

PENGUNAAN ALGORITMA DBSCAN DALAM PENGELOMPOKAN KABUPATEN/KOTA DI SULAWESI TENGGARA BERDASARKAN INDIKATOR PENDIDIKAN

Gusti Arviana Rahman^{1*}, Irham Hasbi², Andi Tenriawaru³, La Surimi⁴, Arif Nur Alfian⁵

^{1,5} Program Studi Statistika, Jurusan Matematika, FMIPA, Universitas Halu Oleo

^{2,3,4} Program Studi Ilmu Komputer, Jurusan Matematika, FMIPA, Universitas Halu Oleo

*¹arviana.rahman@uho.ca.id, ²irhamhasbi06@gmail.com, ³andi.tenriawaru@uho.ac.id,
⁴lasurimi@uho.ac.id, ⁵arifnuralfian@uho.ac.id

Pendidikan merupakan pilar utama dalam membangun kemajuan suatu bangsa serta merupakan investasi strategis bagi peningkatan kualitas sumber daya manusia. Di Indonesia, pendidikan menjadi salah satu fokus utama dalam upaya pembangunan nasional. Meskipun pemerintah telah mengimplementasikan program wajib belajar 12 tahun, ketimpangan pendidikan antara wilayah perkotaan dan pedesaan masih cukup mencolok, mencerminkan belum meratanya akses pendidikan di berbagai daerah. Penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Provinsi Sulawesi Tenggara berdasarkan indikator pendidikan menggunakan algoritma DBSCAN. Algoritma ini dipilih karena kemampuannya dalam mengenali pola, membentuk kluster secara dinamis, serta mendeteksi outlier. Hasil analisis menunjukkan pembentukan satu kluster utama dan satu outlier, yaitu Kota Kendari, yang memiliki nilai indikator pendidikan lebih tinggi dibandingkan wilayah lainnya. Temuan ini diharapkan dapat menjadi dasar pertimbangan dalam merumuskan kebijakan pendidikan yang lebih tepat sasaran.

Kata Kunci — Algoritma DBSCAN, Indikator, Pengelompokan, Pendidikan

I. PENDAHULUAN

Pendidikan merupakan fondasi utama dalam membangun kemajuan bangsa sekaligus menjadi investasi strategis dalam peningkatan kualitas sumber daya manusia. Di Indonesia, pendidikan memegang peranan penting dalam proses pembangunan nasional, sebagaimana tercermin dalam komitmen terhadap Sustainable Development Goals (SDGs), khususnya pada tujuan keempat yang menekankan pemerataan akses terhadap pendidikan berkualitas sepanjang hayat. Untuk mendukung hal tersebut, pemerintah telah mengimplementasikan program wajib belajar selama 12

tahun sebagai upaya meningkatkan kualitas dan akses pendidikan[1].

Namun, realita menunjukkan bahwa kesenjangan akses pendidikan masih menjadi persoalan utama di Indonesia. Ketimpangan antara wilayah perkotaan dan pedesaan masih cukup mencolok, terutama dalam hal kualitas dan pemerataan fasilitas pendidikan [2]. Laporan Badan Pusat Statistik (BPS) tahun 2022 juga menegaskan bahwa disparitas pendidikan terjadi di semua jenjang, mulai dari dasar hingga menengah[3].

Lebih lanjut, untuk memperoleh gambaran yang lebih menyeluruh mengenai kondisi pendidikan antarwilayah, diperlukan metode analisis yang mampu mengelompokkan wilayah-wilayah berdasarkan karakteristik indikator pendidikan. Dengan pendekatan pengelompokan (clustering), daerah-daerah dengan karakteristik yang serupa dapat dimasukkan dalam satu kelompok, sehingga perbedaan dan kesenjangan antar kelompok akan lebih terlihat dan mudah dianalisis oleh pemangku kebijakan [4].

Salah satu metode clustering yang banyak digunakan dalam analisis multidimensi adalah Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Algoritma ini termasuk dalam kategori non-hierarki yang membentuk kluster berdasarkan kepadatan data serta memiliki keunggulan dalam mendeteksi outlier tanpa perlu menentukan jumlah kluster di awal [5]. DBSCAN telah berhasil diimplementasikan dalam berbagai domain, seperti pada pengelompokan penyakit pasien [6], data kesehatan [7], maupun pemetaan sebaran minat mahasiswa pasca-pandemi [8]. Penggunaannya pada aplikasi spasial juga ditunjukkan oleh penelitian mengenai deteksi titik panas kebakaran hutan dan lahan menggunakan parameter DBSCAN [9].

Sebagai pembanding, metode clustering lain seperti K-Means dan Average Linkage juga telah banyak digunakan dalam berbagai studi, seperti analisis penjualan berbasis K-Means menggunakan RapidMiner [10] atau pengklasifikasian kualitas air menggunakan algoritma K-

Nearest Neighbor (KNN), Naïve Bayes, dan Decision Tree[11]. Meskipun metode-metode tersebut memiliki keunggulan tertentu, DBSCAN tetap unggul dalam menangani data dengan kepadatan tidak merata serta dalam mengenali noise secara otomatis.

Penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Provinsi Sulawesi Tenggara berdasarkan indikator pendidikan menggunakan algoritma DBSCAN. Dengan menerapkan metode ini, diharapkan dapat diperoleh informasi yang akurat mengenai distribusi karakteristik pendidikan antarwilayah, sekaligus mendeteksi wilayah dengan karakteristik menyimpang (outlier). Hasil analisis ini diharapkan dapat dijadikan acuan dalam penyusunan kebijakan pendidikan yang lebih tepat sasaran dan berbasis data.

II. METODE PENELITIAN

A. Data

Data pada penelitian ini adalah data sekunder yang berasal dari Badan Pusat Statistik (BPS) Sulawesi Tenggara. Data indeks pendidikan yang didapatkan dari BPS merupakan data tahun 2022. Variabel yang digunakan sebanyak 22 buah, variabel tersebut merupakan variabel yang diperoleh melalui tabel/indikator pendidikan. Variabel tersebut dapat dilihat pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Keterangan
APM_SD	Angka Partisipasi Murni SD
APM_SMP	Angka Partisipasi Murni SMP
APM_SMA	Angka Partisipasi Murni SMA
APS7_12	Angka Partisipasi Sekolah 7-12
APS_13.15	Angka Partisipasi Sekolah 13-15
APS_16.18	Angka Partisipasi Sekolah 16-18
APS_19.24	Angka Partisipasi Sekolah 19-24
APK_SD	Angka Partisipasi Kasar SD
APK_SMP	Angka Partisipasi Kasar SMP
APK_SMA	Angka Partisipasi Kasar SMA
HLS	Harapan Lama Sekolah
RLS	Rata-rata Lama Sekolah
AMH	Angka Melek Huruf
SMP	Jumlah SMP
SMA	Jumlah SMA
SMK	Jumlah SMK
GURU_SMP	Jumlah Guru SMP
GURU_SMA	Jumlah Guru SMA
GURU_SMK	Jumlah Guru SMK
MURID_SMP	Jumlah Murid SMP
MURID_SMA	Jumlah Murid SMA
MURID_SMK	Jumlah Murid SMK

Daftar Kabupaten/kota yang ada di Sulawesi Tenggara dapat dilihat pada Tabel 2.

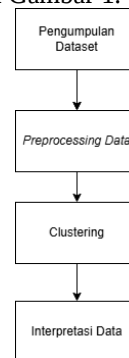
Tabel 2. Kabupaten/Kota di Sulawesi Tenggara

No.	Nama Kabupaten/Kota
1	Buton

2	Muna
3	Konawe
4	Kolaka
5	Konawe Selatan
6	Bombana
7	Wakatobi
8	Kolaka Utara
9	Buton Utara
10	Konawe Utara
11	Kolaka Timur
12	Konawe Kepulauan
13	Muna Barat
14	Buton Tengah
15	Buton Selatan
16	Kota Kendari
17	Kota Baubau

B. Analisis dan Perancangan

Algoritma yang digunakan pada penelitian ini adalah DBSCAN dengan bantuan KNN untuk yang digunakan untuk menentukan nilai epsilon. Implementasi algoritma tersebut dilakukan dengan menggunakan perangkat lunak RStudio. Pengelompokan kabupaten/kota dilakukan melalui tiga tahapan yang diawali dengan pengumpulan data, setelah data berhasil dikumpulkan maka akan dilakukan proses pengelompokan dengan menggunakan algoritma DBSCAN, tahapan terakhir yaitu interpretasi data. Proses untuk melakukan pengelompokan secara umum dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Pengelompokan

DBSCAN

DBSCAN merupakan algoritma pengelompokan yang baik dalam mengatasi *outlier* atau *noise*. DBSCAN menghasilkan kluster yang lebih akurat dan baik pada jumlah data besar dan tidak perlu ditentukan jumlah kluster awalnya. Kepadatan data pada DBSCAN dapat identik dengan inti (*core*), batas (*border*), serta gangguan (*noise*) [9].

Cara kerja algoritma DBSCAN adalah sebagai berikut:

1. Menetapkan nilai parameter, yaitu epsilon (eps) dan minimum points (MinPts)
2. Menentukan titik awal secara acak
3. Menghitung seluruh jarak pada setiap titik dengan menggunakan *euclidean distance*, untuk menghitung *euclidean distance* dapat menggunakan persamaan (1).

$$E(a, b) = \sqrt{\sum_{i=0}^n (a_i - b_i)^2} \quad (1)$$

Keterangan:

$E(a, b)$ = jarak *euclidean* antara dua titik yaitu a dan b

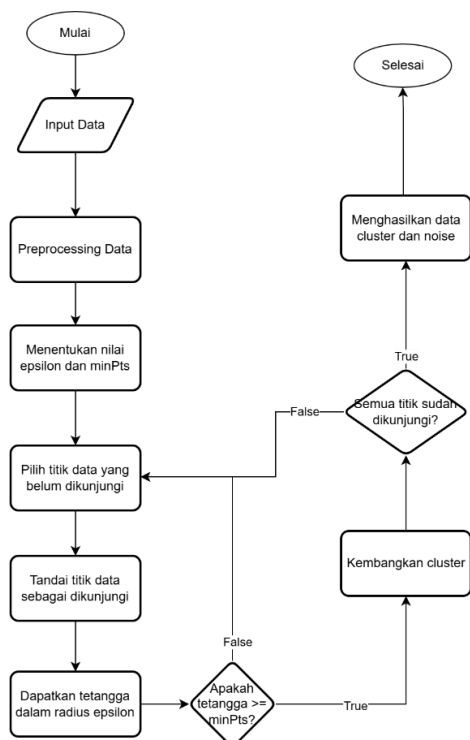
a_i = nilai dari titik data a pada dimensi ke-i

b_i = nilai dari titik data b pada dimensi ke-i

n = jumlah dimensi

4. Klaster tercipta berdasarkan kerapatan dari data.
5. Apabila titik yang tercipta dalam radius epsilon adalah sama atau melebihi nilai MinPts, maka titik p menjadi titik inti, kemudian terciptalah sebuah klaster
6. Apabila titik p merupakan titik batas dan tidak terdapat titik yang kepadatannya dapat digapai oleh titik p, maka proses dilanjutkan pada titik selanjutnya.
7. Tahap 5 dan 6 akan terus diulang sampai semua poin sudah diproses [10].

Proses DBSCAN secara umum dapat dilihat pada Gambar 2.



Gambar 2. Proses DBSCAN

K-Nearest Neighbor

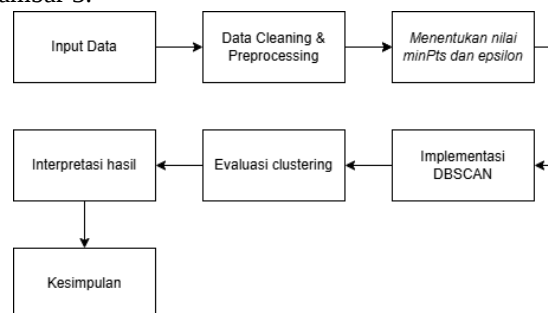
K-Nearest Neighbor (KNN) adalah metode klasifikasi data berdasarkan pembelajaran data yang sudah terklasifikasi sebelumnya. Algoritma KNN merupakan metode mengklasifikasi objek berdasarkan data latih sebelumnya yang memiliki jarak terdekat. Prinsip kerja KNN yaitu menentukan dan mencari jarak terdekat berdasarkan nilai k neighbor terdekat dalam data training dengan data yang akan diuji. Nilai k terbaik pada algoritma ini bergantung pada nilai sebuah data [11].

Pada penelitian ini, untuk menentukan nilai epsilon digunakan bantuan algoritma KNN. Penentuan nilai epsilon dengan bantuan KNN dilakukan dengan mencari *knee point*, yaitu titik di mana kurva pada sebuah plot berubah tajam [12].

Silhouette Score

Silhouette Score atau disebut juga *Silhouette Coefficient* merupakan metrik evaluasi yang digunakan dalam pengukuran untuk mengetahui seberapa jauh hasil dari analisis klasterisasi efektif untuk mencari nilai k yang terbaik. *Silhouette Score* mempunyai rentang nilai dari -1 sampai 1, semakin mendekati nilai 1 menandakan bahwa klaster dipisahkan dengan baik, begitu juga sebaliknya jika nilai mendekati -1 menandakan bahwa klaster dipisahkan dengan buruk [13].

Data yang diperoleh dari BPS akan diolah untuk menghasilkan klaster. Sebelum proses tersebut dilakukan, data akan dipersiapkan terlebih dahulu, tahap persiapan data meliputi pembersihan data kosong serta normalisasi skala pada setiap variabel. Tahap selanjutnya adalah menentukan parameter untuk algoritma DBSCAN yang akan digunakan untuk membagi data menjadi beberapa klaster. Tahapan proses lebih rinci dapat dilihat pada Gambar 3.



Gambar 3. Proses Pembentukan Klaster

Penjelasan tahapan pada Gambar 3 adalah sebagai berikut:

1. **Input data**
Data yang diperoleh dari BPS terdiri dari 22 variabel terkait indikator Pendidikan serta 17 kolom terkait kabupaten/kota yang ada di Sulawesi Tenggara.
2. **Data cleaning dan preprocessing**
Sebelum data digunakan, data harus dibersihkan untuk mencegah adanya nilai kosong. Selain

mengatasi nilai kosong, dilakukan normalisasi data agar data memiliki skala yang sama.

- Menentukan nilai minimum points (MinPts) dan epsilon

Pengelompokan dengan algoritma DBSCAN perlu menggunakan parameter MinPts dan epsilon, oleh karena itu parameter tersebut harus ditentukan terlebih dahulu. Penentuan epsilon dilakukan dengan bantuan algoritma KNN, sedangkan penentuan MinPts akan dilakukan dengan melihat hasil dari *silhouette score*.

- Implementasi DBSCAN

Pengelompokan data akan dilakukan dengan menggunakan algoritma DBSCAN setelah parameter ditentukan. Hasil dari implementasi ini adalah kluster dari kabupaten/kota.

- Evaluasi *Clustering*

Tahap ini merupakan proses untuk mengukur kualitas dari hasil pengelompokan. Pengukuran

dilakukan dengan bantuan *silhouette score*. Tahap ini juga akan membantu dalam pengambilan nilai parameter dengan melakukan beberapa percobaan untuk mengetahui kombinasi parameter mana yang paling baik.

- Interpretasi hasil

Tujuan dari tahap ini adalah untuk memahami kluster yang terbentuk.

- Kesimpulan

Langkah terakhir dari penelitian ini adalah melakukan kesimpulan terkait hasil dari penelitian yang telah dilakukan..

III. HASIL DAN PEMBAHASAN

Data yang digunakan pada penelitian ini merupakan data dalam bentuk numerik. Contoh dari data tersebut dapat dilihat pada Gambar 4.

	APM_SD	APM_SMP	APM_SMA	APS7_12	APS_13.15	APS_16.18	APS19_24	APK_SD	APK_SMP	APK_SMA	HLS	RLS	AMH
1	99.30	78.91	70.47	99.30	97.81	82.70	32.37	113.71	80.08	97.09	13.87	8.25	97.14
2	99.56	82.01	73.33	99.56	96.40	82.94	30.07	109.97	87.03	96.87	14.01	8.52	93.31
3	95.20	76.06	62.54	98.86	93.87	70.61	24.38	108.79	88.51	85.59	13.03	9.30	97.91
4	99.67	71.40	61.40	99.67	94.87	71.29	20.82	114.19	83.63	87.10	13.02	9.06	96.73
5	99.90	76.05	59.95	99.90	97.30	68.77	17.13	105.47	89.30	78.35	12.59	8.27	95.09
6	99.66	74.13	59.58	99.66	88.32	61.94	16.03	111.17	80.65	79.45	11.89	8.05	94.86
	SMP	SMA	SMK	GURU_SMP	GURU_SMA	GURU_SMK	MURID_SMP	MURID_SMA	MURID_SMK				
1	6.546855	7.348243	6.024096	6.355526	7.457930	5.009276	5.160532	5.863994	2.683259				
2	10.269576	11.182109	14.457831	11.221706	11.346252	13.172542	10.753539	9.980680	15.327352				
3	8.600770	8.945687	6.626506	8.125046	8.656298	5.592367	8.338680	9.438228	6.447043				
4	6.803594	4.792332	8.433735	6.023741	5.379908	10.151073	7.486230	6.418123	10.702480				
5	10.783055	9.265176	6.024096	9.017179	8.184600	5.300822	12.011799	10.481731	7.208133				
6	7.573813	8.306709	6.024096	5.655091	5.456400	5.671879	5.495513	5.879918	3.880448				

Gambar 4. Data Sebelum Standarisasi

Data yang telah kumpulan akan diperiksa untuk melihat apakah ada data yang hilang (*missing value*), apakah ada data yang duplikat, serta melakukan

standarisasi. Setelah dilakukan pemeriksaan tidak ditemukan adanya data yang hilang ataupun duplikat seperti yang dapat dilihat pada Gambar 5.

```

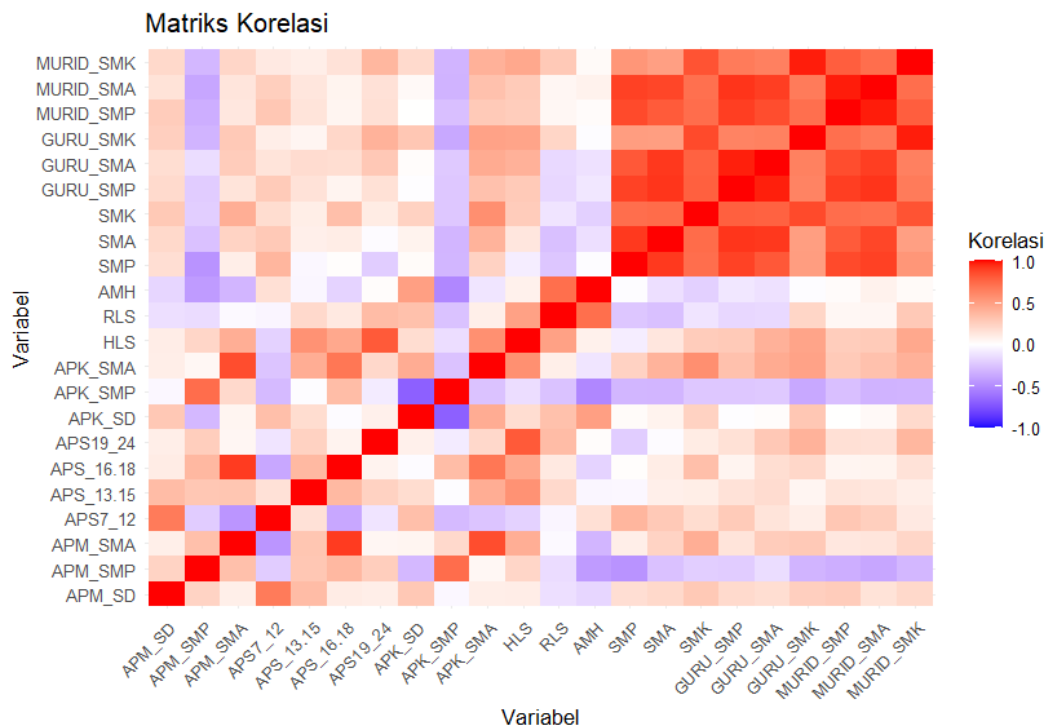
> #cek nilai kosong
> colSums(is.na(data))
KABUPATEN.KOTA      APM_SD      APM_SMP      APM_SMA      APS7_12      APS_13.15      APS_16.18
0                0                0                0                0                0                0
APS19_24      APK_SD      APK_SMP      APK_SMA      HLS      RLS      AMH
0                0                0                0                0                0                0
SMP      SMA      SMK      GURU_SMP      GURU_SMA      GURU_SMK      MURID_SMP
0                0                0                0                0                0                0
MURID_SMA      MURID_SMK
0                0
> #cek duplikat
> duplicated(data)
[1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

```

Gambar 5. Pemeriksaan Data

Berdasarkan Gambar 5, data yang dimiliki saat ini tidak terdapat nilai kosong serta tidak terdapat data yang duplikat. Hubungan antara variable dapat dilihat melalui

korelasi matriks, tampilan korelasi matriks untuk data yang digunakan dapat dilihat pada Gambar 6.

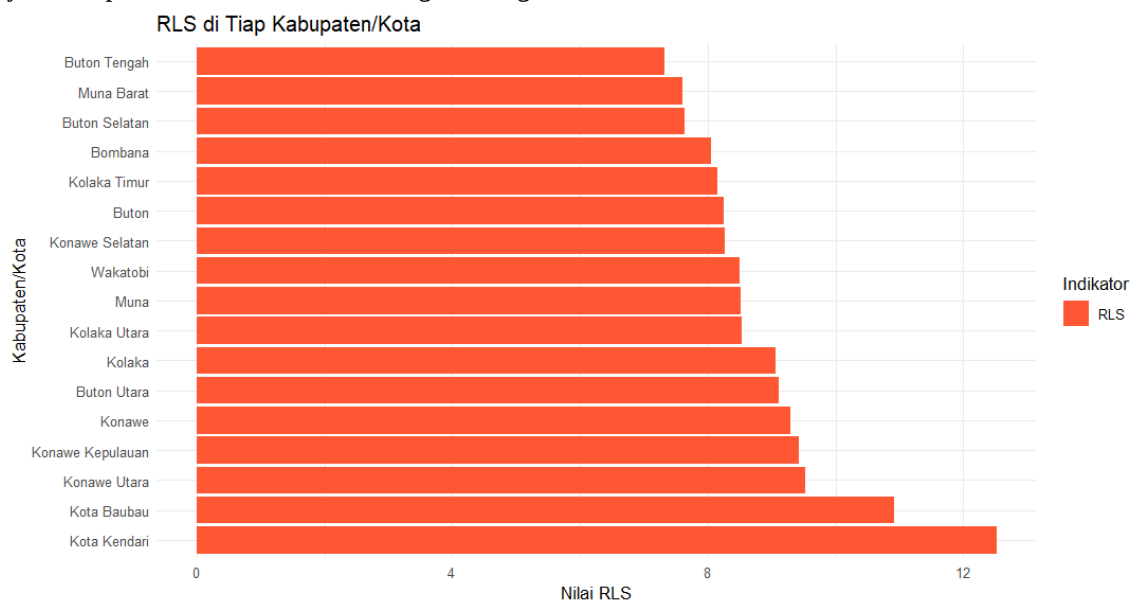


Gambar 6. Korelasi Matriks

Berdasarkan korelasi matriks pada Gambar 6, terlihat bahwa korelasi positif ditandai dengan warna merah yang semakin kuat, korelasi negatif ditandai dengan warna biru, dan korelasi lemah ditandai dengan warna putih. Berdasarkan Gambar 6, terlihat bahwa variabel RLS memiliki korelasi positif yang cukup tinggi dengan variabel HLS dan AMH, Dimana ketiga variabel tersebut sama-sama menggambarkan kualitas Pendidikan secara makro. Sedangkan variabel AMH dengan variabel MURID_SMA memiliki korelasi yang lemah, hal tersebut dapat disebabkan karena AMH mencakup seluruh populasi di wilayah kabupaten/kota tersebut, dibandingkan dengan

variabel MURID_SMA yang lebih berfokus pada jumlah siswa yang berada pada tingkat SMA, sehingga jumlah AMH pada suatu kabupaten/kota bisa jauh lebih tinggi dibandingkan jumlah siswa yang berada pada tingkat SMA.

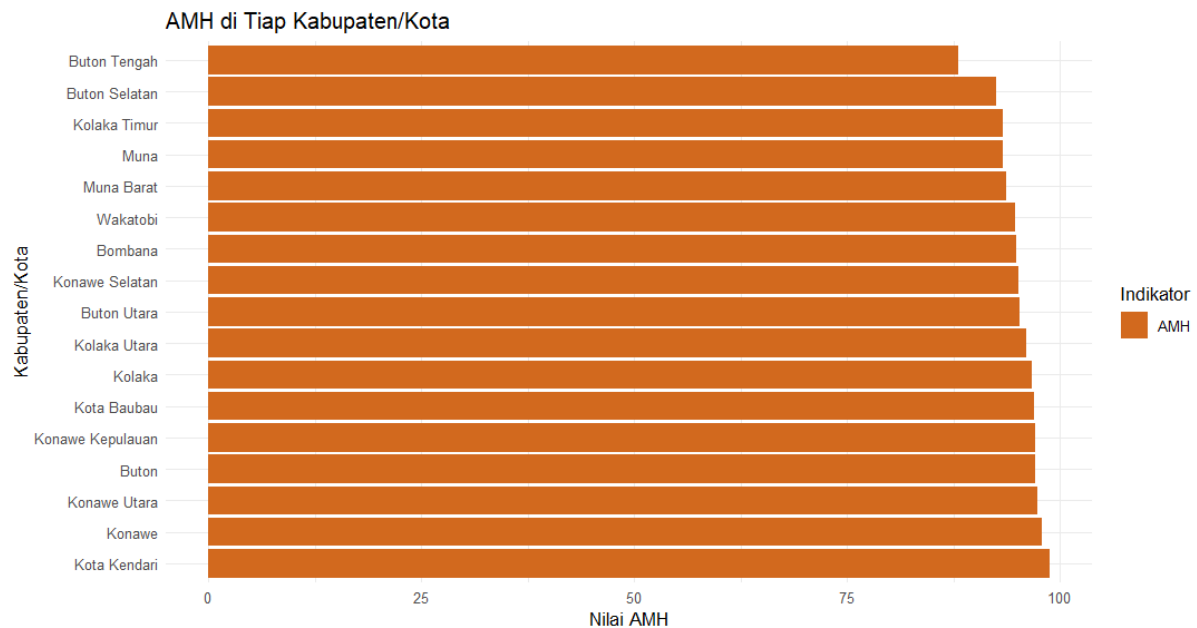
Berdasarkan penjelasan terkait Gambar 6 diketahui bahwa RLS, AMH, dan HLS memiliki korelasi positif yang cukup tinggi. Korelasi positif menandakan bahwa variabel-variabel tersebut berbanding lurus, yaitu semakin tinggi nilai variabel RLS maka semakin tinggi nilai variabel AMH dan HLS, begitu juga sebaliknya.



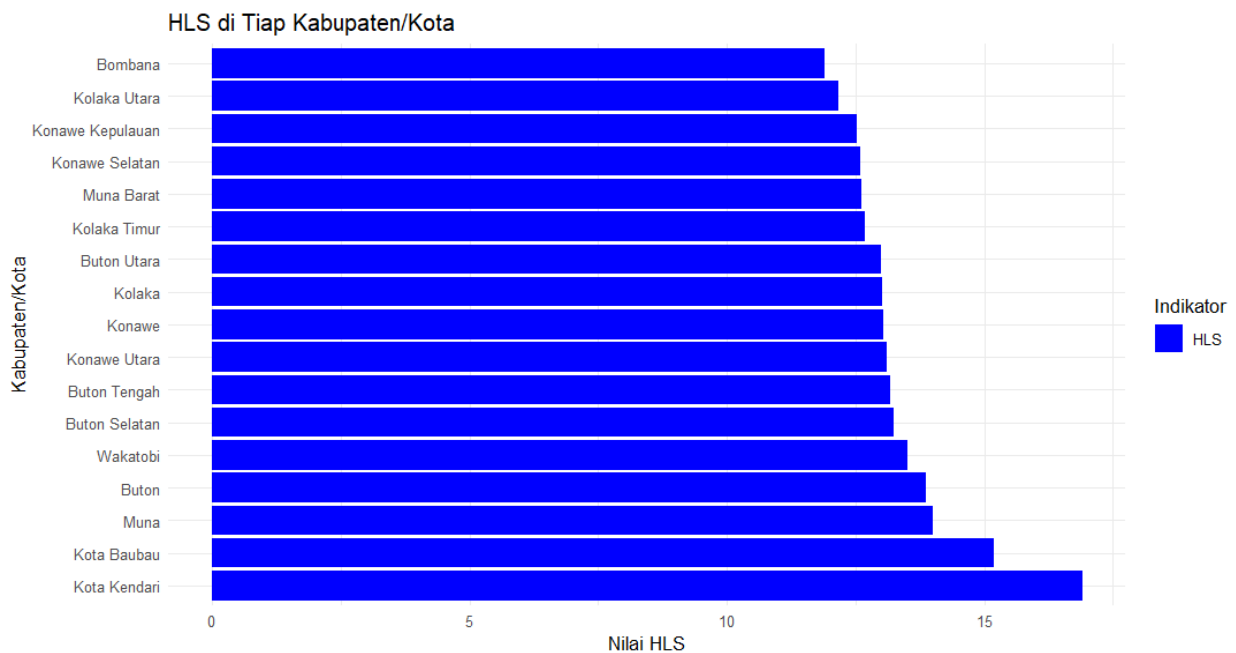
Gambar 7. Perbandingan RLS

Perbandingan variabel RLS di setiap kabupaten/kota dapat dilihat pada Gambar 7. Untuk melihat perbandingan variabel AMH di setiap kabupaten/kota dapat dilihat pada

Gambar 8. Untuk melihat perbandingan variabel HLS di setiap kabupaten/kota dapat dilihat pada Gambar 9.



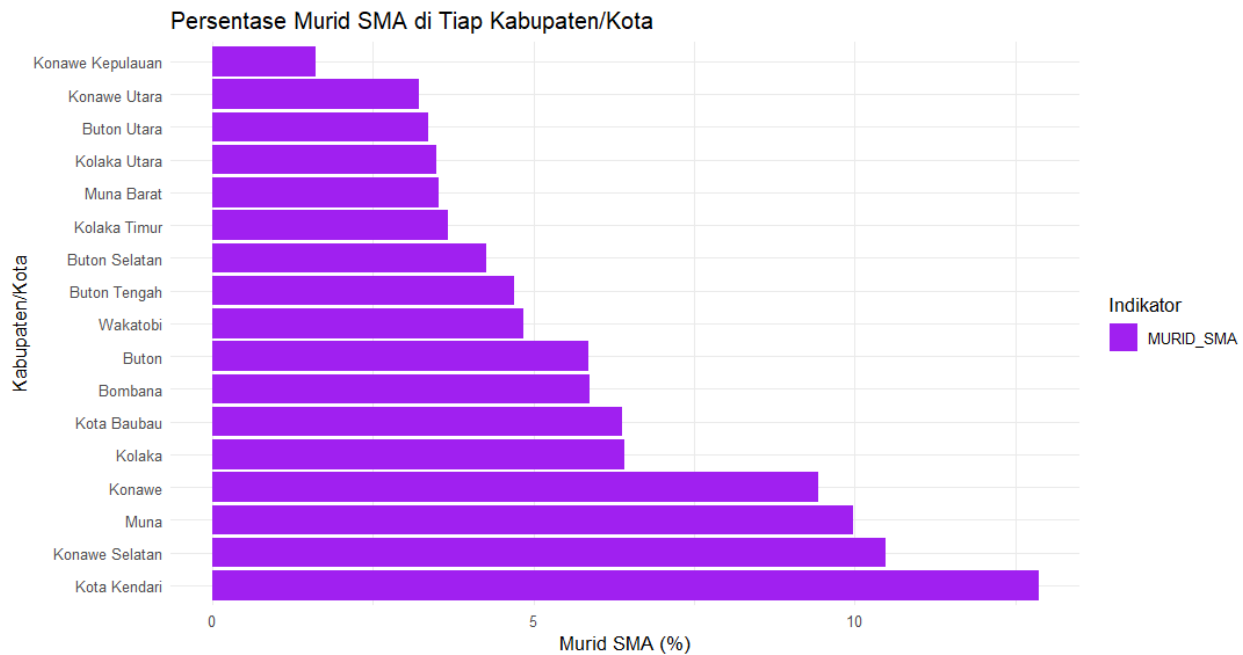
Gambar 8. Perbandingan AMH



Gambar 9. Perbandingan HLS

Korelasi antara variabel yang lemah dapat ditemukan antara variabel AMH dengan variabel MURID_SMA. Korelasi yang lemah menandakan bahwa perubahan pada variabel AMH tidak terlalu mempengaruhi variabel

MURID_SMA, begitu juga sebaliknya. Perbandingan variabel MURID_SMA di setiap kabupaten/kota dapat dilihat pada Gambar 10.



Gambar 11. Perbandingan Murid SMA

Tahap selanjutnya setelah dilakukan pembersihan data adalah dengan melakukan standarisasi data. Standarisasi data dapat dilihat pada Gambar 11.

Standarisasi dilakukan pada seluruh variabel yang

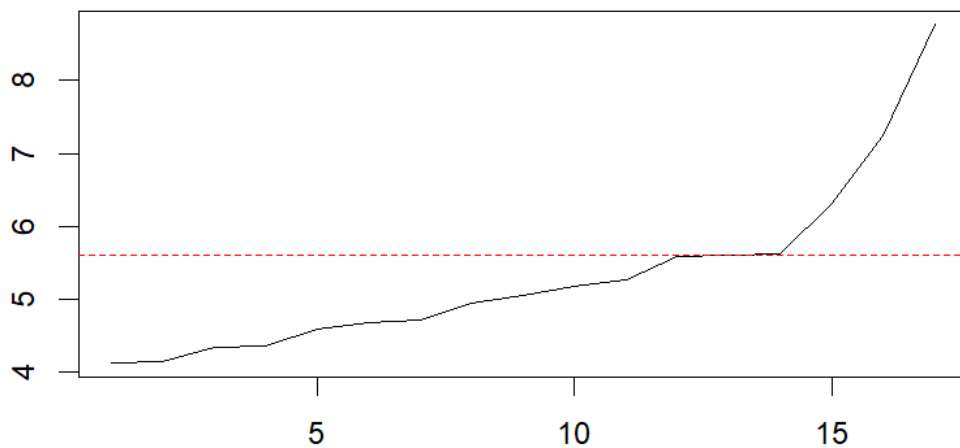
APM_SD	APM_SMP	APM_SMA	APS7_12	APS_13.15	APS_16.18	APS19_24	APK_SD	APK_SMP
0.6348527	-0.1389268	0.6633591	0.3905552	0.95064025	0.7668594	0.4277515	1.72835928	-1.11526280
0.7777996	0.4435267	0.9586038	0.7435724	0.48974125	0.7881004	0.2478152	0.44774999	-0.21483708
-1.6193109	-0.6744082	-0.1552736	-0.2068585	-0.33726193	-0.3031516	-0.1973317	0.04370749	-0.02309175
0.8382772	-1.5499673	-0.2729586	0.8929258	-0.01038321	-0.2429690	-0.4758419	1.89271556	-0.65533312
0.9647302	-0.6762871	-0.4226455	1.2052102	0.78393210	-0.4659986	-0.7645225	-1.09309006	0.07925880
0.8327792	-1.0370325	-0.4608415	0.8793482	-2.15143887	-1.0704796	-0.8505790	0.85864067	-1.04141494
APK_SMA	HLS	RLS	AMH	SMP	SMA	SMK	GURU_SMP	GURU_SMA
0.93692469	0.4558940	-0.4798874	0.73710961	0.2872917	0.5454562	0.04084132	0.21180984	0.5665163
0.91877156	0.5729412	-0.2695070	-0.73688323	1.8967782	1.9720340	2.47089973	2.39009256	1.9646062
-0.01198883	-0.2463893	0.3382587	1.03344760	1.1752843	1.1398636	0.21441692	1.00391265	0.9974030
0.11260764	-0.2547498	0.1512539	0.57931926	0.3982908	-0.4055956	0.73514372	0.06329057	-0.1806596
-0.60939178	-0.6142520	-0.4643036	-0.05184217	2.1187764	1.2587451	0.04084132	1.40326448	0.8277987
-0.51862614	-1.1994881	-0.6357247	-0.14035871	0.7312880	0.9021007	0.04084132	-0.10173085	-0.1531562
GURU_SMK	MURID_SMP	MURID_SMA	MURID_SMK					
-0.20186034	-0.2147731	-0.0059421616	-0.5679530					
1.68553385	1.4493906	1.3265168320	1.6768233					
-0.06704647	0.7308646	1.1509401054	0.1002526					
0.98695289	0.4772234	0.1734141032	0.8557441					
-0.13445341	1.8237778	1.4886933779	0.2353731					
-0.04866276	-0.1151018	-0.0007882459	-0.3554094					

Gambar 11. Data Setelah Standarisasi

Implementasi DBSCAN

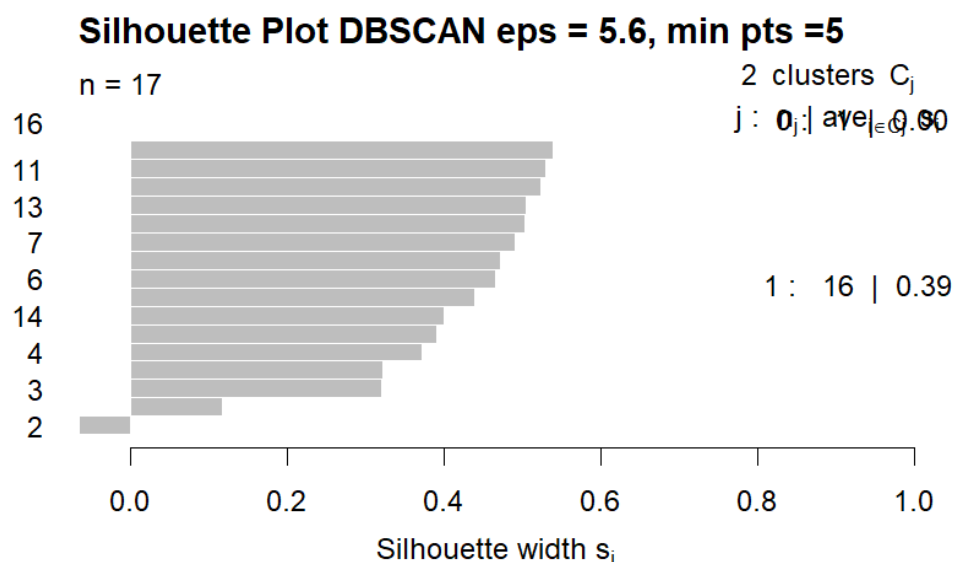
Proses pengelompokan diawali dengan menentukan nilai parameter yaitu epsilon dan MinPts. Penentuan nilai epsilon dilakukan dengan bantuan algoritma KNN untuk

melihat *knee point*. Hasil dari penerapan algoritma KNN untuk melihat *knee point* dapat dilihat pada Gambar 12

Gambar 12. Penentuan *knee point*

Berdasarkan Gambar 12, terlihat bahwa titik patah dari grafiknya terletak di titik 5,6 seperti yang terlihat pada garis putus putus warna merah. Berdasarkan hasil tersebut nilai epsilon yang digunakan untuk melakukan pengelompokan adalah 5,6.

Penentuan MinPts dilakukan dengan bantuan *silhouette score*. Penentuan MinPts dengan bantuan silhouette score dilakukan dengan cara melihat nilainya. Hasil dari penerapan silhouette score dapat dilihat pada Gambar 13.

Gambar 13. Hasil *Silhouette Score*

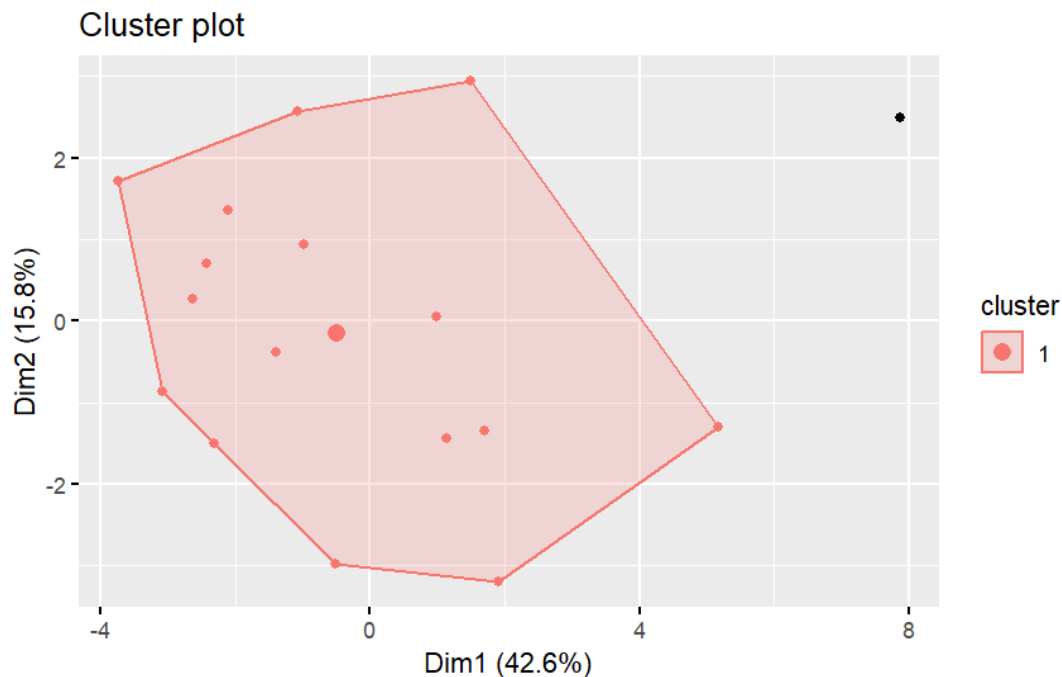
Berdasarkan Gambar 13, terlihat bahwa hasil *silhouette score* adalah 0,37 dengan menggunakan nilai MinPts 5. Selanjutnya dilakukan beberapa percobaan dengan nilai MinPts yang berbeda untuk mengetahui MinPts mana yang terbaik. Hasil dari percobaan yang telah dilakukan dapat dilihat pada Tabel 3.

Tabel 3. Percobaan MinPts

Minimum Points	Hasil
2	0,37
5	0,37
6	0,37
7	0,34
8	0,34
10	0,28

Berdasarkan hasil percobaan pada Table 3, terlihat bahwa nilai *shilouette score* yang tertinggi terdapat pada MinPts 2 sampai 6, sedangkan untuk MinPts bernilai 7, 8, dan 10 menghasilkan nilai yang semakin menurun. Pada penelitian ini nilai MinPts yang akan digunakan adalah 5.

Proses klaster dilakukan dengan menggunakan parameter yang ditentukan yaitu epsilon bernilai 5,6 dan MinPts bernilai 5. Setelah dilakukan pengelompokan menggunakan algoritma DBSCAN dihasilkan klaster dan *outlier*. Hasil dari *cluster* dan *outlier* dapat dilihat pada Gambar 14.



Gambar 14. Hasil Pembentukan Kluster

Berdasarkan hasil pengelompokan menggunakan algoritma DBSCAN pada gambar 9, terlihat bahwa tercipta 1 kluster dan juga 1 outlier, artinya terdapat 16 data yang menjadi 1 buah kluster dan terdapat sebuah data yang menjadi outlier.

Interpretasi data

Berdasarkan hasil yang diperoleh pada proses clustering, ditemukan sebuah kluster dan sebuah *outlier*. Adapun kabupaten/kota yang menjadi kluster dan data yang menjadi *outlier* dapat dilihat pada Tabel 4.

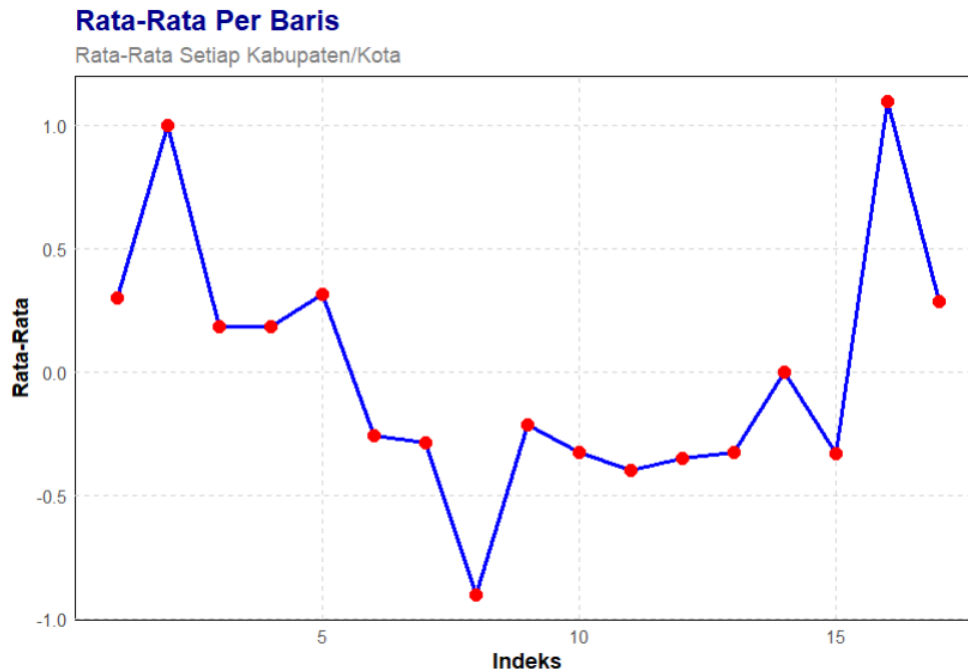
Tabel 4. Hasil Pembentukan Kluster

No.	Kluster 1	Outlier
1	Buton Utara	Kota Kendari
2	Bombana	
3	Konawe	
4	Konawe Selatan	
5	Kolaka Timur	
6	Muna	
7	Wakatobi	
8	Kolaka	
9	Buton	
10	Konawe Kepulauan	

11	Kota Baubau	
12	Konawe Selatan	
13	Muna Barat	
14	Buton Tengah	
15	Buton Selatan	
16	Kolaka Utara	

Berdasarkan Tabel 4, diketahui bahwa kabupaten/kota yang menjadi *outlier* adalah Kota Kendari, sedangkan 16 kabupaten/kota lainnya menjadi satu kluster. Hasil tersebut menandakan bahwa terdapat perbedaan persebaran data antara Kota Kendari dibandingkan dengan kabupaten/kota lainnya. Berdasarkan penjelasan sebelumnya yaitu DBSCAN menentukan kluster berdasarkan kerapatan datanya. Berdasarkan penjelasan tersebut, kota kendari memiliki kerapatan data yang lebih rendah dibanding data kabupaten/kota yang lain.

Kota Kendari menjadi *outlier* menandakan adanya perbedaan karakteristik dalam indikator pendidikan. Kota Kendari menjadi *outlier* dapat disebabkan karena nilai-nilai pada variabel yang lebih besar dibandingkan dengan wilayah yang lain. Nilai rata-rata variabel setiap kabupaten/kota dapat dilihat pada Gambar 15.



Gambar 15. Rata-rata Nilai Data Kabupaten/kota

Berdasarkan Gambar 15 terlihat bahwa nilai rata-rata Kota Kendari (indeks ke-16) lebih tinggi dibanding dengan kabupaten/kota yang lain. Hasil tersebut dapat memberikan gambaran terkait Kota Kendari yang menjadi *outlier*. Hal tersebut juga terlihat pada gambar 16, yaitu terdapat *outlier* yang posisinya lebih jauh dibandingkan dengan titik yang lain.

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

Penelitian ini berhasil mengelompokkan wilayah kabupaten/kota di Provinsi Sulawesi Tenggara ke dalam satu kluster utama serta satu outlier, dengan Kota Kendari sebagai pengecualian yang menunjukkan karakteristik pendidikan berbeda secara signifikan. Hal ini menunjukkan potensi ketimpangan distribusi kualitas pendidikan di provinsi tersebut. DBSCAN terbukti efektif dalam mengidentifikasi struktur data berdasarkan kerapatan, meskipun nilai Silhouette Score yang dihasilkan (0,37) mengindikasikan bahwa kualitas kluster masih dapat ditingkatkan. Untuk penelitian selanjutnya, disarankan mengkaji ulang variabel input dengan pendekatan reduksi dimensi atau penggabungan variabel berkorelasi tinggi, serta membandingkan hasil dengan algoritma lain seperti K-Means atau HDBSCAN.

B. Saran

Penelitian ini hanya menggunakan indikator pendidikan sebagai dasar pengelompokan. Untuk pengembangan ke depan, disarankan mempertimbangkan integrasi variabel tambahan seperti indikator sosial ekonomi, kualitas infrastruktur pendidikan, atau tingkat pengangguran. Selain itu, eksplorasi algoritma lain serta pendekatan evaluasi klusterisasi yang lebih komprehensif dapat meningkatkan validitas hasil analisis

DAFTAR PUSTAKA

- [1] S. Hanifah dan A. H. Primandari, "IMPLEMENTASI METODE K-MEANS CLUSTERING DALAM PENGELOMPOKAN KABUPATEN/KOTA DI PROVINSI NTB BERDASARKAN INDIKATOR PENDIDIKAN," *Emerging Statistics and Data Science Journal*, vol. 1, no. 3, 2023.
- [2] R. Sira Torsina Sihombing, "Pemerataan Pendidikan: Studi Kasus 34 Provinsi di Indonesia," *Parahyanan Economic Development Review (PEDR)*, vol. 1, no. 2, hlm. 143–151, 2022.
- [3] Badan Pusat Statistik, *statistik-pendidikan-2022*. 2022.
- [4] S. P. Simanjorang dan M. Yanti, "Pengelompokan Kabupaten/Kota di Sumatera Utara Menggunakan Algoritma Average Linkage dan K-Means Berdasarkan Indikator Pendidikan Riwayat Artikel," *Emerging Statistics and Data Science Journal*, vol. 1, no. 3, 2023.
- [5] D. Pri Indini, S. R. Siburian, dan D. Putro Utomo, "IMPLEMENTASI ALGORITMA DBSCAN UNTUK CLUSTERING SELEKSI PENENTUAN MAHASISWA YANG BERHAK MENERIMA BEASISWA YAYASAN," 2022.
- [6] D. A. Puspitasari, Y. Cahyana, S. Arum, dan P. Lestari, "Penerapan Algoritma Density Based Spastial Clustering Algorithm With Noise Untuk Pengelompokan Penyakit Pasien," vol. IV, no. 1, 2023.
- [7] R. Anggara, A. Rahman, dan S. Si, "Implementasi Algoritma DBSCAN Dalam Mengelompokan Data Pasien Terdiagnosa Penyakit Ginjal Kronis (PGK)," *Jurnal AlgoritmeCCS*, vol. x, No.x, no. 1, hlm. 114–123, 2022, doi: 10.35957/algoritme.xxxx.
- [8] A. Kristianto, "Implementasi DBSCAN dalam Clustering Data Minat Mahasiswa Setelah Pandemi Covid19," 2022.
- [9] O. Putri Yuli Utami, S. Agustian Hudjimartsu, T. Aurilia Viona, H. Sharfina, J. A. Yani No, dan K. Barat, "Parameter Algoritma DBSCAN untuk Mendeteksi Titik Panas Kebakaran Hutan dan Lahan," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 10, no. 4, 2023, [Daring]. Tersedia pada: <https://hotspot.brin.go.id/>.
- [10] M. R. Nahjan, N. Heryana, dan A. Voutama, "IMPLEMENTASI RAPIDMINER DENGAN METODE CLUSTERING K-MEANS UNTUK ANALISA PENJUALAN PADA TOKO OJ CELL," 2023.
- [11] A. Tangkelayuk dan E. Mailoa, "Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes Dan Decision Tree," vol. 9, no. 2, hlm. 1109–1119, 2022, [Daring]. Tersedia pada: <http://jurnal.mdp.ac.id>
- [12] I. Bagus, F. Arafat, M. Amin Hariyadi, I. B. Santoso, C. Crysdian, dan P. Korespondensi, "CLUSTERING GEMPABUMI DI WILAYAH REGIONAL VII MENGGUNAKAN PENDEKATAN DBSCAN," vol. 10, no. 4, 2023, doi: 10.25126/jtiik.2023106918.
- [13] A. Khalif dkk., "Klasterisasi Tingkat Kemiskinan di Indonesia Menggunakan Algoritma K-Means," 2023.