PRIMER: Jurnal Ilmiah Multidisiplin
2026, Vol. 4, No. 4, 203-210
https://doi.org/10.55681/primer.v4i4.553

Pengelompokkan Provinsi di Indonesia Berdasarkan Faktor-Faktor yang Memengaruhi Angka Partisipasi Kasar Perguruan Tinggi Menggunakan Metode *K-Medoids*

Indah Suci Parwati[1], Admi Salma[2]

[1234] Universitas Negeri Padang
Corresponding Author's e-mail : indahsucipurwati03@gmail.com

e-ISSN: 2985-7996

Article History:
Received: 11-07-2026
Revised : 01-08-2026
Accepted: 26-08-2026

© 2026, The Author(s)

Abstrak: Ketidakmerataan pendidikan di Indonesia terutama di perguruan tinggi masih menjadi tantangan besar. Hal ini terbukti dengan capaian Angka Partisipasi Kasar Perguruan Tinggi (APK-PT) di Indonesia pada tahun 2024 masih belum merata yang disebabkan oleh perbedaan faktor ekonomi dan sosial masyarakat. Penelitian ini bertujuan untuk mengelompokkan 38 provinsi di Indonesia berdasarkan faktor-faktor yang memengaruhi angka partisipasi kasar perguruan tinggi menngunakan metode Clustering K-Medoids. Variabel yang digunakan meliputi PDRB, Jumlah Pengeluaran Perkapita, Persentase Penduduk Miskin, Jumlah Perguruan Tinggi, dan Jumlah Tenaga Pendidik dengan menggunkan data tahun 2024 yang diperoleh dari Badan Pusat Statistik (BPS). Jumlah klaster optimal yang ditentukan menggunakan metode silhouette coefficient, yang menghasilkan dua klaster. Klaster 1 terdiri dari 27 provinsi yang dicirikan dengan PDRB dan PDRB dan jumlah pengeluaran perkapita tertinggi, serta persentase penduduk miskin terendah. Klaster 2 terdiri dari 11 provinsi dengan persentase penduduk miskin tertinggi dan PDRB terendah.

Kata Kunci: Angka Partisipasi Kasar Perguruan Tinggi, *Cluster, Clustering Large Application*

How to cite: Indah Suci Purwati, & Admi Salma. (2026). Pengelompokkan Provinsi di Indonesia Berdasarkan Faktor-Faktor yang Memengaruhi Angka Partisipasi Kasar Perguruan Tinggi Menggunakan Metode K-Medoids. *PRIMER : Jurnal Ilmiah Multidisiplin*, *4*(4), 203–210. https://doi.org/10.55681/primer.v4i4.553

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

PENDAHULUAN

Pendidikan merupakan kebutuhan dasar manusia yang berperan penting dalam meningkatkan kualitas sumber daya manusia (SDM) dan mendukung pembangunan berkelanjutan. Hal ini sejalan dengan tujuan *Sustainable Development Goals* (SDGs) yang menekankan pentingnya pendidikan inklusif, berkualitas, dan merata bagi seluruh masyarakat (Soraya & Suryadi, 2019). Dalam konteks pendidikan tinggi, Angka Partisipasi Kasar Perguruan Tinggi (APK-PT) menjadi salah satu indikator utama untuk mengukur tingkat partisipasi masyarakat pada jenjang perguruan tinggi.

Angka partisipasi kasar merupakan persentase perbandingan antara jumlah penduduk yang sedang menempuh pendidikan pada jenjang tertentu, tanpa memperhatikan Batasan usia, dengan jumlah penduduk yang berada pada kelompok usia resmi untuk jenjang pendidikan tersebut(Mukhaiyar et al., 2022). Berdasarkan data Badan Pusat Statistik (BPS, 2024), APK-PT Indonesia masih berada pada kisaran 31-31% dan belum mencapai target nasional sebesar 37,63%. Selain itu, terdapat kesenjangan yang cukup besar antarprovinsi, dimana Daerah Istimewa Yogyakarta memiliki APK-PT tertinggi, sedangkan beberapa provinsi di kawasan timur Indonesia masih menunjukkan capaian yang rendah. Kondisi ini mengindikasikan adanya ketimpangan akses pendidikan tinggi yang dipengaruhi oleh faktor ekonomi, sosial, serta ketersediaan sarana pendidikan (Bara et al., 2023). Meskipun pendidikan tinggi tidak termasuk dalam program wajib belajar nasional, tingkat partisipasinya tetap memiliki peranan penting dalam mendukung pembangunan. Pendidikan tinggi dipandang sebagai bentuk investasi yang mampu meningkatkan produktivitas individu sekaligus mendorong pertumbuhan ekonomi. Dengan demikian, rendahnya tingkat partisipasi pendidikan tinggi dapat berdampak pada kualitas tenaga kerja serta daya saing suatu negara.

Sejumlah penelitian terdahulu telah menggunakan metode *K-Medoids* untuk melakukan pengelompokkan provinsi di Indonesia berdasarkan indikator Pendidikan. Fialine et al (2021) menerapkan metode *K-Medoids* untuk mengelompokkan provinsi di Indonesia berdasarkan delapan indikator Pendidikan, yaitu, Angka Partisipasi Kasar (APK), Angka Partisipasi Murni (APM), Angka Partisipasi Sekolah (APS), persentase guru layak, Rata-Rata Lama Sekolah (RLS), angka melanjutkkan sekolah, Tingkat penyelesaian sekolah, dan angka putus sekolah. Hasil penelitian tersebut membentuk tiga kelompok provinsi yang memiliki karakteristik indikator Pendidikan tinggi, sedang, dan rendah. Penelitian lainnya dilakukan oleh Apridayanti et al (2024) mengenai pengelompokkan provinsi di Indonesia berdasarkan indikator Pendidikan tahun 2021 dengan menggunakan metode *K-Medoidsi.* Dalam penelitian tersebut, *Davies-Bouldin Index* (DBI) digunakan untuk menentukan jumlah *cluster* yang optimal, dengan jumlah *cluster* yang diuji mulai dari 2 hingga 10. Hasil penelitian menunjukkan bahwa lima *cluster* merupakan jumlah kelompok yang optimal karena menghasilkan nilai DBI terendah, yaitu 1,31. Hasil pengelompokkan tersebut menunjukkan adanya perbedaan karakteristik indikator Pendidikan pada masing-masing *cluster* provinsi.

Berdasarkan penelitian terdahulu yang dilakukan dapat disimpulkan perbedaan karakteristik antarprovinsi menyebabkan permasalahan APK-PT bersifat heterogen sehingga diperlukan pendekatan analisis yang mampu mengidentifikasikan pola kemiripan wilayah. Salah satu metode yang dapat digunakan adalah *Custering*, yaitu teknik pengelompokkan objek berdasarkan kesamaan karakteristik atau kedekatan antarobjek (Irhamni et al., 2014). Dari berbagai metode klaster yang tesedia, *K-Medoid* menjadi pilihan dalam penelitian ini karena metode ini lebih tahan terhadap data yang memiliki nilai sangat tinggi atau rendah dibandingkan data lainnya. Kodisi tersebut membuat hasil pengelompokkan tidak mudah berubah akibat adanya data ekstrem. Mengingat bahwa terdapat variasi yang cukup ekstrem pada data faktor-faktor yang

memengaruhi angka partisipasi kasar perguruan tinggi di Indonesia. Selain itu, *K-Medoids* menggunakan salah satu objek aktual dalam dataset sebagai pusat klaster yang disebut *medoid.* Berbeda dengan *K-Means* yang menggunakan nilai rata-rata (*mean)* sebagai pusat klaster (Kaufman & Rousseeuw, 1990). Melalui pengelompokkan provinsi, dapat diperoleh informasi mengenai wilayah-wilayah dengan karakteristik yang serupa sehingga hasilnya dapat menjadi dasar pengambilan kebijakan untuk meningkatkan pemerataan akses pendidikan tinggi di Indonesia.

METODE PENELITIAN

Sumber Data dan Variabel Penelitian

Pada penelitian ini menggunakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS). Berdasarkan penelitian yang dilakukan oleh Ga et al (2022). Data mencangkup faktor-faktor yang memengaruhi angka partisipai kasar perguruan tinggi di Indonesia tahun 2024. Data yang diambil dari BPS yaitu PDRB per Kapita (X_1), Pengeluaran per Kapita (X_2), Persentase penduduk miskin (X_3), Jumlah perguruan tinggi (X_4), dan Jumlah tenaga pendidik (X_5)

Teknik Analisis Data

Seluruh tahapan pengolahan dan analisis data dalam penelitian ini dilakukan menggunakan bantuan *software* Rstudio. Adapun tahapan analisisnya yaitu sebagai berikut

1. Standarisasi Data

Sebelum analisis klaster dilakukan, standarisasi data terlebih dahulu dikarenakan setiap variabel memiliki skala yang berbeda. Hal ini dilakukan agar variabel dengan skala besar tidak mendominasi perhitungan jarak (Hair et al., 2019). Salah satu metode standarisasi yang umum digunakan adalah metode *Z-score*. Proses standarisasi data dilakukan dengan rumus berikut.

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j} \qquad (1)$$

dengan, Z_{ij} adalah nilai standar variabel ke-*j* pada objek ke-*i* , x_{ij} adalah nilai *x* variabel ke-*j* pada objek ke-*i*, $\bar{x}_j$ adalah rata-rata variabel ke-*j*, S_j adalah simpangan baku variabel ke-*j.*

2. Penentuan Jumlah Klaster Optimal dengan Metode *Silhouette Coefficient*

Metode *Silhouette Coefficient* digunakan untuk menentukan jumlah klaster terbaik. Nilai *Silhouette Coefficient* berada pada rentang (-1) hingga (1). Klaster yang dikatakan terbentuk baik bila nilai mendekati 1 dan sebaliknya jika nilai mendekati angka -1 menunjukkan kemungkinan objek lebih sesuai berada pada klaster lain. Berikut persamaan *Silhouette Coefficient*:

$$S(i) = \frac{b(i)-a(i)}{\max\{a(i).b(i)\}} \qquad (2)$$

Dengan $a(i)$ adalah rata-rata jarak objek i terhadap objek lain dalam klaster yang sama dan $b(i)$ adalah rata-rata jarak minimun objek *i* terhadap klaster lain.

3. Analisis Klaster *K-Medoids*

Adapun cara kerja *K-Medoids* menurut Han and Kamber (2012):

1) Menentukan jumlah klaster sebanyak k (jumlah klaster)
2) Menentukan medoid awal secara acak pada objek amatan
3) Menempatkan objek amatan yang bukan medoid kedalam klaster yang memiliki jarak terdekat terhadap suatu pusat klaster dengan menggunakan jarak *manhattan*
4) Hitung total simpangan (S) menggunakan Persamaan (4), kemudian kembali lakukan langkah 2 – 4, iterasi berhenti jika nilai S > 0

$$S = b - a \quad (3)$$

a : jumlah jarak tedekat antara objek ke medoid awal,
b : jumlah jarak terdekat antara objek ke medoid baru

4. Interpretasi Hasil Klaster

Interpretasi hasil klaster merupakan tahap dalam analisis klaster yang dilakukan dengan mengidentifikasi anggota dan karakteristik masing-masing klaster berdasarkan nilai rata-rata variabel yang digunakan. Karakteristik tersebut dibandingkan untuk mengetahui perbedaan antarklaster. Ukuran yang biasa digunakan dalam proses interpretasi yaitu *centroid* klaster.

HASIL DAN PEMBAHASAN

Statistika Deskriptif

Statistik deskriptif berfungsi untuk memberikan sebuah gambaran mengenai variabel yang digunakan yaitu PDRB (X1), Jumlah Pengeluaran Perkapita (X2), Persentase Penduduk Miskin (X3), Jumlah Perguruan Tinggi (X4), dan Jumlah Tenaga Pendidik (X5). Analisis deskriptif yang diperoleh dapat dilihat pada Tabel 1 berikut ini

Tabel 1. Statistik Deskriptif

Variabel	min	mean	max
PDRB(X1)	18105	84343,74	344350
Jumlah Pengeluaran Perkapita(X2)	5707	11538,29	19953
Persentase Penduduk Miskin(X3)	3,8	10,66474	29,66
Jumlah Perguruan Tinggi(X4)	8	77,28947	367
Jumlah Tenaga Pendidik(X5)	1,077	204,427	997

Berdasarkan Tabel 1. Dapat dilihat bahwa kelima variabel memiliki rentang nilai yang berbeda cukup jauh antara nilai minimum dan maksimum. PDRB per kapita memiliki rentang yang sangat lebar, dari Rp 18105 juta hingga Rp344350 juta, yang mencerminkan ketimpangan kapasitas ekonomi antarwilayah yang signifikan. Variabel jumlah pengeluran perkapita memiliki nilai minimum sebesar Rp5.707, rata-rata sebesar Rp11.538,29, dan maximum sebesar Rp19.953. kondisi ini menunjukkan bahwa tingkat pengeluaran per kapita antarprovinsi memiliki variasi, dengan rentang sebesar Rp14.246. Sementara itu, variabel persentase penduduk miskin menunjukkan adanya perbedaan tingkat kemiskinan yang cukup besar antarprovinsi yang dilihat dari rentang nilai sebesar 28,86% dengan nilai minimun sebesar 3,8%, rata-rata 10,66474%, dan nilai maximum sebesar 29,66%. Jumlah perguruan tinggi juga menunjukkan perbedaan yang cukup besar dengan nilai minimum 8 dan maximum 367. Begitupun dengan jumlah tenaga pendidik berkisar pada 1,077 hingga 997. Adanya variasi ekstrem pada beberapa variabel diatas, mengindikasikan keberadaan *outlier* (pencilan) yang menjadi pertimbangan utama dalam pemilihan metode *K-Medoids* sebagai metode yang memiliki keunggulan *robust* (tahan) terhadap pencilan.

Standarisasi Data

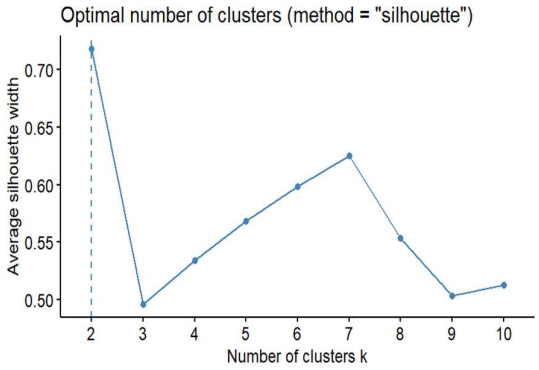
Sebelum analisis klaster dilakukan, seluruh variabel terlebih dahulu distandarisasi menggunakan metode *Z-score* karena masing-masing variabel memiliki satuan dan skala pengukuran yang berbeda-beda. Proses standarisasi menghasilkan nilai dengan rata-rata 0 dan satnadar deviasi 1 untuk setiap variabel. Dengan demikian, setiap variabel memiliki skala yang sebanding sehingga perbedaan satuan dan skala tidak mendominasi perhitungan jarak dalam proses pengelompokkan. Hasil standarisasi disajikan pada Tabel 2.

Tabel 2. Standarisasi Data

Provinsi	x1	x2	x3	x4	x5
1	-0.6521	-0.2944	0.3102	0.0631	-0.5881
2	-0.1731	-0.0316	-0.5457	1.3351	-0.5633
3	-0.4388	0.0727	-0.8237	-0.0253	-0.5861
⋮	⋮	⋮	⋮	⋮	⋮
37	0.5535	-1.5096	2.6599	-0.7332	0.1595
38	-1.0649	-2.3604	2.9835	-0.7553	0.0425

Penentuan Jumlah Klaster

Penentuan jumlah klaster optimal dilakukan menggunakan metode *Silhouette Coefficient*. Hasil metode ini ditunjukkan pada Gambar 1



Gambar 1. Plot Metode *Silhouette*

Berdasarkan metode *Silhouette Coefficient* pada Gambar 1, diperoleh nilai *average silhouette width* untuk jumlah klaster 2 hingga 10. Nilai *silhouette* tertinggi diperoleh pada k=2, yaitu sebesar 0,718, sedangkan nilai terendah terdapat pada k=3 yaitu sekitar 0,495. Nilai *silhoette* pada k=2 merupakan nilai tertinggi dibandingkan jumlah klaster lainnya, sehingga dua klaster ditetapkan sebagai jumlah klaster optimal.

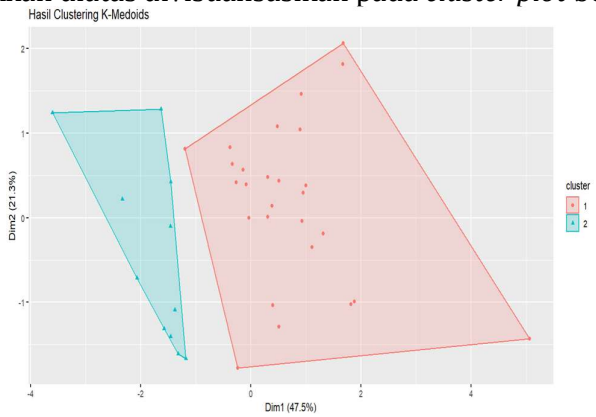
Analisis Klaster *K-Medoids*

Berdasarkan jumlah klaster optimal sebanyak 2 klaster, selanjutnya dilakukan analisis klaster *K-Medoids* menggunakan data yang telah distandarisasi. Hasil pengelompokkan 38 provinsi di Indonesia disajikan pada Tabel 3.

Tabel 3. Hasil Pengelompokkan Analisis Klaster *K-Medoids*

Klaster	Banyak Anggota (Provinsi)	Nama Provinsi
1	27	Aceh, Sumatera Utara,Sumatera Barat, Riau, Jambi, Sumatera Selatan, Lampung, Kep.Bangka Belitung, Kepulauan Riau, Dki Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Kalimatan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Maluku
2	11	Bengkulu, Nusa Tenggara Timur, Gorontalo, Sulawesi Barat, Maluku Utara, Papua Barat, Papua Barat Daya, Papua, Papua Selatan, Papua Tengah, Papua Pegunungan

Hasil pengelompokkan diatas divisualisasikan pada *cluster plot* berikut.



Gambar 2. Plot Hasil Analisis Klaster *K-Medoids*

Berdasarkan plot hasil klaster *K-Medoids,* terlihat terbentuk 2 klaster, yaitu klaster 1 (merah), dan klaster 2 (biru). Klaster 1 memiliki jumlah anggota yang lebih banyak dengan penyebaran data yang relatif luas, sedangkan klaster 2 dengan jumlah anggota yang lebih sedikit dan cenderung memiliki titik-titik yang lebih berdekatan. Dimensi 1 (Dim1) menjelaskan 47,5% keberagaman data, sedangkan Dimensi 2 (Dim2) menjelaskan sebesar 21.3% , sehingga secara kumulatif kedua dimensi mampu merepresentasikan68,8% keragaman total data.

Karakteristik masing-masing klaster berdasarkan nilai rata-rata setiap variabel disajikan pada Tabel 4.

Tabel 4. Karakteristik Klaster Berdasarkan Nilai Rata-Rata

Klaster	1	2
X1	92.725.00	63771.55
X2	12.417.556	9380.091
X3	7.791481	17.717273
X4	100.7778	19.63636
X5	60.38393	557.98718

Berdasarkan Tabel 4, masing-masing klaster menunjukkan karakteristik yang berbeda berdasarkan nilai rata-rata setiap variabel pada data faktor-faktor yang

memengaruhi angka partisipasi kasar perguruan tinggi, dengan interpretasi sebagai berikut.

a. Klaster 1 merupakan kelompok provinsi dengan PDRB(X1), Jumlah Pengeluaran Perkapita(X2), dan Jumlah Perguruan Tinggi(X4) tertinggi, sedangkan Persentase Penduduk Miskin(x3) dan Jumlah Tenaga Pendidik(X5) yang rendah. Hal ini menunjukkan bahwa klaster 1 dapat dikatakan sebagai kelompok provinsi dengan kondisi ekonomi yang relatif lebih tinggi, tingkat kemiskinan yang lebih rendah, serta kesediaan perguruan tinggi yang lebih banyak. Secara umum, karakteristik ini memiliki kodisi yang lebih mendukung peningkatan pasrtisipasi masyarakat dalam pendidikan tinggi.
b. Klaster 2 merupakan kelompok provinsi dengan Persentase Penduduk Miskin(X3) dan Jumlah Tenaga Pendidik(X5) tertinggi. Sedangkan PDBR(X1), Jumlah Pengeluaran Perkapita(X2), dan Jumlah Perguruan Tinggi(X4) yang lebih rendah. Hal ini menunjukkan bahwa provinsi dalam klaster 2 secara umum berada dalam kondisi ekonomi dan akses terhadap pendidikan tinggi yang relatif kurang mendukung dibandingkan dengan klaster 1.

KESIMPULAN DAN SARAN

Pengelompokkan 38 provinsi berdasarkan faktor-faktor yang memengaruhi angka partisipasi kasar perguruan tinggi di Indonesia tahun 2024 menggunakan analisis klaster K-Medoids menghasilkan 2 klaster optimal berdsarkan metode Silhouette Coefficient. Klaster 1 beranggotakan 27 provinsi yaitu Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Lampung, Kep.Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Kalimatan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara dan Maluku dengan rata-rata PDRB, dan Jumlah Pengeluaran Perkapita tertinggi, serta persentase penduduk miskin dan jumlah tenaga pendidik terendah. Klaster 2 beranggotakan 11 provinsi yaitu Bengkulu, Nusa Tenggara Timur, Gorontalo, Sulawesi Barat, Maluku Utara, Papua Barat, Papua Barat Daya, Papua, Papua Selatan, Papua Tengah, Papua Pegunungan dengan rata-rata persentase penduduk miskin tertinggi dan PDRB terendah. Hasil pengelompokkan ini diharapkan dapat dijadikan sebagai acuan pengambilan keputusan bagi pemerintah dalam merumuskan kebijakan peningkatan pendidikan terutama pendidikan tinggi di Indonesia yang lebih terarah sesuai karakteristik setiap klaster.

DAFTAR PUSTAKA

Apridayanti, A. Z., Fathurahman, M., & Prangga, S. (2024). *An Implementation of K-Medoids Method in Provincial Clustering Based on Education Indicator Data*. *7*(2), 191–198.

Bara, E. G., Fitriani, F., Indrasetianingsih, A., & Dukuh. (2023). Pemodelan Geographically Weighted Regression Pada Angka Partisipasi Kasar Perguruan Tinggi Di Indonesia Tahun 2022 Abstrak Pendahuluan Di Indonesia diketahui bahwa kualitas sumber daya manusia masih rendah karena. *Jurnal Sains Matematika Dan Statistika*, *9*(2), 76–88.

BPS. (2024). *Statistik Pendidikan Indonesia 2024*.

Fialine, A. P., Alodia, D. A., Endriani, D., & Widodo, E. (2021). *Implementasi Metode K-Medoids Clustering untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan*. *02*(02), 1–13.

Ga, E., Fitriani, F., Indrasetianingsih, A., Dukuh, J., Xii, M., Surabaya, K., Timur, J., Partisipasi, A., Perguruan, K., & Apk-pt, T. (2022). *Pemodelan Geographically Weighted Regression Pada Angka Partisipasi Kasar Perguruan Tinggi Di Indonesia Tahun 2022 Abstrak*

Pendahuluan Di Indonesia diketahui bahwa kualitas sumber daya manusia masih rendah karena. *9*(2), 76–88.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis Eighth Edition*.

Han and Kamber. (2012). *Data Mining Concepts and Tehniques*.

Irhamni, F., Damayanti, F., K, B. K., & A, M. (2014). Optimalisasi Pengelompokan Kecamatan Berdasarkan Indikator Pendidikan Menggunakan Metode Clustering dan Davies Bouldin Index. *Seminar Nasional Sains Dan Teknologi UMJ*, *11*, 1–5.

Kaufman & Rousseeuw. (1990). *FINDING GRUPS IN DATA*.

Mukhaiyar, U., Rontos, F., Handoko, K., & Kardiyanti, S. (2022). Analisis Faktor-Faktor yang Memengaruhi Angka Partisipasi Kasar SMA/Sederajat di Indonesia Menggunakan Regresi Ridge. *Euler : Jurnal Ilmiah Matematika, Sains Dan Teknologi*, *10*(2), 222–234. https://doi.org/10.34312/euler.v10i2.15903

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 Manajemen Pendidikan*, *6*(2), 28–44.