

## Algoritma K-Means dan Analisis Komponen Utama untuk Mengatasi Multikolinearitas pada Pengelompokan Kabupaten Tertinggal

Firna Aviliana <sup>(1)\*</sup>, Putriaji Hendikawati <sup>(2)</sup>

<sup>1</sup> Departemen Matematika, Universitas Negeri Semarang, Semarang, Indonesia

<sup>2</sup> Departemen Statistika Terapan dan Komputasi, Universitas Negeri Semarang, Semarang, Indonesia

e-mail : aviliana17@students.unnes.ac.id, putriaji.mat@mail.unnes.ac.id.

\* Penulis korespondensi.

Artikel ini diajukan 28 Juli 2024, direvisi 7 Oktober 2024, diterima 8 Oktober 2024, dan dipublikasikan 30 September 2025.

### Abstract

*Underdeveloped areas are regions that frequently face developmental challenges in various aspects such as infrastructure, education, and healthcare. Presidential Regulation Number 63 of 2020 designates 62 regencies in Indonesia as underdeveloped areas. This study categorizes the 62 underdeveloped regencies based on education and health indicators. The methods used are the k-means algorithm and principal component analysis due to multicollinearity in the data. MANOVA is conducted to determine the influence of the cluster results on the Human Development Index (HDI), Average Years of Schooling (AYS), Expected Years of Schooling (EYS), and Life Expectancy (LE). Due to multicollinearity in the education indicator data, principal component analysis was performed, resulting in three main components. The k-means analysis groups the 62 regencies into three clusters based on education indicators and two clusters based on health indicators. Further analysis using MANOVA shows the influence of the education and health clusters on HDI, AYS, EYS, and LE, indicated by statistical test results showing  $p\text{-value} < \alpha(0.05)$ . Thus, education and health indicators influence the categorization of underdeveloped areas.*

**Keywords:** K-Means Algorithm, Principal Component Analysis, Multicollinearity, MANOVA, Underdeveloped Regencies

### Abstrak

Daerah tertinggal merupakan daerah yang sering menghadapi tantangan pembangunan dalam berbagai hal, seperti infrastruktur, pendidikan, dan kesehatan. Peraturan Presiden Nomor 63 tahun 2020 menetapkan 62 kabupaten di Indonesia ke dalam daerah tertinggal. Penelitian ini mengelompokkan 62 kabupaten tertinggal berdasarkan indikator pendidikan dan kesehatan. Metode yang digunakan adalah algoritma k-means dan analisis komponen utama karena terjadi multikolinearitas dalam data. MANOVA dilakukan untuk mengetahui kesesuaian pengaruh hasil *cluster* terhadap nilai Indeks Pembangunan Manusia (IPM), Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), dan Umur Harapan Hidup (UHH). Adanya multikolinearitas dalam data indikator pendidikan sehingga dilakukan analisis komponen utama dan terbentuk tiga komponen utama. Analisis k-means mengelompokkan 62 kabupaten ke dalam tiga *cluster* berdasarkan indikator pendidikan dan dua *cluster* berdasarkan indikator kesehatan. Hasil analisis lanjutan menggunakan MANOVA menunjukkan adanya kesesuaian pengaruh *cluster* pendidikan dan kesehatan terhadap nilai IPM, RLS, HLS, dan UHH ditandai dari hasil uji statistik yang menunjukkan nilai  $p\text{-value} < \alpha(0,05)$ . Dengan demikian, indikator pendidikan dan kesehatan memberikan pengaruh pada penentuan kelompok pada daerah tertinggal.

**Kata Kunci:** Algoritma K-Means, Analisis Komponen Utama, Multikolinearitas, MANOVA, Kabupaten Tertinggal

## 1. PENDAHULUAN

Peningkatan kesejahteraan masyarakat di berbagai wilayah adalah tujuan utama pembangunan nasional. Pembangunan nasional melibatkan banyak pihak, termasuk pemerintah, sektor swasta,



Artikel ini didistribusikan mengikuti lisensi Atribusi-NonKomersial CC BY-NC sebagaimana tercantum pada <https://creativecommons.org/licenses/by-nc/4.0/>.

masyarakat sipil, dan organisasi internasional. Telah banyak program dan kebijakan yang diambil untuk meningkatkan kesejahteraan masyarakat di berbagai bidang, seperti ekonomi, pendidikan, kesehatan, dan infrastruktur. Namun, pada kenyataannya masih terdapat ketidaksetaraan atau kesenjangan pembangunan antarwilayah sehingga masih ditemukan kabupaten tertinggal di berbagai provinsi di Indonesia.

Pendidikan dan kesehatan merupakan komponen yang ikut menentukan kesejahteraan masyarakat. Amrullah et al. (2022) menyatakan bahwa pendidikan merupakan salah satu cara untuk meningkatkan kualitas sumber daya manusia, yang diharapkan dapat meningkatkan kesejahteraan manusia. Badan Pusat Statistik (BPS) menyatakan bahwa Angka Partisipasi Murni (APM) dan Angka Partisipasi Kasar (APK) menjadi indikator dalam keberhasilan pendidikan (Prasetyawati, 2023). Sitinjak et al., (2022) mengungkapkan bahwa pembangunan kesehatan merupakan komponen penting dari pembangunan nasional, yang berarti melakukan upaya kesehatan untuk memastikan bahwa setiap orang memiliki kesempatan untuk hidup dengan kesehatan yang optimal. Adanya sarana dan prasarana kesehatan yang baik, seperti puskesmas, rumah sakit, dan fasilitas kesehatan lainnya, serta tenaga kesehatan yang memadai mampu meningkatkan kesehatan masyarakat secara keseluruhan.

Peraturan Presiden (Perpres) Nomor 63 Tahun 2020 Tentang Penetapan Daerah Tertinggal Tahun 2020-2024 menyatakan bahwa daerah tertinggal adalah daerah kabupaten yang wilayah serta masyarakatnya kurang berkembang dibandingkan dengan daerah lain dalam skala nasional. Selanjutnya dalam Pasal 2 menyebutkan kriteria yang menjadi penetapan daerah tertinggal di antaranya adalah perekonomian masyarakat, sumber daya manusia, sarana dan prasarana, kemampuan keuangan daerah, aksesibilitas, dan karakteristik daerah. Peraturan Presiden tersebut menetapkan 62 kabupaten di Indonesia yang termasuk ke dalam daerah tertinggal. Ditinjau dari data BPS tahun 2022, yang tergolong ke dalam daerah tertinggal memiliki Indeks Pembangunan Manusia (IPM) sedang, bahkan beberapa kabupaten tergolong rendah. IPM didasarkan pada tiga dimensi utama, yaitu umur panjang dan hidup sehat, pengetahuan, dan standar hidup yang layak (Setiawan et al., 2022). Umur panjang dan hidup sehat sangat dipengaruhi oleh kualitas kesehatan yang dicerminkan oleh Umur Harapan Hidup (UHH), pengetahuan dipengaruhi oleh kualitas pendidikan, yang dicerminkan oleh Rata-rata Lama Sekolah (RLS) dan Harapan Lama Sekolah (HLS), sedangkan standar hidup layak dipengaruhi oleh kualitas ketenagakerjaan pada daerah tersebut yang dicerminkan oleh Pendapatan Nasional Bruto (PNB) per kapita (PPP).

Pendidikan dan kesehatan memainkan peran penting dalam peningkatan IPM serta mewujudkan tercapainya kesejahteraan masyarakat. Indikator-indikator seperti fasilitas, akses, serta pelayanan dalam bidang pendidikan dan kesehatan menjadi penentu untuk mengukur kualitas hidup masyarakat. Kabupaten yang masih dalam kategori daerah tertinggal perlu dikelompokkan berdasarkan indikator pendidikan dan kesehatan agar proses pembangunan terlaksana tepat sasaran. Hasil pengelompokan ini berguna untuk mengidentifikasi kebutuhan khusus dari setiap kelompok kabupaten, terutama dalam bidang pendidikan dan kesehatan. Hasil pengelompokan menjadi alat yang penting dalam upaya meningkatkan kesejahteraan masyarakat di kabupaten tertinggal agar tercapai pemerataan pembangunan nasional.

Analisis *cluster* merupakan teknik mengelompokkan objek atau data ke dalam *cluster* berdasarkan kesamaan dan kemiripan karakteristiknya. Dalam satu *cluster* memiliki karakteristik yang sama dan memiliki perbedaan karakteristik pada objek *cluster* lain (Amrullah et al., 2022). Salah satu metode untuk mengelompokkan data yaitu menggunakan algoritma k-means. Algoritma k-means bekerja dengan mengelompokkan data ke dalam *k* kelompok (*cluster*) yang ditentukan sebelumnya, dengan meminimalkan variasi atau jarak antara titik data dalam satu *cluster* (Aryanto et al., 2024). K-means merupakan metode non-hirarki yang paling banyak digunakan untuk pengelompokan karena kemudahan dan kemampuannya untuk mengelompokkan dalam jumlah data yang besar dengan waktu komputasi yang cepat dan efisien (Lenama et al., 2023). Algoritma k-means merupakan metode yang mudah dipahami dan diimplementasikan, serta mampu menghasilkan hasil yang cukup akurat.



Beberapa penelitian terdahulu yang melakukan pengelompokkan menggunakan algoritma k-means, di antaranya Amrullah et al. (2022) mengenai pengelompokkan faktor penunjang pendidikan di Kabupaten Karawang, diperoleh hasil dua *cluster* optimal dengan nilai Davies-Bouldin Index (DBI) 0,408. Penelitian lain dilakukan oleh Octaviyani et al. (2022) mengenai pengelompokkan gizi balita, diperoleh lima *cluster* optimal dengan nilai DBI 0,5224. Yudhistira & Andika (2023) juga melakukan penelitian pengelompokkan data nilai siswa menggunakan k-means, diperoleh tiga *cluster* terbaik dengan nilai *silhouette coefficient* 0,489.

Setelah diketahui kategori pengelompokkan daerah tertinggal tersebut, selanjutnya dilakukan uji kesesuaian pengaruh pengelompokkan tersebut terhadap variabel utama pembangunan nasional, yaitu IPM, RLS, HLS, dan UHH menggunakan MANOVA. MANOVA adalah uji statistik untuk mengukur pengaruh variabel independen (variabel bebas) pada skala kategori terhadap variabel dependen (variabel terikat) pada skala data kuantitatif (Pursitasari et al., 2024). Tujuan penelitian ini adalah menguji hubungan hasil *cluster* terhadap nilai IPM, RLS, HLS, dan UHH. Apakah pengelompokkan berdasarkan indikator pendidikan dan kesehatan mempengaruhi berbagai aspek pembangunan manusia di daerah tertinggal sehingga hasil *cluster* yang diberikan dapat mendukung pembuatan kebijakan yang lebih efektif dan tepat sasaran.

## 2. METODE PENELITIAN

Penelitian ini menggunakan analisis *cluster* dengan algoritma k-means untuk memperoleh hasil *cluster* kabupaten tertinggal berdasarkan indikator pendidikan dan kesehatan. Selanjutnya hasil *cluster* tersebut dianalisis lanjut menggunakan MANOVA untuk menguji apakah terdapat pengaruh yang signifikan dari hasil *cluster* yang terbentuk terhadap Indeks Pembangunan Manusia (IPM), Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah, dan Umur Harapan Hidup (UHH). Analisis MANOVA dilakukan untuk mengetahui kategori hasil *cluster* yang diperoleh memberikan kesesuaian pengaruh pada tingkat nilai IPM, RLS, HLS, dan UHH.

### 2.1 Data dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diambil dari *website* Badan Pusat Statistik (BPS) melalui laman <https://www.bps.go.id>, meliputi data indikator pendidikan dan kesehatan di 62 kabupaten tertinggal di Indonesia pada tahun 2022. Pada analisis *cluster*, data indikator pendidikan memuat 14 variabel, yaitu jumlah TK/RA ( $X_1$ ), jumlah SD/MI ( $X_2$ ), jumlah SMP/MTs ( $X_3$ ), jumlah SMA/SMK/MA ( $X_4$ ), jumlah guru TK/RA ( $X_5$ ), jumlah guru SD/MI ( $X_6$ ), jumlah guru SMP/MTs ( $X_7$ ), jumlah guru SMA/SMK/MA ( $X_8$ ), APM SD/MI ( $X_9$ ), APM SMP/MTs ( $X_{10}$ ), APM SMA/SMK/MA ( $X_{11}$ ), APK SD/MI ( $X_{12}$ ), APK SMP/MTs ( $X_{13}$ ), dan APK SMA/SMK/MA ( $X_{14}$ ), sedangkan indikator kesehatan memuat 5 variabel, yaitu jumlah dokter ( $X_{15}$ ), jumlah perawat ( $X_{16}$ ), jumlah bidan ( $X_{17}$ ), jumlah rumah sakit ( $X_{18}$ ), jumlah puskesmas ( $X_{19}$ ). Analisis MANOVA dilakukan setelah diperoleh hasil analisis *cluster* dua indikator yang masing-masing memuat variabel bebas dan terikat. Pada indikator pendidikan, variabel bebas yaitu hasil *cluster* pendidikan ( $X_{20}$ ), sedangkan variabel terikat meliputi IPM ( $Y_1$ ), RLS ( $Y_2$ ), dan HLS ( $Y_3$ ). Pada indikator kesehatan, variabel bebas yaitu hasil *cluster* kesehatan ( $X_{21}$ ), sedangkan variabel terikat meliputi IPM ( $Y_4$ ) dan UHH ( $Y_5$ ).

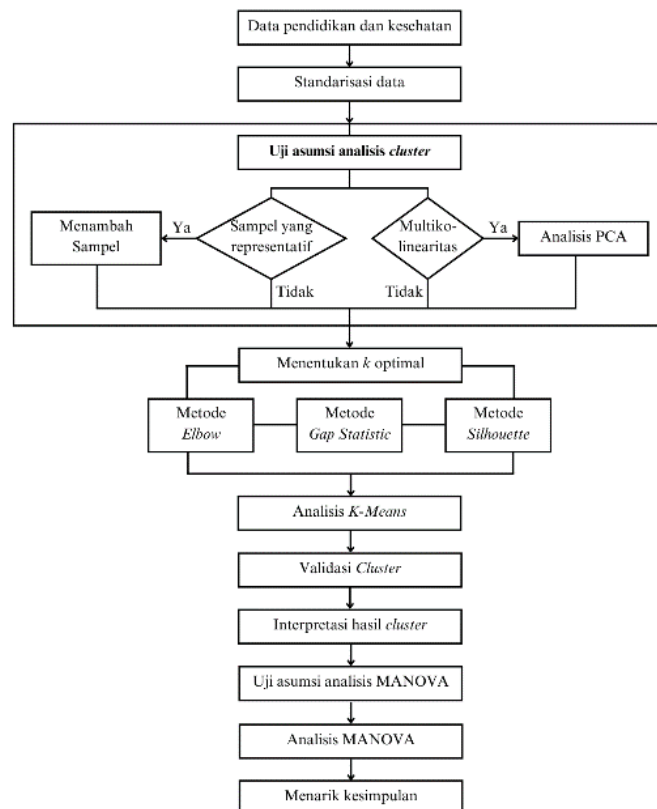
### 2.2 Tahapan Analisis Data

Analisis data dalam penelitian ini dilakukan dengan bantuan software RStudio. Tahapan analisis dapat dilihat pada Gambar 1 yang dimulai dengan pengumpulan data yang bersumber dari publikasi BPS. Setelah itu, dilakukan standarisasi data agar semua variabel memiliki skala yang sama dan tidak menimbulkan dominansi antarvariabel, mengingat data pendidikan dan kesehatan memiliki satuan yang berbeda. Standarisasi ini bertujuan menyamakan distribusi variabel sehingga lebih seragam (Wijaya et al., 2024). Selanjutnya, dilakukan pengujian asumsi *cluster* yang mencakup representativitas sampel dan uji nonmultikolinearitas. Apabila ditemukan multikolinearitas, maka dilakukan analisis komponen utama atau Principal Component Analysis (PCA).



Tahap berikutnya adalah menentukan jumlah cluster optimal ( $k$ ) menggunakan tiga metode, yaitu metode elbow, metode gap statistic, dan metode silhouette. Setelah jumlah cluster optimal diperoleh, analisis dilanjutkan menggunakan algoritma k-means. Validasi hasil clustering kemudian dilakukan dengan menggunakan Silhouette Index (SI) untuk memastikan kualitas pemisahan antarcluster. Selanjutnya, hasil cluster diinterpretasikan secara menyeluruh.

Tahapan berikutnya melibatkan pengujian asumsi MANOVA, termasuk uji normalitas multivariat dan uji homoskedastisitas. Setelah asumsi terpenuhi, analisis dilakukan menggunakan MANOVA untuk menguji perbedaan antarcluster berdasarkan variabel penelitian. Tahapan terakhir adalah menarik kesimpulan dari seluruh rangkaian analisis yang telah dilakukan.



**Gambar 1 Tahapan Penelitian**

### 3. HASIL DAN PEMBAHASAN

Penerapan algoritma k-means dengan analisis komponen utama dan dilanjut analisis MANOVA dalam pengelompokan kabupaten tertinggal berdasarkan indikator pendidikan dan kesehatan dapat membantu meningkatkan efektivitas upaya peningkatan pendidikan dan kesehatan di kabupaten tertinggal. Tahapan dalam analisis dapat diuraikan sebagai berikut.

#### 3.1 Analisis K-Means

Sebelum melakukan analisis *cluster*, uji asumsi *cluster* dilakukan pada masing-masing data indikator agar mendapatkan hasil yang lebih akurat. Hair et al. (2013, dikutip dalam Hamidah et al., 2022) menyebutkan uji asumsi analisis *cluster* meliputi sampel yang representatif dan tidak terjadi multikolinearitas.

- 1) Pengecekan sampel yang representatif dilakukan menggunakan uji KMO (*Kaiser Meyer Olkin*). Menurut Kaiser (1974, dikutip dalam Hamidah et al., 2022), sampel dikatakan mewakili populasi jika nilai KMO lebih dari 0,5. Dalam pengujian ini menunjukkan hasil KMO data



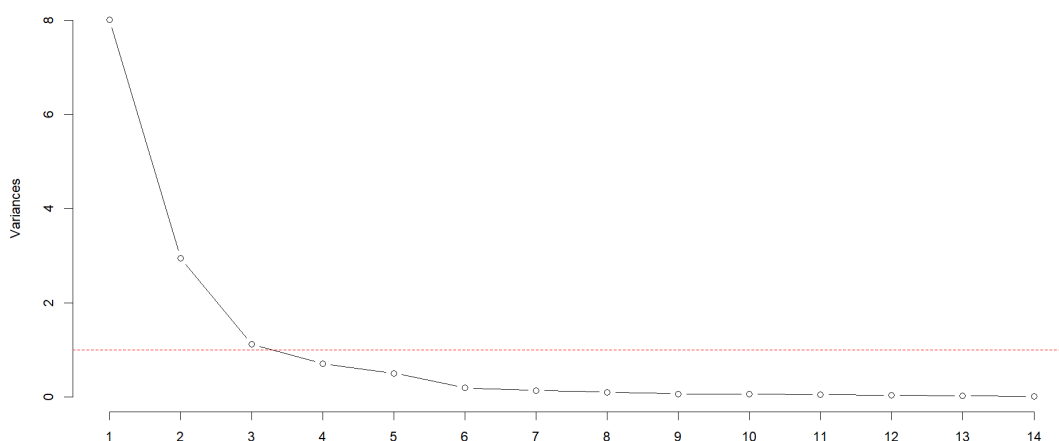
indikator pendidikan sebesar 0,82, sedangkan data indikator kesehatan memiliki nilai KMO sebesar 0,8. Dengan demikian kedua data indikator tersebut memenuhi syarat sampel yang representatif.

- 2) Uji multikolinearitas dilakukan dengan penghitungan nilai VIF antar variabel. Dikatakan memenuhi syarat multikolinearitas apabila nilai VIF lebih kecil dari 10 atau dinotasikan sebagai  $VIF < 10$  (Fajar & Indrawati, 2020). Hasil dari penghitungan nilai VIF pada data indikator pendidikan dapat dilihat dalam Tabel 7 di Lampiran A dan data indikator kesehatan dalam Tabel 1. Pada Tabel 7 di Lampiran A diketahui terdapat nilai  $VIF > 10$  maka data indikator pendidikan terindikasi multikolinearitas. Jika suatu data terjadi multikolinearitas menandakan adanya korelasi yang kuat antar variabel sehingga perlu dilakukan analisis komponen utama atau *Principal Component Analysis* (PCA) (Thamrin & Murni, 2022). Pada Tabel 1 memiliki nilai  $VIF < 10$ , artinya data indikator kesehatan tidak terjadi multikolinearitas.

**Tabel 1 Nilai VIF Antar Variabel Kesehatan**

Nilai VIF	$X_{15}$	$X_{16}$	$X_{17}$	$X_{18}$	$X_{19}$
$X_{15}$		2,735	2,153	2,939	2,941
$X_{16}$	2,992		2,586	3,152	2,917
$X_{17}$	2,552	2,801		3,490	3,495
$X_{18}$	1,129	1,107	1,132		1,115
$X_{19}$	1,433	1,298	1,436	1,414	

Scree Plot

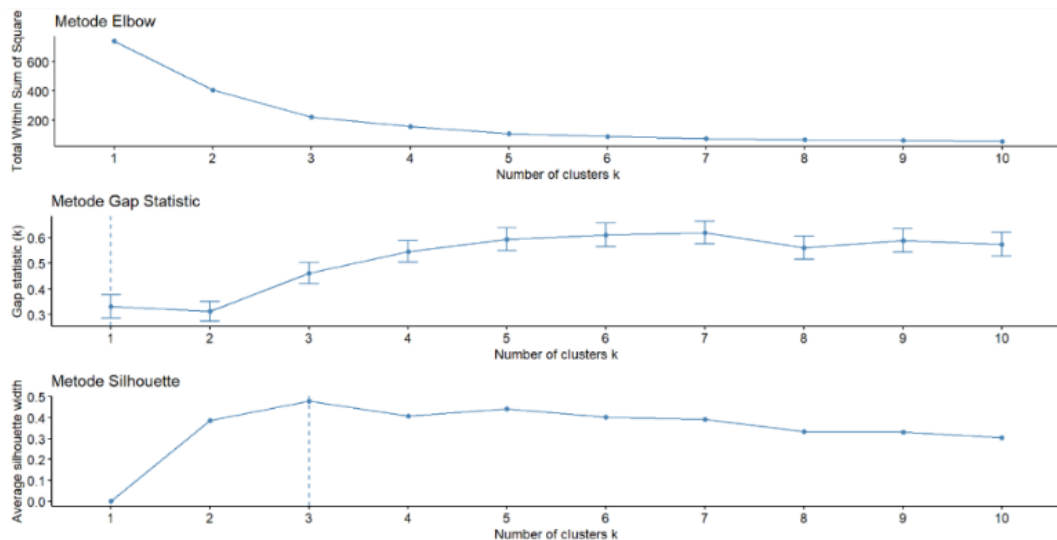


**Gambar 2 Scree Plot**

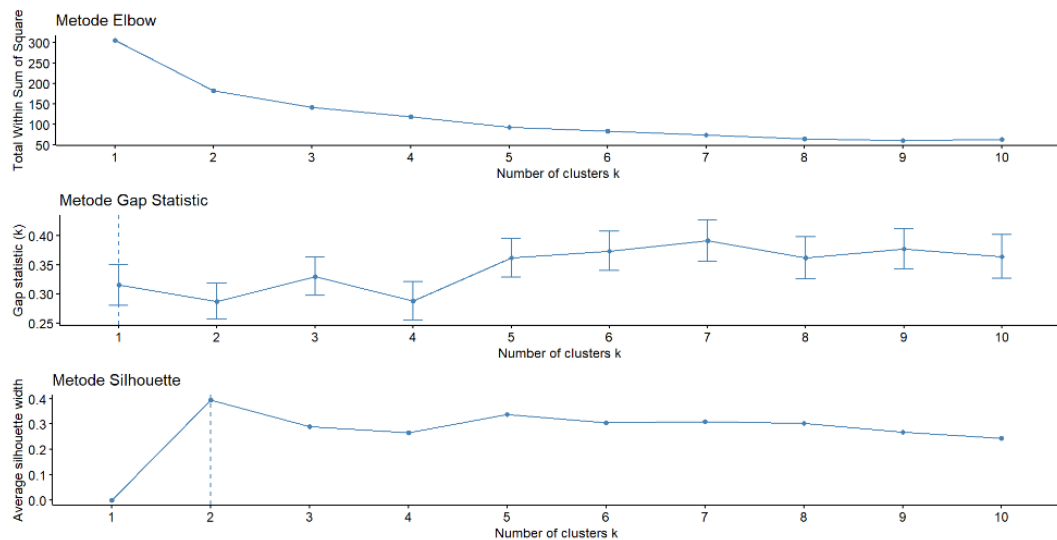
Analisis komponen utama atau *Principal Component Analysis* (PCA) adalah suatu cara yang digunakan untuk mereduksi variabel yang banyak dan mengatasi data yang terdapat multikolinearitas (Amelia & Kholijah, 2023). Menurut Mariana (2013) ada tiga kriteria pemilihan  $k$  komponen utama yaitu memilih akar ciri (nilai eigen) yang lebih dari satu, nilai proporsi varians kumulatif minimal 80%, dan menggunakan *scree plot*. Data indikator pendidikan mengalami multikolinearitas sehingga dilakukan analisis komponen utama terlebih dahulu. Pencarian nilai eigen dan proporsi varians kumulatif data indikator pendidikan dapat dilihat pada Tabel 8 di Lampiran A yang menunjukkan terdapat tiga komponen utama yang memiliki nilai eigen lebih dari satu serta PC 3 menunjukkan proporsi kumulatif sebesar 0,8634 artinya data sudah mampu menjelaskan sebesar 86,33% dari keragaman data aslinya sehingga menurut kriteria ini dipilih tiga komponen utama karena telah melebihi 80%. Berdasarkan *scree plot* pada Gambar 2 menunjukkan titik tiga garis *scree plot* mulai melandai atau bisa dilihat dari garis horizontal putus-putus, yakni terdapat tiga titik yang berada di atas garis sehingga dipilih komponen utama sebanyak tiga. Ketiga kriteria pemilihan menunjukkan hasil yang sama, yaitu tiga komponen



utama sehingga tiga komponen utama tersebut yang akan digunakan dalam analisis *cluster* menggunakan k-means.



**Gambar 3** Plot Penentuan  $k$  Optimal Indikator Pendidikan



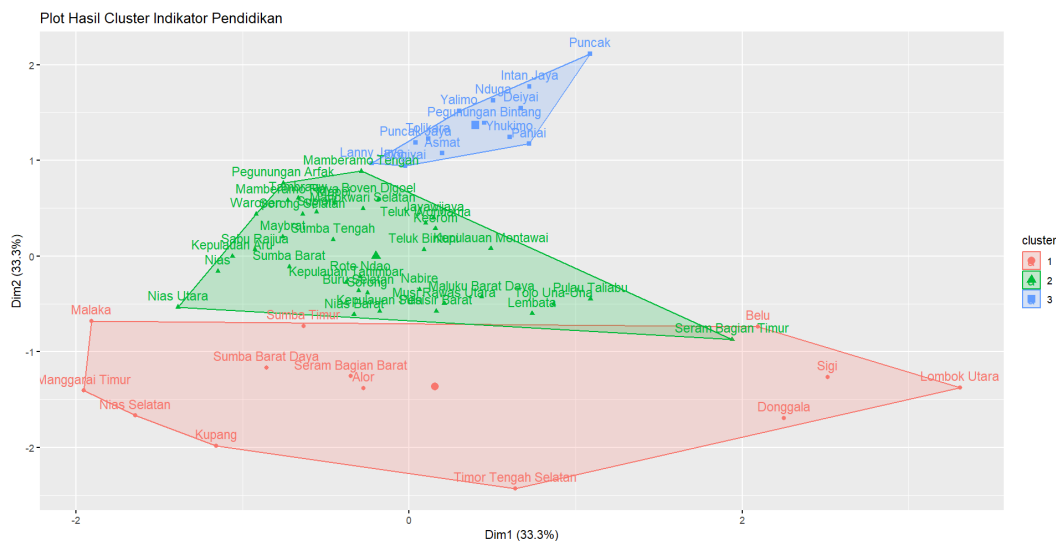
**Gambar 4** Plot Penentuan  $k$  Optimal Indikator Kesehatan

Penentuan  $k$  optimal dapat dilakukan dengan metode *elbow*, metode *gap statistic*, dan metode *silhouette* (Yuan & Yang, 2019). Metode *elbow* dijelaskan bahwa untuk menentukan  $k$  optimal dilihat pada persentase grafik yang membentuk sudut siku (Lashiyanti et al., 2023). Metode *gap statistic* dilihat dari garis yang dihubungkan dengan garis putus-putus (Zulyani et al., 2023). Metode *silhouette* menunjukkan nilai yang paling baik dilihat pada presentasi garis yang paling tinggi (Lashiyanti et al., 2023). Gambar 3 menunjukkan plot penentuan  $k$  optimal pada data indikator pendidikan, yaitu nilai  $k$  optimal pada metode *elbow* menunjukkan tiga *cluster*, metode *gap statistic* menunjukkan satu *cluster*, sedangkan metode *silhouette* menunjukkan tiga *cluster*. Dua metode menunjukkan hasil yang sama sehingga tiga dipilih sebagai banyaknya *cluster* optimal. Gambar 4 menunjukkan plot penentuan  $k$  optimal pada data indikator kesehatan, yaitu pada metode *elbow* menunjukkan dua *cluster*, metode *gap statistic* menunjukkan satu *cluster*, sedangkan metode *silhouette* menunjukkan dua *cluster*. Dua metode menunjukkan hasil yang sama sehingga dua dipilih sebagai banyaknya *cluster* optimal.

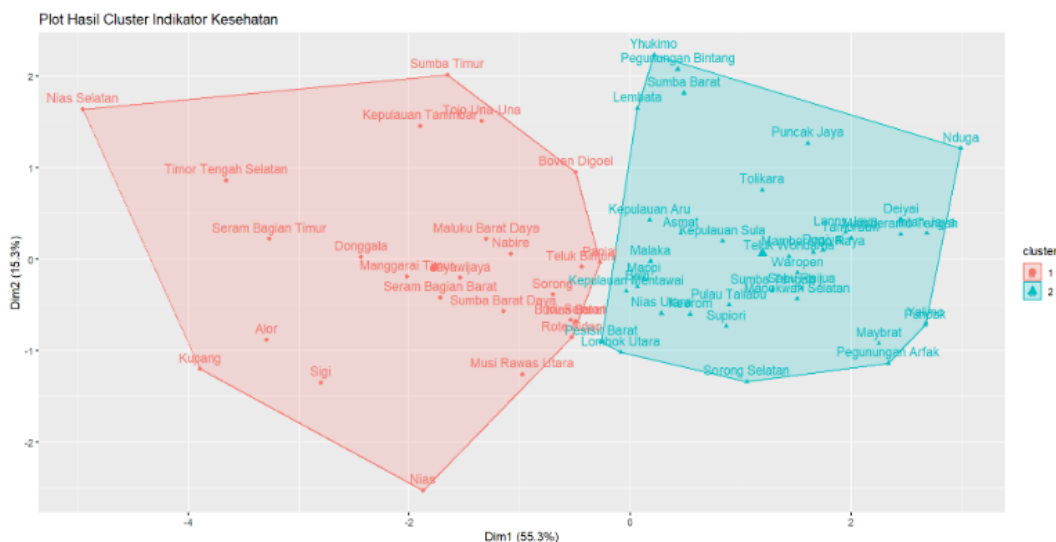




Bholowalia et al. (Faujia & Subarkah, 2022) menjelaskan bahwa algoritma k-means diawali dengan penentuan jumlah *cluster* yang akan terbentuk dan mendefinisikan nilai inisiasi *centroid* (pusat *cluster*). Dengan menggunakan jarak terdekat dari *centroid*, k-means mengelompokkan data hanya pada satu *cluster* sehingga satu data tidak akan menempati dua *cluster* atau lebih. Hasil pembentukan k-means divisualisasikan menggunakan Rstudio yang dapat dilihat pada Gambar 5 dan 6. Gambar 5 menunjukkan hasil visualisasi *cluster* pendidikan yang mengelompokkan 62 kabupaten tertinggal ke dalam tiga *cluster*, yakni *cluster* 1 terdiri dari 13 kabupaten, *cluster* 2 terdiri dari 36 kabupaten, dan *cluster* 3 terdiri dari 13 kabupaten. Gambar 6 menunjukkan hasil visualisasi *cluster* kesehatan yang mengelompokkan 62 kabupaten tertinggal ke dalam dua *cluster*, yakni *cluster* 1 terdiri dari 25 kabupaten dan *cluster* 2 terdiri dari 37 kabupaten.



Gambar 5 Visualisasi *Cluster* Pendidikan



Gambar 6 Visualisasi *Cluster* Kesehatan

Tahap validasi *cluster* yang telah terbentuk dilakukan menggunakan Silhouette Index (SI). Silhouette Index akan mengevaluasi penempatan setiap objek dalam setiap *cluster* dengan membandingkan jarak rata-rata objek dalam satu *cluster* dan jarak antara objek dengan *cluster* berbeda. Semakin besar nilai koefisien *silhouette* akan semakin baik kualitas suatu kelompok



(Nahdliyah et al., 2019). Hasil validasi *cluster* menggunakan *silhouette index* dapat dilihat pada Tabel 2. Hasil pengelompokan terbaik data pendidikan menunjukkan 3 *cluster* dengan nilai *silhouette index* sebesar 0,3696. Hal ini sesuai dengan penentuan banyaknya *k* optimal menggunakan metode *elbow* dan metode *silhouette* pada tahap sebelumnya sehingga pengelompokan indikator pendidikan ke dalam 3 *cluster* memiliki kualitas yang baik dan kuat. Namun, terdapat perbedaan pada indikator kesehatan yang mana pada validasi *silhouette index* menunjukkan banyaknya *cluster* terbaik 4, sedangkan dengan menggunakan metode *elbow* dan metode *silhouette* pada tahap sebelumnya menunjukkan 2 *cluster*. Namun demikian, pengelompokan 2 *cluster* memiliki nilai *silhouette index* sebesar 0,2351 menduduki urutan ke dua setelah pengelompokan 4 *cluster* yang memiliki nilai *silhouette index* sebesar 0,2491. Dengan demikian, pengelompokan 2 *cluster* pada indikator kesehatan juga memiliki kualitas pengelompokan yang baik dan kuat.

**Tabel 2 Validasi *Cluster* Pendidikan dan Kesehatan**

Banyak Cluster	Indikator Pendidikan	Indikator Kesehatan
2	0,3578	0,2351
3	0,3696	0,2009
4	0,3293	0,2491
5	0,3437	0,1733

Karakteristik masing-masing *cluster* ditandai dengan besaran rata-rata tiap *cluster* sehingga diketahui kategori tiap *cluster*-nya. Hasil rata-rata tiap *cluster* ditunjukkan pada Tabel 3 untuk *cluster* pendidikan dan Tabel 4 untuk *cluster* kesehatan. Dengan demikian diperoleh rincian sebagai berikut.

- 1) Hasil analisis *cluster* pendidikan menggunakan metode k-means dengan tiga *cluster* diperoleh hasil sebagai berikut:
  - a) *Cluster* 1 terdiri dari 13 kabupaten, yaitu Nias Selatan, Lombok Utara, Sumba Timur, Kupang, Timor Tengah Selatan, Belu, Alor, Sumba Barat Daya, Manggarai Timur, Malaka, Donggala, Sigi, dan Seram Bagian Barat. Anggota *cluster* 1 memiliki indikator pendidikan yang lebih tinggi atau memadai.
  - b) *Cluster* 2 terdiri dari 36 kabupaten, yaitu Nias, Nias Utara, Nias Barat, Kepulauan Mentawai, Musi Rawas Utara, Pesisir Barat, Sumba Barat, Lembata, Rote Ndao, Sumba Tengah, Sabu Raijua, Tojo Una-Una, Kepulauan Tanimbar, Kepulauan Aru, Seram Bagian Timur, Maluku Barat Daya, Buru Selatan, Kepulauan Sula, Pulau Taliabu, Teluk Wondama, Teluk Bintuni, Sorong Selatan, Sorong, Tambrauw, Maybrat, Manokwari Selatan, Pegunungan Arfak, Jayawijaya, Nabire, Boven Digoel, Mappi, Keerom, Waropen, Supiori, Mamberamo Raya, dan Mamberamo Tengah. Anggota *cluster* 2 memiliki indikator pendidikan yang sedang atau cukup memadai.
  - c) *Cluster* 3 terdiri dari 13 kabupaten, yaitu Paniai, Puncak Jaya, Asmat, Yahukimo, Pegunungan Bintang, Tolikara, Nduga, Lanny Jaya, Yalimo, Puncak, Dogiyai, Intan Jaya, dan Deiyai. Anggota *cluster* 3 memiliki indikator pendidikan yang rendah atau terbatas sehingga perlu perhatian khusus dari pemerintah.
- 2) Hasil analisis *cluster* kesehatan menggunakan metode k-means dengan dua *cluster* diperoleh hasil sebagai berikut:
  - a) *Cluster* 1 terdiri dari 25 kabupaten, yaitu Nias, Nias Selatan, Nias Barat, Musi Rawas Utara, Sumba Timur, Kupang, Timor Tengah Selatan, Alor, Rote Ndao, Sumba Barat Daya, Manggarai Timur, Donggala, Tojo Una-Una, Sigi, Kepulauan Tanimbar, Seram Bagian Barat, Seram Bagian Timur, Maluku Barat Daya, Buru Selatan, Teluk Bintuni, Sorong, Jayawijaya, Nabire, Paniai, Boven Digoel. Anggota *cluster* 1 memiliki jumlah sarana prasarana dan tenaga medis yang memadai sehingga memiliki indikator kesehatan yang tinggi.
  - b) *Cluster* 2 terdiri dari 37 kabupaten, yaitu Nias Utara, Kepulauan Mentawai, Pesisir Barat, Lombok Utara, Sumba Barat, Belu, Lembata, Sumba Tengah, Sabu Raijua, Malaka, Kepulauan Aru, Kepulauan Sula, Pulau Taliabu, Teluk Wondama, Sorong Selatan, Tambrauw, Maybrat, Manokwari Selatan, Pegunungan Arfak, Puncak Jaya, Mappi, Asmat, Yahukimo, Pegunungan Bintang, Tolikara, Keerom, Waropen,





Supiori, Mamberamo Raya, Nduga, Lannya Jaya, Mamberamo Tengah, Yalimo, Puncak, Dogiyai, Intan Jaya, Deiyai. Anggota *cluster* 2 memiliki jumlah sarana prasarana dan tenaga medis yang terbatas sehingga memiliki indikator kesehatan yang rendah.

**Tabel 3 Karakteristik *Cluster* Pendidikan**

Variabel	Cluster 1	Cluster 2	Cluster 3
Jumlah Sekolah TK/RA	126,615385	52,444444	8,923077
Jumlah Sekolah SD/MI	297,8462	110,3056	71,0000
Jumlah Sekolah SMP/MTs	117,53846	40,52778	17,15385
Jumlah Sekolah SMA/SMK/MA	62,15385	19,86111	6,00000
Jumlah Guru TK/RA	385,53846	162,83333	28,69231
Jumlah Guru SD/MI	1949,9231	1001,3611	394,3846
Jumlah Guru SMP/MTs	1778,6923	584,7778	181,4615
Jumlah Guru SMA/SMK/MA	1414,69231	391,50000	93,23077
APM SD/MI (%)	94,54077	94,21778	68,21308
APM SMP/MTs (%)	70,65769	70,94861	48,34615
APM SMA/SMK/MA (%)	56,00000	56,92278	28,39769
APK SD/MI (%)	111,07769	109,57000	81,60308
APK SMP/MTs (%)	89,78077	93,45306	69,65769
APK SMA/SMK/MA (%)	81,35000	87,29694	46,66231

**Tabel 4 Karakteristik *Cluster* Kesehatan**

Variabel	Cluster 1	Cluster 2
Jumlah Dokter	48,88000	19,72973
Jumlah Perawat	475,880	174,973
Jumlah Bidan	370,5600	119,1351
Jumlah Rumah Sakit Umum	1,440000	1,054054
Jumlah Puskesmas	21,52000	13,43243

### 3.2 Analisis MANOVA

Setelah dilakukan analisis *cluster*, selanjutnya akan dilakukan analisis MANOVA dengan tujuan menguji apakah terdapat kesesuaian pengaruh dari hasil *cluster* yang terbentuk terhadap nilai IPM, RLS, HLS, dan UHH. Namun, sebelum melakukan analisis MANOVA harus memenuhi syarat uji asumsi MANOVA. Terdapat dua asumsi yang harus dipenuhi, yaitu normalitas multivariat dan homoskedastisitas (Hamidah et al., 2022). Pengujian asumsi ini dilakukan untuk memastikan bahwa MANOVA dapat digunakan dan menghasilkan keputusan yang tepat (Sutrisno & Wulandari, 2018).

a) Uji normalitas multivariat dengan hipotesis sebagai berikut:

$H_0$  : Data berdistribusi normal multivariat

$H_1$  : Data tidak berdistribusi normal multivariat

Pengujian normalitas multivariat dilakukan menggunakan uji Mardia menunjukkan  $p - value > \alpha(0,05)$  pada pengujian hasil *cluster* pendidikan terhadap IPM, RLS, dan HLS, serta pengujian hasil *cluster* kesehatan terhadap IPM dan UHH. Dengan demikian, keduanya memiliki keputusan terima  $H_0$  sehingga data berdistribusi normal multivariat. Oleh karena itu, uji asumsi normalitas multivariat terpenuhi.

b) Uji homoskedastisitas dengan hipotesis sebagai berikut:

$H_0$  : Data bersifat homoskedastisitas

$H_1$  : Data tidak bersifat homoskedastisitas

Pengujian homoskedastisitas dilakukan menggunakan uji Levene menunjukkan  $p - value(0,3332) > \alpha(0,05)$  pada pengujian hasil *cluster* pendidikan terhadap IPM, RLS, dan HLS, sedangkan pada pengujian hasil *cluster* kesehatan terhadap IPM dan UHH memiliki  $p - value(0,03473) < \alpha(0,05)$ . Dengan demikian, data pendidikan memiliki keputusan terima  $H_0$  sehingga data bersifat homoskedastisitas, tetapi data kesehatan memiliki keputusan tolak  $H_0$ .



sehingga data tidak bersifat homoskedastisitas. Menurut Hamidah et al. (2022), jika terdapat uji asumsi yang tidak terpenuhi, maka dapat menggunakan uji statistik Wilk's lambda karena kriteria pada uji tersebut dianggap lebih *robust* dibanding kriteria pada uji lainnya. Dengan demikian, analisis MANOVA dapat dilanjutkan.

Analisis MANOVA pada data pendidikan menggunakan empat uji statistik, yaitu Pillai's trace, Wilk's lambda, Hotelling trace, dan Roy's largest root, sedangkan data kesehatan menggunakan uji statistik Wilk's lambda karena terdapat salah satu uji asumsi yang tidak terpenuhi. Hipotesis penelitian MANOVA yang digunakan sebagai berikut.

$H_0$ : Hasil *cluster* tidak berpengaruh signifikan secara multivariat

$H_1$ : Hasil *cluster* berpengaruh signifikan secara multivariat

Tabel 5 menunjukkan uji statistik pengaruh *cluster* pendidikan terhadap nilai IPM, RLS, dan HLS. Hasil uji statistik menyatakan nilai  $p - value < \alpha(0,05)$  yang berarti tolak  $H_0$ . Dengan demikian, dapat disimpulkan bahwa terdapat kesesuaian pengaruh yang signifikan antara hasil *cluster* (*cluster* 1, *cluster* 2, dan *cluster* 3) terhadap nilai IPM, RLS, dan HLS. *Cluster* dengan kategori tinggi memiliki nilai IPM, RLS, dan HLS yang tinggi juga, begitupun sebaliknya.

Tabel 6 menunjukkan uji statistik pengaruh *cluster* kesehatan terhadap nilai IPM dan UHH. Hasil uji *Wilk's lambda* menunjukkan  $p - value < \alpha(0,05)$  yang berarti tolak  $H_0$ . Dengan demikian, dapat disimpulkan bahwa terdapat pengaruh yang signifikan antara hasil *cluster* (*cluster* 1 dan *cluster* 2) terhadap nilai IPM dan UHH. *Cluster* dengan kategori tinggi memiliki nilai IPM dan UHH yang tinggi juga, begitupun sebaliknya.

Hasil *cluster* pendidikan dan kesehatan memang secara nyata berpengaruh pada IPM, RLS, HLS, dan UHH. Antara lain ditunjukkan pada Kabupaten Nias Selatan yang tergolong kategori tinggi pada indikator pendidikan dan kesehatan. Kabupaten Nias Selatan memiliki nilai IPM yaitu 63,17, nilai RLS 6,23, nilai HLS 12,48, dan nilai UHH 69,21. Dibandingkan dengan Kabupaten Puncak Jaya yang tergolong dalam kategori rendah di kedua indikator tersebut memiliki nilai IPM 49,84, nilai RLS 4,03, nilai HLS 7,50, dan nilai UHH 65,66. Terbukti bahwa kabupaten dengan *cluster* tinggi akan memiliki nilai IPM, RLS, HLS, dan UHH yang tinggi dibandingkan dengan *cluster* rendah. Dengan demikian kabupaten-kabupaten yang termasuk dalam kategori rendah, baik indikator pendidikan maupun kesehatan harus mendapatkan perhatian lebih dari pemerintah agar tercipta peningkatan kesejahteraan masyarakat.

**Tabel 5 Uji Statistik MANOVA *Cluster* Pendidikan**

Uji Statistik	p-value	Keputusan	Keterangan
Pillai's trace	1,367e-09	Tolak $H_0$	Berbeda signifikan
Wilk's lambda	1,367e-09	Tolak $H_0$	Berbeda signifikan
Hotelling trace	1,367e-09	Tolak $H_0$	Berbeda signifikan
Roy	1,367e-09	Tolak $H_0$	Berbeda signifikan

**Tabel 6 Uji Statistik MANOVA *Cluster* Kesehatan**

Uji Statistik	p-value	Keputusan	Keterangan
Pillai's trace	0,000282	Tolak $H_0$	Berbeda signifikan

#### 4. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan dapat disimpulkan bahwa algoritma k-means dapat dilakukan untuk mengelompokkan kabupaten tertinggal berdasarkan data indikator pendidikan dan kesehatan. Pengelompokan 62 kabupaten tertinggal berdasarkan indikator pendidikan terbagi menjadi tiga *cluster* dengan nilai *silhouette index* sebesar 0,3696. *Cluster* 1 mencakup 13 kabupaten dengan sarana prasarana dan tenaga pendidik yang memadai sehingga memiliki indikator pendidikan yang tinggi. *Cluster* 2 mencakup 36 kabupaten dengan jumlah sarana prasarana dan tenaga pendidik yang cukup memadai sehingga memiliki indikator



pendidikan yang sedang. *Cluster* 3 mencakup 13 kabupaten dengan sarana prasarana dan tenaga pendidik yang terbatas sehingga memiliki indikator pendidikan yang rendah. Hasil analisis MANOVA menunjukkan adanya pengaruh hasil *cluster* pendidikan terhadap IPM, RLS, dan HLS.

Sementara indikator kesehatan terbagi ke dalam dua *cluster dengan nilai silhouette index sebesar 0,2351*. *Cluster* 1 mencakup 25 kabupaten dengan jumlah sarana prasarana dan tenaga medis yang memadai sehingga memiliki indikator kesehatan yang tinggi. *Cluster* 2 mencakup 37 kabupaten dengan jumlah sarana prasarana kesehatan dan tenaga medis yang terbatas sehingga memiliki indikator kesehatan yang rendah. Hasil analisis MANOVA menunjukkan adanya pengaruh hasil *cluster* kesehatan terhadap IPM dan UHH. Dengan kata lain masing-masing pertumbuhan bidang pendidikan dan kesehatan memiliki pengaruh yang signifikan terhadap Indeks Pembangunan Manusia, Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), dan Umur Harapan Hidup (UHH). Hasil penelitian ini sejalan dengan penelitian yang dilakukan oleh Hasibuan et al. (2023).

## DAFTAR PUSTAKA

- Amelia, D., & Kholijah, G. (2023). Analisis Cluster Pengelompokan Provinsi di Indonesia Berdasarkan Sub Sektor Nilai Tukar Petani. *DEMOS: Journal of Demography, Ethnography and Social Transformation*, 3(1), 1–12. <https://doi.org/10.30631/demos.v3i1.1812>
- Amrullah, A., Purnamasari, I., Sari, B. N., Garno, G., & Voutama, A. (2022). Analisis Cluster Faktor Penunjang Pendidikan Menggunakan Algoritma K-Means (Studi Kasus: Kabupaten Karawang). *Jurnal Informatika dan Rekayasa Elektronik*, 5(2), 244–252. <https://doi.org/10.36595/jire.v5i2.701>
- Aryanto, R. P., Nilogiri, A., & Wardoyo, A. E. (2024). Klasterisasi Jumlah Penduduk Provinsi Jawa Timur Tahun 2021-2023 Menggunakan Algoritma K-Means. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 9(2), 134–146. <https://doi.org/10.14421/jiska.2024.9.2.134-146>
- Fajar, M. A., & Indrawati, L. (2020). Pengaruh Belanja Pendidikan, Belanja Kesehatan dan Belanja Perumahan dan Fasilitas Umum Terhadap Indeks Pembangunan Manusia (Studi Kasus pada Pemerintah Daerah Kabupaten Cianjur). *Indonesian Accounting Research Journal*, 1(1), 108–118. <https://jurnal.polban.ac.id/iarj/article/view/2366>
- Faujia, R. A., & Subarkah, M. Z. (2022). Analisis Klaster K-Means dan Visualisasi Data Spasial Berdasarkan Karakteristik Persebaran Covid-19 dan Pelanggaran Protokol Kesehatan di Jawa Tengah. *Seminar Nasional Official Statistics*, 2022(1), 813–822. <https://doi.org/10.34123/semnasoffstat.v2022i1.1222>
- Hamidah, N., Santoso, R., & Rusgiyono, A. (2022). Klasterisasi Provinsi di Indonesia Berdasarkan Faktor Penyebaran Covid-19 Menggunakan Model-Based Clustering t-Multivariat. *Jurnal Gaussian*, 11(1), 56–66. <https://doi.org/10.14710/j.gauss.v11i1.33999>
- Hasibuan, S. R., Harahap, I., & Tambunan, K. (2023). Pengaruh Pertumbuhan Ekonomi, Pendidikan dan Kesehatan Terhadap Indeks Pembangunan Manusia di Provinsi Sumatera Utara. *Jurnal Manajemen Akuntansi (JUMSI)*, 3(1), 272–285. <https://doi.org/10.36987/jumsi.v3i1.4023>
- Lashiyanti, A. R., Munthe, I. R., & Nasution, F. A. (2023). Optimisasi Klasterisasi Nilai Ujian Nasional dengan Pendekatan Algoritma. *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, 6(1), 14–20. <https://ejournal.sisfokomtek.org/index.php/jikom/article/view/1550>
- Lenama, N., Kleden, M. A., & Pasangka, I. G. (2023). K-Means Clustering Analysis pada Pengelompokan Kabupaten/Kota di Provinsi Nusa Tenggara Timur Berdasarkan Indikator Pendidikan. *Jurnal Cakrawala Ilmiah*, 2(9), 3365–3376. <https://doi.org/10.53625/jcijurnalcakrawalailmiah.v2i9.5653>
- Mariana, M. (2013). Analisis Komponen Utama. *Jurnal Matematika dan Pembelajarannya*, 1(2), 99–114. <https://doi.org/10.33477/mp.v1i2.304>
- Nahdliyah, M. A., Widiari, T., & Prahutama, A. (2019). Metode K-Medoids Clustering dengan Validasi Silhouette Index dan C-Index (Studi Kasus Jumlah Kriminalitas Kabupaten/Kota di Jawa Tengah Tahun 2018). *Jurnal Gaussian*, 8(2), 161–170. <https://doi.org/10.14710/j.gauss.8.2.161-170>



- Octaviyani, N. R., Mayasari, R., & Susilawati, S. (2022). Implementasi Algoritma K-Means Clustering Status Gizi Balita. *Jurnal Ilmiah Wahana Pendidikan*, 2022(13), 370–381. <https://doi.org/10.5281/zenodo.6962588>
- Peraturan Presiden (Perpres) Nomor 63 Tahun 2020 Tentang Penetapan Daerah Tertinggal Tahun 2020-2024, Pub. L. No. 63, Pemerintah Pusat Indonesia (2020).
- Prasetyawati, M. D. (2023). *Statistik Daerah Provinsi Jawa Tengah 2023* (D. Nursetyohadi & M. Samuharwadi, Eds.). Badan Pusat Statistik Provinsi Jawa Tengah.
- Pursitasari, I. D., Harianto, B., Wulan, S. S., Hermanto, D., & Ardianto, D. (2024). Multivariate Analysis of Variance (MANOVA) di Bidang Kesehatan dan Pendidikan MIPA. *Jurnal Ilmiah Kanderang Tingang*, 15(1), 117–126. <https://doi.org/10.37304/jikt.v15i1.307>
- Setiawan, I. N., Kristiani, S. N., Nurarifin, N., Delyana, S., Setyawati, N., & Arsyi, F. A. (2022). *Indeks Pembangunan Manusia 2022* (W. Winardi, Y. Karyono, M. Mutijo, & D. H. Santoso, Eds.). Badan Pusat Statistik.
- Sitinjak, D. K., Pangestu, B. A., & Sari, B. N. (2022). Clustering Tenaga Kesehatan Berdasarkan Kecamatan di Kabupaten Karawang Menggunakan Algoritma K-Means. *Journal of Applied Informatics and Computing*, 6(1), 47–54. <https://doi.org/10.30871/jaic.v6i1.3855>
- Sutrisno, S., & Wulandari, D. (2018). Multivariate Analysis of Variance (MANOVA) untuk Memperkaya Hasil Penelitian Pendidikan. *AKSIOMA: Jurnal Matematika dan Pendidikan Matematika*, 9(1), 37. <https://doi.org/10.26877/aks.v9i1.2472>
- Thamrin, D. R., & Murni, D. (2022). Analisis Cluster Hierarki Metode Single Linkage pada Kabupaten/Kota di Provinsi Sumatera Barat Berdasarkan Indikator Kesehatan. *Journal of Mathematics UNP*, 7(3), 45–51. <https://doi.org/10.24036/unpjomath.v7i3.12988>
- Wijaya, T. A., Utami, E., & Fatta, H. (2024). Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Data Gangguan PT. PLN UID Kalselteng. *Innovative: Journal of Social Science Research*, 4(1), 8846–8854. <https://doi.org/10.31004/innovative.v4i1.8920>
- Yuan, C., & Yang, H. (2019). Research on K-Value Selection Method of K-Means Clustering Algorithm. *J—Multidisciplinary Scientific Journal*, 2(2), 226–235. <https://doi.org/10.3390/j2020016>
- Yudhistira, A., & Andika, R. (2023). Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering. *Journal of Artificial Intelligence and Technology Information (JAITI)*, 1(1), 20–28. <https://doi.org/10.58602/jaiti.v1i1.22>
- Zulyani, A. A., Irawan, A. S. Y., & Jamaludin, A. (2023). Penerapan Data Mining Menggunakan Algoritma K-Means untuk Menentukan Tingkat Vaksinasi pada Kecamatan Tambun Selatan. *Innovative: Journal of Social Science Research*, 3(3), 7037–7050. <https://j-innovative.org/index.php/Innovative/article/view/2946>



LAMPIRAN A

Tabel 7 Nilai VIF Antar Variabel Pendidikan

Nilai VIF	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>	X <sub>10</sub>	X <sub>11</sub>	X <sub>12</sub>	X <sub>13</sub>	X <sub>14</sub>
X <sub>1</sub>		12,430	12,895	12,990	2,106	12,645	12,983	13,152	12,705	13,157	13,136	13,112	12,981	13,168
X <sub>2</sub>	24,883		16,750	25,037	24,894	9,430	21,500	25,642	26,376	25,017	26,341	26,367	26,210	26,325
X <sub>3</sub>	30,893	20,046		21,368	30,613	27,883	20,616	31,219	31,282	30,662	29,718	31,550	31,308	31,149
X <sub>4</sub>	21,170	20,384	14,536		21,418	20,903	20,108	15,436	20,969	21,106	20,446	21,368	21,349	21,307
X <sub>5</sub>	2,086	12,318	12,657	13,017		12,241	12,933	12,605	12,684	12,886	12,981	12,982	13,004	13,002
X <sub>6</sub>	43,926	16,364	40,430	44,544	42,928		36,050	39,950	45,725	43,871	45,735	45,715	45,347	45,634
X <sub>7</sub>	20,886	17,278	13,844	19,848	21,005	16,695		19,389	21,046	21,174	21,197	21,087	20,832	20,982
X <sub>8</sub>	21,112	20,562	20,917	15,203	20,427	18,461	19,346		21,146	21,019	21,107	20,762	21,070	21,124
X <sub>9</sub>	9,781	10,144	10,052	9,905	9,859	10,134	10,071	10,142		8,557	10,064	3,884	9,781	10,008
X <sub>10</sub>	7,106	6,750	6,913	6,995	7,027	6,821	7,109	7,073	6,004		5,943	6,424	4,597	7,069
X <sub>11</sub>	7,441	7,072	7,027	7,106	7,424	7,458	7,464	7,449	7,405	6,233		7,381	7,299	3,968
X <sub>12</sub>	6,672	6,702	6,701	6,671	6,669	6,696	6,669	6,581	2,567	6,052	6,629		5,807	6,084
X <sub>13</sub>	3,026	3,052	3,046	3,054	3,060	3,043	3,019	3,060	2,962	1,984	3,004	2,660		3,066
X <sub>14</sub>	5,577	5,577	5,507	5,537	5,560	5,564	5,524	5,574	5,506	5,543	2,967	5,501	5,572	

Tabel 8 Nilai Eigen dan Proporsi Kumulatif Data Pendidikan

Komponen Utama	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14
Nilai Eigen	8,0097	2,9519	1,1257	0,7090	0,5077	0,1934	0,1373	0,1040	0,0670	0,0637	0,0564	0,0399	0,0226	0,1116
Proporsi Kumulatif	0.5721	0.7830	0.8634	0.9140	0.9503	0.9641	0.9739	0.9814	0.9862	0.9907	0.9947	0.9976	0.9992	1.0000

