuk dalam *partitioning clustering* yang memisahkan data ke k daerah bagian yang terpisah. Algoritma K-Means sangat terkenal karena kemudahan dan

kemampuannya untuk mengklaster data besar dan outlier dengan sangat cepat [11].

Algoritma K-Means pada dasarnya melakukan dua proses yaitu proses pendeteksian lokasi pusat klaster (*centroid*) dan proses pencarian anggota dari tiap-tiap klaster [1]. Berikut proses algoritma K-Means.

1. Tentukan k sebagai jumlah klaster terbaik, untuk menentukan banyaknya klaster k dilakukan dengan beberapa pertimbangan seperti pertimbangan teoritis dan konseptual yang mungkin diusulkan untuk menentukan berapa banyak klaster.
2. Tentukan k *centroid* awal secara random, penentuan *centroid* awal dilakukan secara random/acak dari objek-objek yang tersedia sebanyak k klaster, kemudian untuk menghitung *centroid* klaster ke-i berikutnya. Rumus perhitungan untuk menghitung *centroid* klaster dapat dilihat pada Persamaan 6

$$Centroid = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (6)$$

Keterangan:

$Centroid$: titik pusat dari sebuah klaster.

$\frac{1}{n}$: konstanta pembagi, menyatakan bahwa total akan dibagi sebanyak n data.

$\sum$: penjumlahan total.

$i = 1$: indeks awal penjumlahan.

n : jumlah total data dalam satu klaster.

x_i :data ke-i dalam klaster pada dimensi tertentu.

3. Hitung jarak tiap objek ke masing masing klaster, untuk menghitung jarak antara objek dengan *centroid* dapat menggunakan *euclidean distance.* Rumus perhitungan dapat dilihat pada Persamaan 7.

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (7)$$

Keterangan:

$d(x, y)$: Jarak *euclidean* antara dua titik.

$\|x, y\|$: Notasi Panjang (norma) dari vektor selisih antara x dan y.

$\sum_{i=1}^{n}$: Penjumlahan dari indeks i = 1 sampai n.

x_i : Nilai atribut ke-i dari data x.

y_i : Nilai atribut ke-i dari data y.

n :Jumlah dimensi atau atribut yang dibandingkan.

$(x_i - y_i)^2$: Kuadrat dari selisih antara atribut x_i dan y_i.

$\sqrt{\dots}$: Akar kuadrat dari jumlah kuadrat selisih.

4. Alokasikan masing-masing objek ke dalam *centroid* yang paling dekat. Untuk melakukan pengalokasian objek dalam masing-masing klaster pada saat iterasi secara umum dapat dilakukan dengan cara *hard* K-Means dimana secara tegas setiap objek dinyatakan sebagai anggota klaster dengan mengukur jarak kedekatan sifatnya terhadap titik pusat klaster tersebut.
5. Lakukan iterasi, kemudian tentukan posisi *centroid* baru dengan menggunakan Persamaan (6).
6. Ulangi langkah 3 jika posisi *centroid* baru tidak sama.

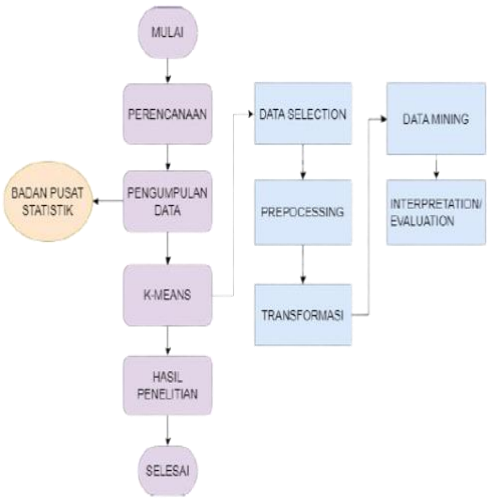
2.4 Sistem Informasi Geografis

Geographic Information System atau Sistem Informasi Geografis (SIG) merupakan suatu sistem informasi berbasis komputer yang digunakan untuk mengelola, menganalisis, dan memvisualisasikan data yang memiliki referensi spasial atau keruangan. SIG menggabungkan fungsi pengolahan data, analisis statistik, dan visualisasi spasial untuk mendukung pengambilan keputusan berbasis lokasi [12].

3. METODE PENELITIAN

3.1 Prosedur Penelitian

Penelitian ini dimulai dengan tahap perencanaan, dilanjutkan dengan pengumpulan data sekunder dari BPS. Setelah itu, data diolah melalui proses seleksi, praproses, dan transformasi sebelum dilakukan analisis klasterisasi. Diagram prosedur penelitian disajikan untuk mempermudah alur kerja penelitian ini. Prosedur penelitian dapat dilihat pada Gambar 1.



Gambar 1 Prosedur Penelitian

3.2 Pengumpulan Data

Pada penelitian ini data diperoleh dari situs resmi BPS Sulawesi Tenggara, dengan fokus pada data indikator pendidikan periode 2019–2022. Data yang telah dikumpulkan dapat dilihat pada Tabel 1

Tabel 1 Data yang Dikumpulkan

No	Wilayah	Periode	RLS	AMH	HLS	APS	APK
1	Buton	2019	7,51	92,1	13,74	77.37	99.71
		2020	7,71	92,08	13,75	77.07	99.36
		2021	7,92	91,12	13,76	77.75	99.71
		2022	8,25	97,14	13,87	78.05	96.96
2	Muna	2019	8,35	90,77	13,78	73.80	96.29
		2020	8,36	93,11	13,79	74.66	96.72
		2021	8,46	88,93	13,8	76.40	97.34
		2022	8.52	93.31	14.01	77.24	97.96
…	…	…	…	…	…	…	…
17	Kota Bau-bau	2022	10.92	97	15.18	75.05	95.20

3.3 Data *Selection*

Data Selection secara selektif menententukan variabel-variabel yang akan digunakan dalam proses analisis berdasarkan ketersediaan data dan kesesuaiannya dengan indikator yang mencerminkan kondisi pendidikan.

3.4. *Prepocessing*

Preprocessing merupakan proses membersihkan dan mempersiapkan data mentah agar siap digunakan dalam analisis. Tahapan ini meliputi.

1. Pengecekan *missing value*
2. Aggregasi data tahunan menjadi rerata periode 2019-2022 per kabupaten.

Data yang telah dilakukan *prepocessing* dapat dilihat pada Tabel 2.

Tabel 2 Data *Prpocessing*

Wilayah	RLS	AMH	HLS	APS	APK
Buton	7.848	93.11	13.78	98.934	77.559
Muna	8.423	91.53	13.845	97.078	75.524
Konawe	9.2	96.008	13.008	92.758	71.49
Kolaka	8.948	96.058	12.865	92.288	71.307
Konawe Selatan	7.995	94.445	12.39	90.491	70.056
Bombana	7.965	93.105	11.853	89.157	65.569
Wakatobi	8.083	92.89	13.43	94.261	73.493
Kolaka Utara	8.185	94.208	12.13	86.52	66.803
Buton Utara	8.928	94.055	12.855	94.002	74.239
Konawe Utara	9.253	96.053	12.89	96.752	72.711
Kolaka Timur	7.725	93.52	12.485	92.746	70.433
Konawe Kepulauan	9.36	96.16	12.18	95.987	71.986
Muna Barat	7.17	90.556	12.455	94.967	71.361
Buton Tengah	7.308	87.99	13.15	99.846	77.846
Buton Selatan	7.508	92.208	13.165	90.064	70.399
Kota Kendari	12.293	98.288	16.673	99.968	85.449
Kota Baubau	10.721	96.29	15.08	96.529	79.865

3.4 Transformasi Data

Transformasi data dalam penelitian ini dilakukan dengan normalisasi Min-Max untuk menyamakan skala variabel sebelum klasterisasi. Hal ini penting dalam klasterisasi K-Means, karena algoritma tersebut sensitif terhadap perbedaan skala antar variabel.

Pemilihan metode ini didasarkan pada hasil visualisasi histogram kelima variabel, terlihat bahwa sebagian besar variabel memiliki distribusi yang tidak normal dan menunjukkan kemiringan (skewness) pada rentang nilai tertentu. perbedaan rentang antarvariabel menjadikan normalisasi (Min-Max *Scaling*) lebih sesuai untuk memastikan seluruh variabel berada pada skala yang seragam dalam algoritma K-Means.

3.5 Penerapan K-Means

Algoritma K-Means digunakan untuk mengelompokkan kabupaten/kota di Provinsi Sulawesi Tenggara berdasarkan indikator pendidikan. Proses ini bertujuan meminimalkan variasi dalam klaster dan memaksimalkan perbedaan antar klaster.

3.6 Proses Data *Mining*

Proses *data mining* difokuskan pada penerapan algoritma K-Means untuk mengidentifikasi pola pengelompokan 17 kabupaten/kota berdasarkan variabel yang ternormalisasi.

3.7 Interpretasi/Visualisasi Hasil Penelitian

Visualisasi dilakukan menggunakan SIG, yang diimplementasikan secara langsung dalam sistem berbasis web melalui integrasi pustaka pemetaan seperti Leaflet.js atau Folium. Visualisasi ini bertujuan untuk menunjukkan distribusi spasial antar klaster secara jelas, interaktif, dan *realtime*, sehingga dapat memudahkan proses interpretasi serta mendukung pengambilan keputusan berbasis lokasi.

4. HASIL DAN PEMBAHASAN

4.1 Analisis Hasil Normalisasi

Visualisasi histogram kelima variabel, terlihat bahwa sebagian besar variabel memiliki distribusi yang tidak normal dan menunjukkan kemiringan (skewness) pada rentang nilai tertentu.

Perbedaan rentang antarvariabel menjadikan normalisasi (Min-Max *Scaling*) lebih sesuai untuk memastikan seluruh variabel berada pada skala yang seragam dalam algoritma K-Means. Nilai normalisasi berkisar antara 0 hingga 1, di mana nilai 1 menunjukkan kondisi terbaik dan 0 menunjukkan kondisi terendah. Adapun hasil dari normalisasi dapat dilihat pada Tabel 3.

Tabel 3 Hasil Normalisasi

Data ke-i	Wilayah	Variabel 2019-2022				
		RLS	AMH	HLS	APS	APK
1	Buton	0,132	0,497	0,4	0,923	0,603
2	Muna	0,245	0,344	0,413	0,785	0,501
3	Konawe	0,396	0,279	0,4	0,464	0,298
4	Kolaka	0,347	0,283	0,201	0,429	0,289
5	Konawe Selatan	0,161	0,227	0,111	0,295	0,226
6	Bombana	0,155	0,497	0,0	0,196	0,0
7	Wakatobi	0,178	0,376	0,327	0,576	0,399
8	Kolaka Utara	0,198	0,204	0,057	0,0	0,062
9	Buton Utara	0,343	0,289	0,208	0,556	0,436
10	Konawe Utara	0,407	0,283	0,215	0,461	0,359
11	Kolaka Timur	0,108	0,237	0,131	0,463	0,245
12	Konawe Kepulauan	0,404	0,293	0,268	0,504	0,323
13	Muna Barat	0,0	0,205	0,125	0,606	0,202
14	Buton Tengah	0,270	0,0	0,238	0,689	0,218
15	Buton Selatan	0,066	0,104	0,072	0,264	0,243
16	Kota Kendari	1,0	1,0	1,0	1,0	1,0
17	Kota Baubau	0,691	0,806	0,67	0,744	0,719

4.2 Analisis Hasil DBI

Evaluasi klastering dilakukan menggunakan DBI, evaluasi klastering objektif menggunakan tiga metrik utama, S_i (dispersi) yang mengukur kerapatan klaster (semakin rendah semakin baik), D_{ij} (*distance*) mengukur pemisahan antar klaster (semakin tinggi semakin baik), dan DBI yang merupakan rasio dari keduanya (semakin rendah semakin baik). DBI sebagai fokus utama karena kemampuannya dalam menyeimbangkan kedua aspek tersebut dalam satu nilai tunggal. Adapun DBI untuk keseluruhan skenario dapat dilihat pada Tabel 4.

Tabel 4 Hasil Evaluasi DBI

k	DBI	Rata-rata Dispersi (S_i)	Jarak Min Antar *Centroid* (D_{ij})
2	0,6035	0,3490	1,1568
3	0,8655	0,3044	0,5827
4	0,7343	0,2497	0,5100
5	0,7914	0,2245	0,4446
6	0,6847	0,1382	0,4347

Hasil evaluasi menunjukkan bahwa nilai DBI terendah diperoleh pada k = 2 dengan nilai sebesar 0,6035. Nilai DBI yang lebih rendah menandakan bahwa konfigurasi klaster tersebut memiliki kombinasi terbaik antara tingkat dispersi dalam klaster dan keterpisahan antar klaster.

Penentuan jumlah klaster dalam penelitian ini sepenuhnya mengacu pada nilai DBI, karena DBI secara langsung menilai tingkat kemiripan terburuk antar klaster. Berdasarkan hasil evaluasi, k = 2 memberikan nilai DBI terendah (0.6035) sehingga dipilih sebagai jumlah klaster terbaik.

4.3 Analisis Hasil K-Means Klastering

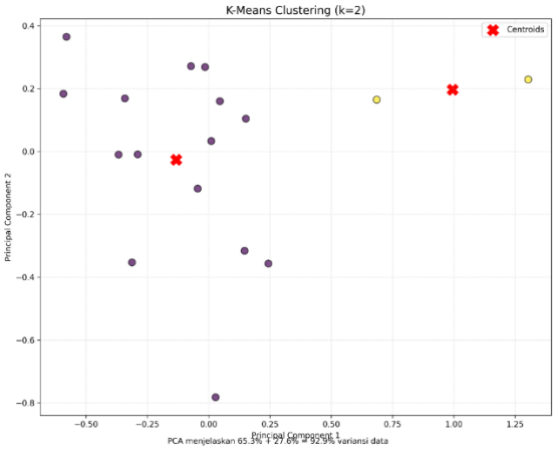
Hasil evaluasi objektif sebelumnya menunjukkan bahwa jumlah klaster optimal yang digunakan adalah k=2 dengan nilai DBI terendah (0,6035). Selanjutnya, algoritma K-Means klastering diterapkan pada data yang telah dinormalisasi untuk mengelompokkan 17 wilayah di Sulawesi Tenggara berdasarkan lima variabel. Hasil analisis K-Means klastering dapat dilihat pada Tabel 5.

Tabel 5 Hasil Evaluasi K-Means

Data ke-i	Wilayah	Variabel 2019-2022					Klaster	Jarak ke *Centroid*	Koordinat *Centroid*
		RLS	AMH	HLS	APS	APK			
1	Buton	0,1	0,4	0,4	0,9	0,6	0	0,52355	[0,2128 0,5312 0,2011

Data ke-i	Wilayah	Variabel 2019-2022					Klaster	Jarak ke *Centroid*	Koordinat *Centroid*
		RLS	AMH	HLS	APS	APK			
2	Muna	32000,24500	97000,34400	00000,41300	23000,78500	03000,50100	0	3693 0,41711655 9	0,5355 0,3261] [0,2128 0,5312 0,2011 0,5355 0,3261]
3	Konawe	39600,39600	77900,77900	24000,24000	46400,46400	29800,29800	0	0,31985908 7	[0,2128 0,5312 0,2011 0,5355 0,3261]
4	Kolaka	34700,34700	78300,78300	21000,21000	42900,42900	28900,28900	0	0,30745392 7	[0,2128 0,5312 0,2011 0,5355 0,3261]
5	Konawe Selatan	16100,16100	62700,62700	11100,11100	29500,29500	22600,22600	0	0,29603556	[0,2128 0,5312 0,2011 0,5355 0,3261]
6	Bombana	15550,15550	49700,49700	00000,00000	19600,19600	00000,00000	0	0,51619530 8	[0,2128 0,5312 0,2011 0,5355 0,3261]
7	Wakatobi	17780,17780	47760,47760	32700,32700	57600,57600	39900,39900	0	0,16441103 6	[0,2128 0,5312 0,2011 0,5355 0,3261]
8	Kolaka Utara	19800,19800	60400,60400	05700,05700	00000,00000	06200,06200	0	0,61852027 6	[0,2128 0,5312 0,2011 0,5355 0,3261]
9	Buton Utara	34300,34300	58900,58900	20800,20800	55600,55600	43600,43600	0	0,18146985 4	[0,2128 0,5312 0,2011 0,5355 0,3261]
10	Konawe Utara	40700,40700	78300,78300	21500,21500	76100,76100	35900,35900	0	0,39123344 6	[0,2128 0,5312 0,2011 0,5355 0,3261]
11	Kolaka Timur	10800,10800	53700,53700	13100,13100	46300,46300	24500,24500	0	0,16648577	[0,2128 0,5312 0,2011 0,5355 0,3261]
12	Konawe Kepulauan	42700,42700	79300,79300	06800,06800	70400,70400	32300,32300	0	0,40123689 5	[0,2128 0,5312 0,2011 0,5355 0,3261]
13	Muna Barat	00,00	02,02	00,00	01,01	06,06	0	0,37437112 6	[0,2128 0,5312 0,2011
14	Buton Tengah	00000,00270	50000,00700	25000,23800	28000,99000	91000,61800	0	0,77994113 5	0,5355 0,3261] [0,2128 0,5312 0,2011 0,5355 0,3261]
15	Buton Selatan	00000,06600	41000,41000	22000,17200	22000,64000	24300,24300	0	0,34965777 5	[0,2128 0,5312 0,2011 0,5355 0,3261]
16	Kota Kendari	11000,00000	11000,00000	11000,00000	11000,00000	11000,00000	1	0,31090365 9	[0,8455 0,903 0,8348 0,8721 0,8596]
17	Kota Baubau	00000,69100	00000,80160	00000,67000	00000,74400	00000,71900	1	0,31090365 9	[0,8455 0,903 0,8348 0,8721 0,8596]

Berdasarkan hasil klaster menggunakan K-Means diperoleh hasil bahwa pada klaster 0 terdapat 15 kabupaten kota, klaster 1 terdapat 2 kabupaten kota. Adapun visualisasi klastering divisualisasikan pada Gambar 2.



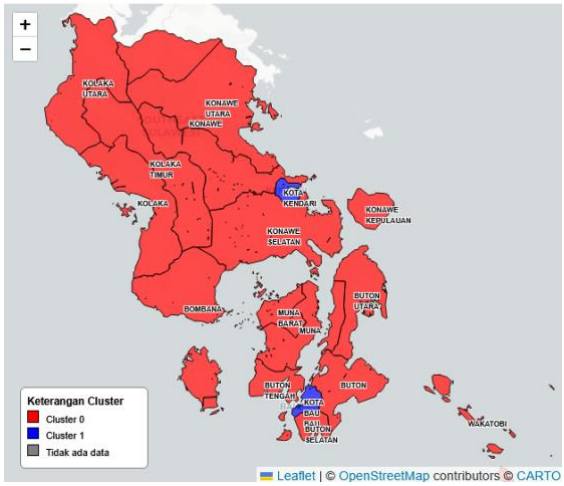
Gambar 2 Visualisasi Klasterisasi K-Means

Terlihat dua kelompok terpisah dengan jelas: Klaster 1 (Kota Kendari dan Kota Baubau) serta Klaster 0 (15 kabupaten). Pemisahan ini konsisten dengan nilai DBI pada k=2, yang menegaskan kekompakan internal dan keterpisahan antar klaster.

4.4 Analisi Hasil Visualisasi Spasial

Upaya memperkuat analisis sekaligus memberikan perspektif spasial dilakukan dengan memvisualisasikan hasil pengelompokan klaster

secara geografis menggunakan SIG. Visualisasi geografis dapat dilihat pada Gambar 3.



Gambar 3 Visualisasi SIG

Peta yang dihasilkan menunjukkan bahwa kedua klaster terpisah dengan jelas secara geografis. Klaster 1 (yang beranggotakan Kota Kendari dan Kota Baubau) terkonsentrasi pada lokasi yang berdekatan di pesisir pantai. Klaster 0 menyebar di berbagai pulau dan wilayah daratan. Hasil visualisasi ini secara efektif mengonfirmasi temuan dari analisis klastering sebelumnya bahwa algoritma K-Means dengan k=2 berhasil memisahkan kedua kelompok dengan baik. Pemisahan yang jelas ini selaras dengan nilai DBI yang rendah (0,6035), yang telah mengindikasikan kualitas pemisahan klaster yang sangat baik.

4.5 Interpretasi Hasil

Interpretasi hasil klastering dilakukan untuk memahami karakteristik pendidikan wilayah berdasarkan pengelompokan yang terbentuk. Tabel profiliasi klaster digunakan sebagai dasar analisis untuk mengidentifikasi perbedaan capaian pendidikan antara Klaster 0 (15 kabupaten) dan Klaster 1 (2 kota), sehingga hasil klastering dapat ditafsirkan secara substantif dalam konteks pendidikan.

Tabel 6 Profiliasi Hasil Klastering

Variabel	Klaster 0	Klaster 1
RLS	8.260	11.5015
AMH	93.460	97.289
HLS	12.822	15.877
APS	72.052	82.657
APK	93.721	98.249

Berdasarkan tabel profiliasi, seluruh indikator pendidikan menunjukkan nilai rata-rata yang lebih tinggi pada Klaster 1. Pada indikator capaian pendidikan, Klaster 1 memiliki nilai RLS sebesar 11,50 tahun dan HLS sebesar 15,88 tahun, lebih tinggi dibandingkan Klaster 0 yang masing-masing sebesar 8,26 tahun dan 12,82 tahun. AMH pada Klaster 1 juga lebih tinggi, yaitu 97,29%, dibandingkan Klaster 0 sebesar 93,46%, yang mencerminkan akumulasi capaian pendidikan yang lebih baik di wilayah perkotaan.

Perbedaan serupa terlihat pada indikator partisipasi pendidikan. Klaster 1 mencatat nilai APS sebesar 82,66% dan APK sebesar 98,25%, lebih tinggi dibandingkan Klaster 0 dengan nilai APS 72,05% dan APK 93,72%. Temuan ini menunjukkan bahwa wilayah perkotaan memiliki tingkat keterlibatan pendidikan yang lebih tinggi, baik dari sisi keberlanjutan maupun cakupan partisipasi pendidikan.

Secara keseluruhan, hasil klastering mencerminkan adanya ketimpangan pendidikan antara wilayah perkotaan dan kabupaten di Provinsi Sulawesi Tenggara. Klastering yang dihasilkan tidak hanya konsisten secara statistik, tetapi juga relevan secara substantif, sehingga dapat memberikan gambaran awal bagi perumusan kebijakan pendidikan daerah yang lebih terarah.

5. KESIMPULAN

Penggunakan algoritma K-Means, 17 wilayah di Provinsi Sulawesi Tenggara berhasil dikelompokkan ke dalam dua klaster utama, yaitu Klaster 1 yang terdiri dari Kota Kendari dan Kota Baubau dengan karakteristik pendidikan yang relatif homogen dan lebih unggul, serta Klaster 0 yang terdiri dari 15 kabupaten dengan karakteristik pendidikan yang lebih beragam.

Perbedaan karakteristik antar klaster mencerminkan adanya kesenjangan capaian pendidikan antara wilayah perkotaan dan kabupaten. Wilayah perkotaan menunjukkan nilai yang lebih tinggi pada indikator pendidikan, seperti RLS, HLS, AMH, serta tingkat partisipasi pendidikan. Kondisi ini menunjukkan bahwa akses terhadap sarana pendidikan, fasilitas pendukung, dan lingkungan pembelajaran di wilayah perkotaan cenderung lebih baik dibandingkan wilayah kabupaten. Hal ini mengindikasikan bahwa kombinasi normalisasi Min–Max, algoritma K-Means, dan evaluasi DBI efektif dalam menghasilkan pengelompokan wilayah yang konsisten dan relevan untuk menggambarkan perbedaan kondisi pendidikan antarwilayah di Provinsi Sulawesi Tenggara.

6. DAFTAR PUSTAKA

[1] N. K. S. Julyantari, K. Budiarta, and N. M. D. K. Putri, "Implementasi K-Means Untuk Pengelompokan Status Gizi Balita (Studi Kasus Banjar Titih)," *J. Janitra Inform. dan Sist. Inf.*, vol. 1, no. 2, pp. 92–101, 2021, doi: 10.25008/janitra .v1i2.134.

[2] O. Pasin and S. Gönenç, "An Investigation

Into Epidemiological Situations of Covid-19 with Fuzzy K-means and K-prototype Clustering Methods," *Sci. Rep.*, vol. 13, 2023, doi: 10.1038/s41598-023-33214-y.

[3] S. Hanifah and A. H. Primandari, "Implementasi Metode K-Means Clustering dalam Pengelompokan Kabupaten/Kota di Provinsi NTB," *Emerg. Stat. Data Sci. J.*, vol. 1, no. 3, pp. 378–393, 2023.

[4] G. C. Messakh, M. N. Hayati, and Sifriyani, "Penerapan Metode K-Means dalam Pengelompokan Kabupaten/Kota di Kalimantan Berdasarkan Indikator Pendidikan," *J. EKSPONENSIAL*, vol. 14, no. 2, pp. 57–66, 2023.

[5] E. Muningsih, I. Maryani, and V. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *EVOLUSI - J. Sains dan Manaj.*, vol. 9, no. 1, pp. 95–100, 2021, doi: 10.31294/evolusi.v9i1.10428.

[6] R. J. Kasim, S. Bahri, and S. Amir, "Implementasi Metode K-Means untuk Clustering Data Penduduk Miskin dengan Systematic Random Sampling," in *Prosiding Seminar Nasional Sistem Informasi dan Teknologi (SISFOTEK)*, 2021, pp. 95–101.

[7] V. V Starovoitov and Y. I. Golub, "Data Normalization in Machine Learning," *Informatics*, 2021, doi: 10.37661/1816-0301-2021-18-3-83-96.

[8] M. Mughnyanti, S. Efendi, and M. Zarlis, "Analysis of Determining Centroid Clustering X-means Algorithm with Davies-Bouldin Index Evaluation," 2020, doi: 10.1088/1757-899X/725/1/012128.

[9] I. T. Umagapi, B. Umaternate, Hazriani, and Yuyun, "Uji Kinerja K-Means Clustering Menggunakan Davies-Bouldin Index pada Pengelompokan Data Prestasi Siswa," in *Prosiding Seminar Nasional Sistem Informasi dan Teknologi (SISFOTEK)*, 2023, pp. 303–308.

[10] M. Norshahlan, H. Jaya, and R. Kustini, "Penerapan Metode Clustering dengan Algoritma K-means pada Pengelompokan Data Calon Siswa Baru," *J. Sist. Inf. TGD*, vol. 2, no. 6, pp. 1042–1053, 2023.

[11] K. P. Sinaga and M. Yang, "Unsupervised K-Means Clustering Algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

[12] Sodikin and E. R. Susanto, "Sistem Informasi Geografis (GIS) Tempat Wisata di Kabupaten Tanggamus," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 3, pp. 125–135, 2021, [Online]. Available: http://jim.teknokrat.ac.id/index.php/sisteminformasi/article/view/881/391.