

Implementasi Metode *K-Medoids Clustering* untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan

Astrika Praokta Fialine¹, Daffa Alya Alodia^{2*}, Desy Endriani³, Edy Widodo⁴

^{1,2,3,4} Prodi Statistika FMIPA Universitas Islam Indonesia
18611087@students.uii.ac.id¹, 18611131@students.uii.ac.id^{2*}, 18611007@students.uii.ac.id³,
edywidodo@uui.ac.id⁴

Abstract

One of the goals of the Sustainable Development Goals (SDGs) is to ensure inclusive and equitable quality education and to support lifelong learning opportunities for all. To reach this purpose, it can be done by running a compulsory education program to upgrade the quality of education in Indonesia. The distribution of compulsory education programs can be seen through clustering analysis on several education indicators. The data used in this study was obtained from the publication of Education Statistics for the 2019/2020 academic year by the Central Bureau of Statistics with 8 indicators of SMA/MA equivalent education by province in Indonesia, namely Gross Enrollment Rate (APK), Net Enrollment Rate (APM), School Participation Rate (APS), percentage of eligible teachers, Average Length of School (RLS), Continuation Rate (AM), school completion rate (SMA), and dropout rates. The method used is *K-Medoids Clustering*, which is a clustering method that uses the partition grouping method to group a set of n objects into several k clusters. From the results of the analysis, obtained three clusters, the first cluster contains 6 provinces with higher education indicators, the second cluster contains 13 provinces with moderate education indicators, and the third cluster contains 15 provinces with low education indicators.

Keyword: *Cluster, Education, K-Medoids, SDGs.*

Abstrak

Salah satu tujuan dalam *Sustainable Development Goals* (SDGs) adalah memastikan pendidikan yang inklusif dan berkualitas setara, serta mendukung kesempatan belajar seumur hidup bagi semua. Untuk mencapai tujuan tersebut, hal yang dapat dilakukan dengan menjalankan program wajib belajar untuk meningkatkan kualitas pendidikan di Indonesia. Pemerataan program wajib belajar ini dapat ditinjau melalui analisis *clustering* pada beberapa indikator pendidikan di Indonesia. Data yang digunakan diperoleh dari publikasi Statistik Pendidikan tahun ajaran 2019/2020 oleh Badan Pusat Statistik dengan 8 variabel indikator pendidikan SMA/MA sederajat menurut provinsi di Indonesia, yaitu Angka Partisipasi Kasar (APK), Angka Partisipasi Murni (APM), Angka Partisipasi Sekolah (APS), persentase guru layak, Rata-Rata Lama Sekolah (RLS), Angka Melanjutkan (AM), tingkat menyelesaikan sekolah (SMA), dan angka putus sekolah. Metode yang digunakan adalah *K-Medoids Clustering*. *K-Medoids* merupakan salah satu metode *clustering* yang menggunakan teknik pengelompokan partisi dalam mengelompokkan kumpulan n objek menjadi ke dalam k cluster. Dari hasil analisis, diperoleh tiga cluster. Cluster pertama terdapat 6 provinsi dengan indikator pendidikan tinggi, cluster kedua terdapat 13 provinsi dengan indikator pendidikan sedang, dan cluster ketiga terdapat 15 provinsi dengan indikator pendidikan rendah.

Kata Kunci: *Cluster, K-Medoids, Pendidikan, SDGs*

1. Pendahuluan

Pendidikan memiliki kedudukan penting untuk merancang sumber daya manusia dalam pembangunan suatu negara (Aprilia, Srinade, & Sari, 2016). Kualitas negara yang baik dapat dilihat dari semakin baiknya kualitas pendidikan yang diselenggarakan. Salah satu tujuan dalam *Sustainable Development Goals* (SDGs) adalah “Memastikan pendidikan yang inklusif dan berkualitas setara, serta mendukung kesempatan belajar seumur hidup bagi semua” (Soraya & Suryadi, 2019). Untuk menggapai hal tersebut, dapat diupayakan dengan menjalankan program wajib belajar 12 tahun untuk memajukan pendidikan di Indonesia yang mencakup pendidikan dasar dan menengah. Beberapa indikator pendidikan yang dapat digunakan adalah Angka Partisipasi Kasar (APK), Angka Partisipasi Murni (APM), Angka Partisipasi Sekolah (APS), persentase guru layak, Rata-Rata Lama Sekolah (RLS), Angka Melanjutkan (AM), tingkat menyelesaikan sekolah SMA (TMS), dan Angka Putus Sekolah (APtS) (Badan Pusat Statistik, 2020).

Berdasarkan data dari Survei Sosial Ekonomi pada bulan Maret 2020 dalam Statistik Pendidikan 2020 pada jenjang SMA/MA sederajat, Indonesia belum mencapai kriteria Tuntas Paripurna dalam penuntasan wajib belajar karena nilai APKnya belum mencapai 95%. Sedangkan dilihat dari Angka Putus Sekolahnya masih terdapat 1.13% penduduk yang putus sekolah dalam menempuh pendidikan jenjang SMA/MA sederajat.

Berdasarkan kondisi tersebut penting dilakukan pengelompokan provinsi di Indonesia berdasarkan indikator pendidikan dengan menggunakan metode *clustering*. Penelitian sebelumnya yang telah mengangkat tema serupa yaitu penelitian Hanna Silia Karti dan Irhamah (2013) yang didapatkan APM dan APK terendah berada di Kabupaten Sampang, yaitu sebesar 26,44 persen dan 35,26 persen. Hasil pengelompokan paling optimum sebanyak 2 kelompok dengan metode *c-means* dan *fuzzy c-means*. Namun, pada penelitian tersebut belum terdapat visualisasi dari hasil kelompok yang terbentuk.

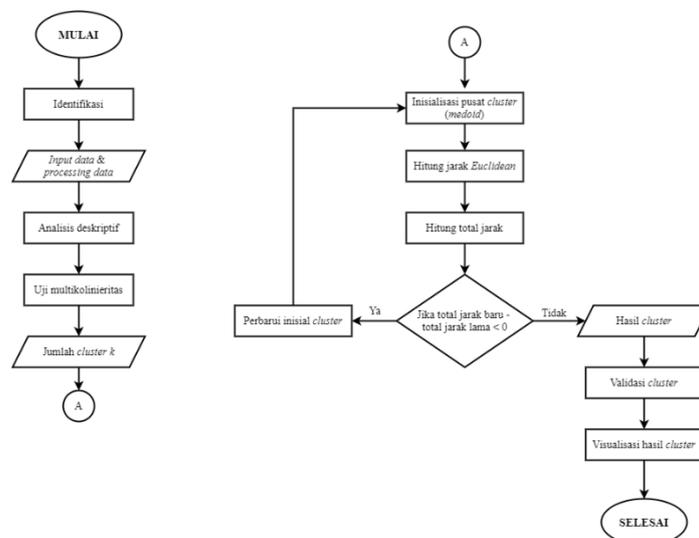
Menurut pemaparan di atas, peneliti tertarik untuk melakukan penelitian dengan melakukan pengelompokan dan pemetaan provinsi di Indonesia berdasarkan indikator pendidikan untuk mengetahui kelompok-kelompok provinsi mana yang membutuhkan perbaikan masalah dalam pendidikan. Data indikator pendidikan jenjang SMA/MA sederajat di Indonesia pada tahun ajaran 2019/2020 diperoleh dari Survei Sosial Ekonomi pada bulan Maret 2020 dalam Statistik Pendidikan 2020. Data tersebut terdiri dari 8 variabel yaitu APK, APM, APS, persentase guru layak, RLS, AM, TMS (SMA), dan APtS.

Pengelompokan dilakukan dengan menggunakan metode *K-Medoids Clustering* karena pada data indikator pendidikan di Indonesia pada tahun ajaran 2019/2020 terdapat *outlier* dan juga belum banyak penelitian mengenai tema serupa dengan menggunakan metode *K-Medoids clustering*. *K-Medoids* adalah teknik pengelompokan partisi yang mengelompokkan kumpulan n objek ke dalam k cluster dengan memanfaatkan objek pada kumpulan objek guna mewakili sebuah *cluster* yang dinamakan *medoid* (Purba, Saifullah, & Dewi, 2019). Sehingga dengan implementasi metode tersebut akan diperoleh hasil pengelompokan provinsi berdasarkan indikator pendidikan di Indonesia yang dapat dijadikan sebagai dasar dalam perencanaan pembangunan pendidikan di Indonesia yang dapat disesuaikan dengan karakteristik setiap provinsi pada tiap *cluster*.

2. Metode Penelitian

Penelitian ini dilakukan dengan menggunakan *K-Medoids Clustering* untuk mengelompokkan provinsi di Indonesia berdasarkan indikator pendidikan tahun ajaran 2019/2020. Data yang digunakan diperoleh dari dari Survei Sosial Ekonomi pada bulan Maret 2020 dalam Statistik Pendidikan tahun ajaran 2019/2020 yang diakses pada tanggal 26 Maret 2021, dengan menggunakan 8 variabel yaitu APK, APM, APS, persentase guru layak, RLS, AM, TMS, dan APtS.

Metode penelitian yang digunakan adalah analisis deskriptif yang digunakan untuk melihat data indikator pendidikan di Indonesia tahun ajaran 2019/2020 secara umum, serta analisis *K-Medoids* yang digunakan untuk mengelompokkan provinsi di Indonesia berdasarkan indikator pendidikan tahun ajaran 2019/2020 dengan menggunakan perangkat lunak *Rstudio* dan memvisualisasikan hasil *cluster* dengan perangkat lunak *QGIS*. **Gambar 1** merupakan diagram alir langkah-langkah penelitian yang dilakukan.



Gambar 1. Diagram Alir Penelitian

Adapun tahapan penelitian yang dilakukan berdasarkan **Gambar 1** adalah sebagai berikut:

1. Mengidentifikasi dan merumuskan masalah yang akan diangkat dalam penelitian.
2. Menginputkan data pendidikan di Indonesia tahun ajaran 2019/2020 dan standarisasi data karena satuan dari variabel yang digunakan berbeda-beda.
3. Melakukan analisis deskriptif untuk mengetahui nilai *min*, *max*, dan *mean*.
4. Melakukan uji multikolinieritas untuk melihat apakah terdapat multikolinearitas pada data yang akan di teliti, dimana multikolinearitas terjadi jika antar variabel bebas nilai korelasinya > 0.95 dengan hipotesis:

H_0 : Tidak terdapat multikolinieritas

H_1 : Terdapat multikolinieritas

(Hocking, 1997).

5. Menentukan jumlah *cluster k* dengan metode *Silhouette*.
6. Inisialisasi pusat *cluster (medoid)* sebanyak *k*.
7. Mengalokasikan setiap objek ke *cluster* paling dekat dengan jarak *Euclidian* menggunakan persamaan berikut (Ningrum et al., 2021):

$$d_{ij} = \sqrt{\sum_{a=1}^p (x_{ia} - x_{ja})^2} \quad (1)$$

dengan :

d_{ij} = jarak objek i dan j

x_{ia} = nilai objek i saat variabel ke a

x_{ja} = nilai objek j saat variabel ke a

p = total variabel yang diamati

8. Pilih sembarang objek pada setiap *cluster* menjadi kandidat *medoid* baru (*non-medoid*).
9. Menghitung jarak antar objek yang terletak di setiap *cluster* dan kandidat *medoid* baru (*non-medoid*).
10. Menghitung total simpangan (S), yaitu dengan pengurangan total jarak baru dengan total jarak lama. Apabila $S < 0$, maka ubah objek menggunakan data *cluster* guna membangun gerombolan *k* objek baru menjadi *medoid*.
11. Lakukan langkah 8-10 kembali sampai tidak ada pergantian *medoid*, sehingga diperoleh *cluster* dan anggota setiap *cluster* (Pulungan et al., 2019).
12. Melakukan validasi dari nilai variansi dalam *cluster* (V_w) dan variansi antar *cluster* (V_b).

Berikut merupakan rumus untuk V_w :

$$V_w = \frac{1}{N-k} \sum_{i=1}^k (n_i - 1) \times V_i^2 \quad (2)$$

dengan :

- V_w = variansi dalam *cluster*
- k = banyaknya *cluster*
- n_i = banyaknya data pada suatu *cluster*
- N = data ke – i pada suatu *cluster*
- V_i^2 = variansi pada suatu *cluster-i*

Berikut merupakan rumus untuk V_b :

$$V_b = \frac{1}{k-1} \sum_{i=1}^k n_i (\bar{x}_i - \bar{x})^2 \quad (3)$$

dengan :

- V_b = variansi antar *cluster*
- k = banyaknya *cluster*
- n_i = banyaknya data pada suatu *cluster*
- \bar{x}_i = rata-rata data ke– i di suatu *cluster*
- \bar{x} = rata-rata seluruh data

Berikut merupakan rumus untuk mencari variansi seluruh *cluster*:

$$V = \frac{V_w}{V_b} \quad (4)$$

dengan:

- V = variansi
- V_w = variansi dalam *cluster*
- V_b = variansi antar *cluster*

(Dini & Fauzan, 2020)

13. Interpretasi hasil cluster, memvisualisasikan hasil *cluster*, dan penarikan kesimpulan.

3. Hasil Penelitian dan Pembahasan

3.1. Analisis Deskriptif

Analisis deskriptif dapat dimanfaatkan untuk menggambarkan karakteristik setiap provinsi di Indonesia berdasarkan indikator pendidikan tahun ajaran 2019/2020. Analisis deskriptif dari setiap variabel indikator pendidikan di Indonesia dapat dilihat pada **Tabel 1** berikut.

Tabel 1. Analisis Deskriptif

	APK (%)	APM (%)	APS (%)	Guru Layak (%)	RLS (Tahun)	AM (%)	TMS (%)	APtS (%)
<i>Min</i>	73.35	44.73	64.83	84.2	6.69	79.25	30.92	0.18
<i>Mean</i>	87.08	62.32	74.97	90	8.65	90.09	64.58	1.23
<i>Max</i>	98.31	73.45	88.95	97.86	11.1	96.9	87.99	3.53

Berdasarkan **Tabel 1**, terlihat bahwa secara keseluruhan untuk variabel AM memiliki rata-rata terbesar jika dibandingkan yang lainnya yaitu sebesar 90,09%. Provinsi yang memiliki nilai rata-rata dibawah 90,09% pada AM adalah Provinsi Jawa Timur, Provinsi Sumatera Selatan, Provinsi Jambi, Provinsi Lampung, Provinsi Sulawesi Selatan, Provinsi Jawa Tengah, Provinsi Aceh, Provinsi Banten, Provinsi Bali, Provinsi Kalimantan Tengah, Provinsi Jawa Barat, dan Provinsi Papua. Namun jika dilihat berdasarkan nilai maksimum yang paling tinggi adalah indikator APK dengan nilai persentase sebesar 98,31% berada pada provinsi Kalimantan Utara. Sedangkan untuk indikator pendidikan paling rendah adalah APtS dengan persentase 0.18% berada pada provinsi Yogyakarta.

3.2. Uji Multikolinieritas

Uji multikolinieritas dilakukan menggunakan nilai korelasi antar variabel yang dapat dilihat pada **Tabel 2**.

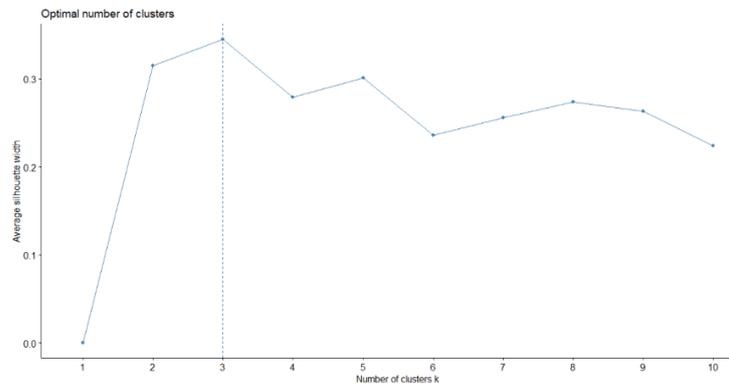
Tabel 2. Korelasi Antar Variabel

	APK	APM	APS	Guru Layak	RLS	AM	TMS	APtS
APK	1.00	0.60	0.66	0.27	0.17	0.60	0.25	-0.21
APM		1.00	0.87	-0.08	0.62	0.56	0.77	-0.30
APS			1.00	-0.13	0.50	0.53	0.63	-0.45
Guru Layak				1.00	-0.10	0.24	-0.05	0.15
RLS					1.00	0.35	0.84	-0.01
AM						1.00	0.43	0.09
TMS							1.00	-0.12
APtS								1.00

Uji multikolinieritas untuk melihat apakah data terjadi multikolinieritas atau tidak dimana H_0 : tidak multikolinieritas dan H_1 : multikolinieritas, dengan tingkat signifikansi 5% dan daerah kritis nilai korelasi $pearson > 0,95$ maka tolak H_0 diperoleh keputusan gagal tolak H_0 karena nilai korelasi $pearson < 0,95$. Sehingga dengan menggunakan tingkat kepercayaan 95% didapatkan bahwa tidak terjadi multikolinieritas maka asumsi multikolinieritas terpenuhi.

3.3. Penentuan Jumlah Cluster

Pada penelitian ini menggunakan metode *silhouette coefficient* untuk menentukan jumlah *cluster* (k). Metode *silhouette coefficient* dapat digunakan untuk melihat tingkat baik buruknya suatu *cluster*, seberapa baik suatu objek diletakkan didalam suatu *cluster* (Wira et al., 2019). Hasil dari metode *silhouette* dapat dilihat pada **Gambar 2**.



Gambar 2. Penentuan jumlah *cluster*

Berdasarkan **Gambar 2** didapatkan hasil k optimal pada saat $k = 3$, hal ini dapat diketahui dengan garis vertikal putus-putus yang berada pada sumbu x saat $k = 3$.

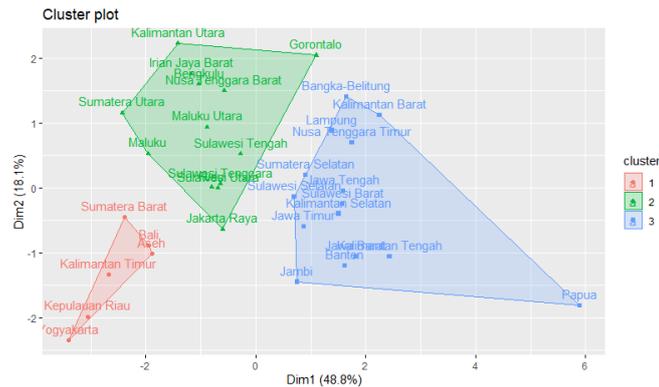
3.4. Analisis *K-Medoids Clustering*

K-Medoids adalah teknik pengelompokan partisi untuk mengelompokkan kumpulan n objek ke dalam k *cluster* dengan memanfaatkan objek pada kumpulan objek guna mewakili sebuah *cluster* yang dinamakan *medoid* (Purba, Saifullah, & Dewi, 2019). Dengan menggunakan k optimal sebanyak 3, maka didapatkan hasil pengelompokan dari data yang sudah distandarisasi dan diperoleh hasil pusat *medoids* seperti pada **Tabel 3**.

Tabel 1. Pusat *Medoids*

Pusat <i>Medois</i>	<i>Cluster</i>	APK	APM	APS	Guru Layak	RLS	AM	TMS	APtS
Kepulauan Riau	1	0.07	1.81	1.62	-1.63	1.58	0.48	1.39	-0.69
Maluku Utara	2	1.04	0.31	0.31	0.97	0.42	0.04	0.19	0.06
Sulawesi Barat	3	-0.45	-0.69	-0.86	-0.88	-0.82	-0.09	-0.79	0.33

Berdasarkan tabel tersebut diperoleh pusat *medoids* pada *cluster* 1 yaitu pada Provinsi Kepulauan Riau, *cluster* 2 pada Provinsi Maluku Utara, dan *cluster* 3 terletak pada Provinsi Sulawesi Barat. Selanjutnya dari hasil pengelompokan tersebut dibuat visualisasi berupa plot *cluster* yang terdiri dari tiga *cluster* seperti pada **Gambar 3**.



Gambar 3. Output K-Medoids Cluster

Gambar 3 menunjukkan bahwa terdapat tiga *cluster* yang didapatkan berdasarkan hasil perhitungan similaritas antar indikator pendidikan di Indonesia. Plot hasil *cluster* terdapat 3 warna, yaitu warna merah menunjukkan hasil *cluster* 1, warna hijau menunjukkan hasil *cluster* 2, dan warna biru menunjukkan hasil *cluster* 3. Setiap warna memiliki ciri khas tersendiri yang berbeda-beda. **Tabel 4** merupakan tabel hasil *cluster* indikator pendidikan.

Tabel 2. Hasil Cluster K-Medoids

Cluster	Jumlah	Anggota Cluster
1	6	Aceh, Kalimantan Timur, Kepulauan Riau, D.I Yogyakarta, Bali, Sumatera Barat
2	13	Sumatra Utara, Riau, Bengkulu, D.K.I Jakarta, Nusa Tenggara Barat, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Maluku, Maluku Utara, Irian Jaya Barat
3	15	Jambi, Sumatera Selatan, Lampung, Bangka-Belitung, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Selatan, Sulawesi Barat, Papua

Dari *cluster* menggunakan algoritma K-Medoids didapatkan hasil bahwa pada *cluster* 1 terdapat 6 provinsi, *cluster* 2 terdapat 13 provinsi, dan *cluster* 3 terdapat 15 provinsi di Indonesia yang memiliki similaritas atau kemiripan karakteristik yang tinggi. Peneliti melakukan segmentasi terhadap perhitungan rata-rata nilai tiap *cluster* dan didapatkan hasil seperti pada **Tabel 5** berikut.

Tabel 3. Profilisasi Cluster K-Medoids

Variabel	Cluster 1	Cluster 2	Cluster 3
APK (%)	90.3	90.7	82.7

Variabel	Cluster 1	Cluster 2	Cluster 3
APM (%)	71.1	64	57.4
APS (%)	84.2	76.6	69.9
Guru Layak (%)	86.4	92.8	89.1
RLS (tahun)	9.45	8.98	8.04
AM (%)	91.1	92	88.1
TMS (%)	75.1	67	58.3
APtS (%)	0.567	1.25	1.49

Berdasarkan **Tabel 5**, warna hijau menunjukkan indikator pendidikan dengan kategori tinggi yang berarti indikator pendidikan pada daerah tersebut sudah baik, warna kuning menunjukkan indikator pendidikan dengan kategori sedang yang berarti indikator pendidikan pada daerah tersebut sudah cukup baik, dan warna merah menunjukkan hasil indikator pendidikan yang rendah dengan artian indikator pendidikan pada daerah tersebut masih kurang. Terdapat tiga kelompok dari seluruh provinsi di Indonesia dengan karakteristik sebagai berikut.

- 1) *Cluster 1* yang beranggotakan 6 provinsi memiliki karakteristik indikator pendidikan yang tinggi pada variabel APM, APS, RLS, TMS, dan APtS dibandingkan dengan *cluster* lainnya, kemudian pada variabel APK dan AM lebih tinggi dari *cluster 3* dan lebih rendah dari *cluster 1*, sedangkan pada persentase guru layak lebih rendah dari *cluster* lainnya.
- 2) *Cluster 2* yang beranggotakan 13 provinsi memiliki karakteristik indikator pendidikan yang relatif sedang pada APM, APS, RLS, TMS, dan APtS yang lebih tinggi dari *cluster 3* dan lebih rendah dari *cluster 1*, kemudian memiliki kategori tinggi untuk variabel persentase guru layak, AM, dan APK jika dibandingkan dengan *cluster* lainnya.
- 3) *Cluster 3* yang beranggotakan 15 provinsi memiliki indikator pendidikan yang relatif rendah jika dibandingkan dengan *cluster* lainnya, dan kategori sedang untuk variabel guru layak yang lebih tinggi dari *cluster 1* dan lebih rendah dari *cluster 2*.

3.5. Validasi Cluster

Validasi *cluster* dilakukan dengan perhitungan nilai variansi *cluster*. Semakin kecil nilai variansi antar anggota maka akan diperoleh hasil yang semakin baik, dan semakin besar nilai variansi antar *cluster* maka hasil yang diperoleh akan semakin baik (Alfina, Santosa, & Barakbah, 2012). Hasil dari perhitungan variansi *cluster* dapat dilihat pada **Tabel 6**.

Tabel 4. Variansi Cluster

	Vw	Vb	Vw/Vb	Var K-Medoids
APK	0.6113	7.0253	0.0870	0.0952

	Vw	Vb	Vw/Vb	Var K-Medoids
APM	0.3305	11.3776	0.0290	
APS	0.2148	13.1710	0.0163	
Guru Layak	0.4507	9.5147	0.0474	
RLS	0.6613	6.2503	0.1058	
AM	0.7329	5.1399	0.1426	
TMS	0.6470	6.4711	0.1000	
APtS	0.8339	3.5751	0.2332	

Dari **Tabel 6** menunjukkan bahwa hasil nilai variansi *cluster* faktor pendidikan di Indonesia tahun ajaran 2019/2020 dengan menggunakan metode *K-Medoids* yaitu sebesar 0.0952. Hal ini dapat dikatakan bahwa *cluster* memiliki variansi yang baik karena sudah mendekati angka 0, karena semakin kecil variansi *cluster* maka semakin baik.

3.6. Visualisasi dengan QGIS

Visualisasi hasil *cluster* dilihat pada **Gambar 4** berikut.



Gambar 4. Visualisasi *Clustering*

Berdasarkan **Gambar 4**, dapat dilihat visualisasi hasil analisis *cluster* pada masing-masing provinsi di Indonesia dengan warna yang berbeda. Perbedaan warna tersebut menunjukkan karakteristik dari setiap daerah dimana untuk warna hijau menunjukkan bahwa daerah tersebut berada pada *cluster* 1 dengan karakteristik indikator pendidikan kategori tinggi, warna kuning menunjukkan bahwa daerah tersebut berada pada *cluster* 2 dengan karakteristik indikator pendidikan kategori sedang, dan warna merah menunjukkan bahwa daerah tersebut berada pada *cluster* 3 dengan karakteristik indikator pendidikan kategori tinggi.

Pada *cluster* 1 terdapat 6 anggota provinsi dimana provinsi-provinsi tersebut merupakan wilayah dengan indikator pendidikan dengan kategori tinggi hampir untuk setiap variabel kecuali pada variabel guru layak dengan karakteristik rendah, variabel APK dan AM dengan karakteristik sedang. Provinsi Kalimantan Timur merupakan provinsi dengan jumlah

persentase guru layak paling rendah jika dibandingkan dengan provinsi lain yaitu 84.2 persen sehingga inilah yang menyebabkan variabel guru layak pada *cluster* 1 memiliki karakteristik dengan kategori rendah. Akan tetapi jika dilihat dari semua variabel indikator pendidikan, *cluster* 1 dapat dikatakan daerah yang sudah memiliki kondisi pendidikan yang layak. *Cluster* 2 beranggotakan 13 provinsi yang merupakan wilayah dengan indikator pendidikan dengan kategori sedang. Namun, jumlah persentase guru layak pada *cluster* 2 termasuk kategori tinggi dengan rata-rata sebesar 92.8 persen. *Cluster* 3 beranggotakan 15 provinsi yang merupakan wilayah dengan indikator pendidikan dengan kategori rendah dimana masih terdapat variabel guru layak yang termasuk kategori sedang dengan rata-rata sebesar 89.1 persen.

4. Kesimpulan dan Saran

4.1. Kesimpulan

Berdasarkan hasil analisis data indikator Pendidikan di Indonesia tahun ajaran 2019/2020 diperoleh kesimpulan bahwa gambaran umum provinsi dengan nilai persentase indikator pendidikan paling tinggi adalah AM dengan rata-rata 90,09%. Namun jika dilihat dari nilai maksimum yang paling tinggi adalah indikator APK dengan nilai persentase sebesar 98,31% berada pada provinsi Kalimantan Utara. Sedangkan untuk indikator pendidikan paling rendah adalah APtS dengan persentase 0.18% berada pada provinsi Yogyakarta.

Berdasarkan hasil dari metode *silhouette*, diperoleh jumlah *cluster* optimum yang terbentuk adalah 3 yang memiliki similaritas antar indikator pendidikan di Indonesia, sehingga dengan analisis *K-Medoids clustering* diperoleh hasil 3 *cluster* yakni cluster tinggi, sedang, dan rendah. *Cluster* 1 terdapat 6 provinsi dengan karakteristik indikator pendidikan kategori tinggi, *cluster* 2 yang beranggotakan 13 provinsi dengan karakteristik indikator pendidikan yang relatif sedang, dan *cluster* 3 terdiri dari 15 provinsi memiliki indikator pendidikan yang relatif rendah.

4.2. Saran

Dari hasil penelitian, peneliti berharap agar pemerintah dapat memaksimalkan sistem pendidikan di Indonesia dengan memaksimalkan pemerataan program wajib belajar dengan memperhatikan daerah yang memiliki indikator pendidikan yang masih rendah. Hal tersebut dapat dilakukan dengan meningkatkan pembangunan infrastruktur pendidikan dan telekomunikasi berupa fasilitas serta meningkatkan kualitas pendidik melalui penetapan standar kualifikasi minimal sarjana dan memiliki sertifikat pendidik, serta mengadakan pelatihan-pelatihan dengan memanfaatkan teknologi. Harapannya untuk penelitian

selanjutnya dapat menggunakan metode lain dalam melakukan pengelompokan dan analisis *cluster* untuk hasil yang lebih optimal serta dapat menambahkan indikator-indikator pendidikan yang lainnya.

5. Referensi

- Alfina, T., Santosa, B., & Barakbah, A. R. (2012). Analisa Perbandingan Metode Hierarchical Clustering, K-means dan Gabungan Keduanya dalam Cluster Data (Studi kasus: Problem Kerja Praktek Jurusan Teknik Industri ITS). *Teknik Its*, 1(1), 521–525.
- Aprilia A.P, N. W. A., Srinade, I. G. A. M., & Sari, K. (2016). Pengelompokan Desa/Kelurahan Di Kota Denpasar Menurut Indikator Pendidikan. *E-Jurnal Matematika*, 5(2), 38–43. <https://doi.org/10.24843/mtk.2016.v05.i02.p119>
- Dini, S. K., & Fauzan, A. (2020). Clustering Provinces in Indonesia based on Community Welfare Indicators. *EKSAKTA: Journal of Sciences and Data Analysis*, 1(1), 56–63. <https://doi.org/10.20885/eksakta.vol1.iss1.art9>
- Hocking, R. R. (1997). Methods and Application of Linear Models. *European Journal of Orthodontics*, 19(2). <https://doi.org/https://doi.org/10.1093/ejo/19.2.232>
- Karti, H. S., & Irhamah. (2013). Pengelompokan Kabupaten / Kota di Provinsi SMA / SMK / MA dengan Metode C-Means dan Fuzzy C-Means. *Jurnal Sains Dan Seni Pomits*, 2(2), 288–293. <https://doi.org/10.12962/j23373520.v2i2.4860>
- Ningrum, H., Irawan, E., & Lubis, M. R. (2021). Implementasi Metode K-Medoids Clustering Dalam Pengelompokan Data Penyakit Alergi Pada Anak. *Jurasik (Jurnal Riset Sistem Informasi Dan Teknik Informatika)*, 6(1), 130. <https://doi.org/10.30645/jurasik.v6i1.277>
- Pulungan, N., Suhada, & Suhendro, D. (2019). Penerapan Algoritma K-Medoids Untuk Mengelompokkan Penduduk 15 Tahun Keatas Menurut Lapangan Pekerjaan Utama. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 3(1), 329–334. <https://doi.org/10.30865/komik.v3i1.1609>
- Purba, L., Saifullah, & Dewi, R. (2019). Pengelompokan Kasus Penyakit Aids Berdasarkan Provinsi Dengan Data Mining K-Medoids Clustering. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 3(1), 687–694. <https://doi.org/10.30865/komik.v3i1.1679>
- Soraya, E., & Suryadi. (2019). Pengembangan Lembaga Pendidikan sebagai Organisasi Pengembangan Lembaga Pendidikan dan Pembelajaran (studi pada penyelenggaraan PPG di LPTK). *IMProvement: Jurnal Ilmiah Untuk Peningkatan Mutu Pendidikan*, 6(2), 28–44.

Statistik, B. P. (2020). *Statistik Pendidikan 2020* (S. S. P. dan K. Sosial (ed.)). Badan Pusat Statistik.

[https://www.bps.go.id/publication/2020/11/27/347c85541c34e7dae54395a3/statistik-
pendidikan-2020.html](https://www.bps.go.id/publication/2020/11/27/347c85541c34e7dae54395a3/statistik-
pendidikan-2020.html)

Wira, B., Budianto, A. E., & Wiguna, A. S. (2019). Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 Di Universitas Kanjuruhan Malang. *Jurnal Terapan Sains & Teknologi*, 1(3), 54–69.