



Pemodelan Ketimpangan Energi dan Kemiskinan di Indonesia Menggunakan *Hybrid* K-Means–PCA dalam Mendukung SDGs 1 dan 7

Carissa Azarine Henry^{a,*}, Nadine Lenora^b, Rika Pramestia Putri^c, Iqbal Kharisudin^d, Virgania Sari^c

^aUniversitas Negeri Semarang, Jl. Sekaran, Kota Semarang, 50229, Indonesia

* Alamat Surel: carissaazarine36@students.unnes.ac.id

Abstrak

Energi bersih dan pengentasan kemiskinan merupakan fokus utama SDG 1 dan 7 yang berperan penting dalam mendukung pembangunan berkelanjutan di Indonesia. Penelitian ini bertujuan untuk memodelkan ketimpangan akses energi dan tingkat kemiskinan antarprovinsi di Indonesia menggunakan pendekatan hybrid antara *Principal Component Analysis* (PCA) dan algoritma *K-Means Clustering* yang dilakukan menggunakan perangkat lunak R. Data penelitian bersumber dari Badan Pusat Statistik (BPS) dan Kementerian Energi dan Sumber Daya Mineral (ESDM) tahun 2024, mencakup enam variabel yang merepresentasikan aspek kemiskinan dan energi listrik. Hasil PCA menunjukkan bahwa dua komponen utama yang mampu menjelaskan 71.2% variasi data. PC1 merepresentasikan dimensi kesejahteraan sosial dan akses energi, sedangkan PC2 menggambarkan dimensi kapasitas dan pemerataan energi. Hasil penelitian juga menunjukkan bahwa tingkat dukungan terhadap pencapaian SDGs 1 (PC1) memiliki nilai lebih tinggi dengan rerata nilai sebesar 0.40 dibandingkan dengan tingkat dukungan terhadap pencapaian SDGs 7 dengan nilai rerata sebesar 0.34. Tahap *K-Means Clustering* menghasilkan tiga klaster dengan jumlah masing-masing 4, 6, dan 28 provinsi. Klaster 1 memiliki nilai PC1 = -3.78 dan PC2 = -1.18, menunjukkan wilayah dengan kemiskinan tinggi dan akses energi listrik rendah. Klaster 2 menunjukkan PC1 = 2.16 dan PC2 = -1.89, merepresentasikan provinsi dengan kesejahteraan tinggi dan kapasitas energi tinggi. Sementara Klaster 3 memiliki PC1 = 0.08 dan PC2 = 0.57, menggambarkan wilayah dengan kondisi sosial-ekonomi menengah dan distribusi energi yang relatif efisien. Temuan ini menegaskan adanya ketimpangan spasial antarprovinsi di Indonesia dalam hal kesejahteraan sosial dan kapasitas energi listrik. Hasil penelitian ini dapat menjadi dasar perumusan kebijakan pemerataan energi dan pengentasan kemiskinan di Indonesia.

Kata kunci:

SDGs 1, SDGs 7, PCA, K-Means, Ketenagalistrikan, Kemiskinan.

© 2026 Dipublikasikan oleh Jurusan Matematika, Universitas Negeri Semarang

1. Pendahuluan

Energi merupakan pendorong utama pembangunan ekonomi dan sosial yang berkelanjutan. Akses terhadap energi modern dan bersih memiliki peran penting dalam meningkatkan kualitas hidup masyarakat, produktivitas ekonomi, serta pencapaian kesejahteraan yang merata (Pokhriyal et al., 2022). Laporan *Tracking SDG7: Energy Progress Report 2024* menyebutkan bahwa sekitar 675 juta penduduk dunia masih hidup tanpa akses listrik dan 2,1 miliar orang belum memiliki akses terhadap bahan bakar memasak bersih. Kondisi ini mengindikasikan bahwa pencapaian *Sustainable Development Goals* (SDG) ke-7, yaitu “akses energi yang terjangkau, andal, berkelanjutan, dan modern untuk semua”, masih menghadapi tantangan serius terutama di negara berkembang seperti Indonesia (IEA, IRENA, UNSD, World Bank, 2024).

To cite this article:

Henry, C.A., Lenora, N., & Putri, R.P. (2026). Pemodelan Ketimpangan Energi dan Kemiskinan di Indonesia Menggunakan *Hybrid K-Means-PCA* dalam Mendukung SDGs 1 dan 7. *PRISMA, Prosiding Seminar Nasional Matematika* 9, 392-401

Hubungan antara akses energi dan kemiskinan telah menjadi fokus berbagai penelitian global. Akses listrik terbukti berkontribusi terhadap peningkatan pendapatan, perluasan kesempatan kerja, dan kesejahteraan masyarakat (Asghar et al., 2022). Studi empiris di Indonesia menunjukkan bahwa elektrifikasi berdampak signifikan terhadap penurunan kemiskinan melalui peningkatan pengeluaran non-pangan dan efisiensi ekonomi rumah tangga (Ikhsan & Amri, 2022). Selain itu, akses energi juga meningkatkan kebahagiaan dan kesejahteraan subjektif masyarakat di daerah tertinggal. (Ria Jayanthi, 2021) menegaskan bahwa pembangunan infrastruktur kelistrikan di Indonesia turut mendorong pertumbuhan ekonomi dan pengurangan kemiskinan, meskipun manfaatnya belum terdistribusi merata di seluruh provinsi. Temuan-temuan tersebut memperkuat relevansi integrasi antara SDG 1 (*No Poverty*) dan SDG 7 (*Affordable and Clean Energy*) dalam kebijakan pembangunan nasional.

Selain itu, peningkatan kualitas pembangunan manusia juga berperan penting dalam menekan tingkat kemiskinan. (Pardita et al., 2025) menunjukkan bahwa peningkatan IPM di Indonesia secara signifikan menurunkan tingkat kemiskinan melalui perbaikan akses pendidikan, kesehatan, dan standar hidup. Oleh karena itu, peningkatan IPM berarti meningkatnya kualitas hidup yang secara langsung menekan angka kemiskinan. Di sisi lain, pengangguran terbuka memiliki hubungan positif dengan kemiskinan, sehingga penciptaan lapangan kerja produktif menjadi faktor penting dalam mempercepat pengentasan kemiskinan di tingkat provinsi (Hasiholan et al., 2022).

Meskipun rasio elektrifikasi nasional Indonesia telah mencapai lebih dari 99% pada tahun 2024 (Direktorat Jenderal Ketenagalistrikan, 2025), kesenjangan spasial akses energi antara wilayah barat dan timur masih cukup tinggi. Multidimensional energi *poverty* di Indonesia tetap signifikan, terutama di wilayah pedesaan dan kepulauan timur yang memiliki keterbatasan infrastruktur dan keterjangkauan energi (Nur et al., 2024). Ketimpangan ini memperlihatkan adanya korelasi dua arah antara kemiskinan dan kekurangan energi, di mana wilayah dengan kemiskinan tinggi cenderung memiliki tingkat akses energi yang rendah, dan sebaliknya.

Dalam konteks tersebut, analisis berbasis data menjadi sangat penting untuk mengidentifikasi pola ketimpangan energi dan kemiskinan di tingkat provinsi. Pendekatan statistik konvensional sering kali terbatas ketika berhadapan dengan data multivariat berukuran besar dan saling berkorelasi. Oleh karena itu, metode unsupervised learning seperti *Principal Component Analysis* (PCA) dan *K-Means Clustering* menjadi alternatif yang efektif untuk mengekstraksi pola dan mengelompokkan wilayah berdasarkan karakteristik yang serupa (Mohammed & Sumari, 2024). Integrasi PCA dan K-Means mampu meningkatkan efisiensi dan interpretabilitas kluster melalui reduksi dimensi dan ekstraksi komponen utama (Hassanpour et al., 2024).

Penelitian ini berfokus pada dimensi akses listrik, karena indikator ini merupakan yang paling representatif dan memiliki data lengkap hingga tingkat provinsi tahun 2024. Sementara itu, indikator lain seperti akses memasak bersih dan energi terbarukan tidak diikutsertakan karena keterbatasan ketersediaan data terbaru di tingkat provinsi. Dengan demikian, aspek kelistrikan digunakan sebagai proksi utama dari SDG 7 untuk menggambarkan kondisi nyata pemerataan energi di Indonesia. Sedangkan untuk SDG 1 direpresentasikan dengan variabel persentase penduduk miskin, indeks pembangunan manusia dan tingkat pengangguran terbuka. Berdasarkan hal tersebut, penelitian ini bertujuan untuk memodelkan ketimpangan energi dan kemiskinan di Indonesia dengan menggunakan pendekatan *Hybrid K-Means-PCA* berbasis data tahun 2024.

Tabel 1. Dataset yang Digunakan

Provinsi	Persentase Penduduk Miskin	IPM	Tingkat Pengangguran Terbuka	Daya Tersambung	Rasio Elektrifikasi	Produksi Tenaga Listrik
Aceh	12.64	74.03	5.75	2323428886	99.99	3946.11
Sumatera Utara	7.19	74.02	5.60	7868060614	99.99	16527.30
Sumatera Barat	5.42	74.49	5.75	2569109070	99.99	4600.37
Riau	6.36	74.79	3.70	7280035485	99.99	6514.38
Jambi	7.26	73.43	4.48	6775638698	99.99	2128.29
Sumatera Selatan	10.51	72.30	3.68	6775638698	99.99	13432.97
Bengkulu	12.52	73.39	3.11	6775638698	99.99	2245.61

Lampung	10.62	71.81	4.19	3913337775	99.99	53290.29
Kep. Bangka Belitung	5.08	73.33	4.63	1213180250	99.99	4878.55
Kep. Riau	4.78	77.97	6.39	7280035485	99.99	4913.51
Dki Jakarta	4.14	83.08	6.21	2241912919 1	100.00	17607.71
Jawa Barat	7.08	74.43	6.75	3369011599 8	99.99	44382.21
Jawa Tengah	9.58	73.88	4.78	2085599212 6	99.99	71068.16
Di Yogyakarta	10.4	81.55	3.48	2085599212 6	99.99	3.64
Jawa Timur	9.56	74.09	4.19	2640659133 6	99.67	46912.51
Banten	5.70	74.48	6.68	2241912919 1	99.99	17607.71
Bali	3.80	77.76	1.79	4965104831	100.00	5556.33
Nusa Tenggara Barat	11.91	70.93	2.73	2306980787	99.99	3274.09
Nusa Tenggara Timur	19.02	67.39	3.02	1715953763	96.35	1691.88
Kalimantan Barat	6.25	70.13	4.86	2493277346	99.85	4184.51
Kalimantan Tengah	5.26	72.73	4.01	3756959198	98.05	2679.67
Kalimantan Selatan	4.02	73.03	4.20	3756959198	99.99	4225.46
Kalimantan Timur	5.51	78.83	5.14	3652091352	99.99	5985.47
Kalimantan Utara	5.38	73.02	3.90	3652091352	99.99	764.87
Sulawesi Utara	6.70	75.03	5.85	3328372543	99.99	3100.47
Sulawesi Tengah	11.04	71.56	2.94	3328372543	99.99	2934.72
Sulawesi Selatan	7.77	74.05	4.19	7316240147	99.99	9798.53
Sulawesi Tenggara	10.63	73.48	3.09	7316240147	99.78	811.95
Gorontalo	13.87	71.23	3.13	3328372543	99.99	963.03
Sulawesi Barat	10.71	68.20	2.68	7316240147	99.9	4.49
Maluku	15.78	71.57	6.11	1277871150	99.08	914.56
Maluku Utara	6.03	71.03	4.03	1277871150	99.99	783.39
Papua Barat	21.09	67.02	4.13	1776328035	99.99	304.00
Papua Barat Daya	16.95	68.63	6.48	1776328035	99.99	483.96
Papua	18.09	73.00	6.48	1776328035	99.81	915.23
Papua Selatan	19.35	67.9	4.05	1776328035	94.08	244.96
Papua Tengah	27.6	59.75	2.75	1776328035	99.49	417,98
Papua Pegunungan	29.66	53.42	1.32	1776328035	94.02	53.20

2. Metode

2.1. Pengumpulan Data

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan tujuan untuk memodelkan ketimpangan akses energi bersih dan tingkat kemiskinan antarprovinsi di Indonesia dalam mencapai SDGs 1 dan SDGs 7. Data yang digunakan merupakan data sekunder yang bersumber dari Badan Pusat Statistik (BPS) dan Kementerian Energi dan Sumber Daya Mineral (ESDM) tahun 2024. Untuk merepresentasikan dimensi SDGs 1 (Tanpa Kemiskinan), penelitian ini menggunakan beberapa indikator yang dikeluarkan oleh BPS, yaitu persentase penduduk miskin, Indeks Pembangunan Manusia (IPM), dan tingkat pengangguran terbuka. Sementara itu, untuk merepresentasikan salah satu capaian SDGs 7 (Akses Energi Terjangkau dan Bersih), penelitian ini menggunakan indikator kelistrikan pada tingkat provinsi di Indonesia meliputi rasio elektrifikasi (%), produksi tenaga listrik (GWh), dan daya tersambung pelanggan (MW). Pemilihan indikator tersebut didasarkan pada ketersediaan data yang lengkap dan keterkaitannya dengan dimensi akses energi.

2.2 Pra-Pengolahan Data

Tahapan pra-pengolahan data dilakukan untuk memastikan bahwa data yang digunakan layak dan konsisten sebelum dianalisis lebih lanjut. Pertama, dilakukan penyelarasan cakupan data, di mana hanya tingkat provinsi yang dipertahankan, sedangkan baris agregat nasional seperti “Indonesia” atau “Total Indonesia” dihapus agar hasil analisis tidak bias terhadap ukuran populasi nasional. Untuk memastikan skala antar variabel seragam, seluruh variabel numerik dinormalisasi menggunakan standarisasi z-score agar variabel berbeda skala tidak mendominasi analisis.

2.3 Principal Component Analysis (PCA)

Dalam penelitian ini digunakan metode Principal Component Analysis (PCA) sebagai teknik reduksi dimensi untuk menyederhanakan data multivariat dengan tetap mempertahankan sebagian besar informasi penting. PCA mengidentifikasi kesamaan antar objek yang diteliti (Fourar et al., 2025). Setelah melakukan standarisasi setiap variabel, dilakukan analisis PCA yang menghasilkan sejumlah komponen utama yang merupakan kombinasi linier dari seluruh variabel dan saling bebas korelasi (orthogonal), sehingga mampu merepresentasikan pola variasi utama dalam data secara lebih ringkas dan efisien (Alirio et al., 2022). Komponen utama dinyatakan sebagai berikut:

$$PC_1 = \sum_{j=1}^p a_{1j} x'_j \quad (1)$$

$$PC_2 = \sum_{j=1}^p a_{2j} x'_j \quad (2)$$

$$PC_n = \sum_{j=1}^p a_{nj} x'_j \quad (3)$$

Untuk mengetahui berapa banyak komponen utama yang cukup mewakili data, digunakan persentase varians yang dijelaskan oleh:

$$\omega_i = \frac{\text{eigenvalue}_i}{\sum \text{eigenvalues}} \quad (4)$$

Nilai ini menunjukkan berapa persen informasi data asli yang dapat dipertahankan oleh setiap komponen utama.

2.4 K-Means Clustering

Tahap selanjutnya adalah klasterisasi menggunakan algoritma K-Means, dengan input berupa skor komponen utama hasil PCA. K-Means merupakan algoritma *unsupervised learning* yang membagi data ke dalam k klaster berdasarkan kemiripan karakteristik (Hassanpour et al., 2024). Dalam penelitian ini, pemilihan jumlah klaster (k) tidak hanya menggunakan metode Elbow melainkan berdasarkan koefisien Silhouette, yaitu patahan gradien terbesar yang didapat pada *Elbow chart* dan nilai rata-rata Silhouette tertinggi dalam rentang nilai k yang diuji (misalnya $k = 2$ hingga $k = 8$). Kemudian, dilakukan

tinjauan terhadap ukuran jarak dan kesamaan. Di antara ukuran jarak terdapat hal yang disebut jarak Minkowski, termasuk jarak Euclidean, jarak Manhattan atau City block, jarak Chebyshev, dan jarak Hamming menurut formulasi berikut, yang merepresentasikan jarak antara dua titik $g_i = (g_1^1, g_1^2, \dots, g_1^n)$ dan $g_2 = (g_2^1, g_2^2, g_2^n)$ dalam ruang berdimensi n dirumuskan sebagai:

$$d(g_1, g_2) = \left(\sum_{i=1}^n |g_1^i - g_2^i|^p \right)^{\frac{1}{p}} \quad (5)$$

Nilai parameter p menentukan jenis ukuran jarak yang digunakan, yaitu:

- $p = 1$ untuk Manhattan Distance (rectilinear),
- $p = 2$ untuk Euclidean Distance, dan
- $p = \infty$ untuk Chebyshev Distance.

Dalam penelitian ini, digunakan ukuran jarak Euclidean ($p = 2$), karena paling umum diterapkan pada algoritma K-Means untuk mengukur kedekatan antar objek dalam ruang multidimensi dan menghasilkan pembentukan kluster yang optimal (Alirio et al., 2022). Pendekatan ini sesuai dengan literatur yang menekankan bahwa kombinasi PCA dan K-Means dapat menghasilkan struktur kluster yang lebih stabil dan interpretatif (Ultsch, 2024). Label kluster yang terbentuk kemudian dianalisis karakteristiknya berdasarkan nilai rata-rata komponen utama pada masing-masing kluster, sehingga interpretasi yang dihasilkan merepresentasikan pola umum dari seluruh variabel awal.

2.5 Batasan Studi

Penelitian ini hanya meneliti tentang ketimpangan energi dan kemiskinan di Indonesia menggunakan *Hybrid K-Means-PCA* dalam Mendukung SDGs 1 dan 7. SDGs 1 direpresentasikan oleh dimensi persentase penduduk miskin, Indeks Pembangunan Manusia (IPM) dan tingkat pengangguran. Pada bagian lain, SDGs 7 direpresentasikan melalui dimensi akses listrik, karena indikator lain seperti akses memasak bersih, porsi energi terbarukan, efisiensi energi belum tersedia lengkap pada tingkat provinsi untuk tahun 2024.

3. Hasil dan Pembahasan

Sebelum melakukan analisis *clustering*, dilakukan uji asumsi terlebih dahulu yaitu uji kecukupan dan kelayakan data. Uji kecukupan data dilakukan dengan metode KMO, sedangkan uji kelayakan setiap variabel digunakan kriteria MSA. Uji KMO merupakan indeks perbandingan jarak antara nilai koefisien korelasi terhadap korelasi parsial, dengan pendugaan awal jumlah sampel tidak memenuhi kriteria untuk analisa lanjut.

Tabel 2. Hasil Uji KMO dan MSA

Variabel	Nilai KMO	Nilai MSA
Persentase Penduduk Miskin	0.67	0.69
IPM		0.64
Tingkat Pengangguran Terbuka		0.86
Daya Tersambung		0.59
Rasio Elektrifikasi		0.87
Produksi Tenaga Listrik		0.53

Berdasarkan **Tabel 2** diperoleh nilai KMO lebih besar dari 0.5, begitu pula nilai MSA setiap variabelnya. Oleh karena itu, dapat disimpulkan bahwa seluruh variabel masih layak untuk dilakukan analisis PCA, meskipun variabel Daya Tersambung dan Produksi Tenaga Listrik memiliki kontribusi paling kecil.

Setelah dipastikan seluruh variabel layak untuk dianalisis, dilakukan standarisasi menggunakan *Z-Score* karena setiap variabel memiliki satuan dan rentang nilai yang berbeda. Proses standarisasi ini

bertujuan untuk menyamakan bobot antar variabel sehingga data siap digunakan dalam analisis *Principal Component Analysis* (PCA).

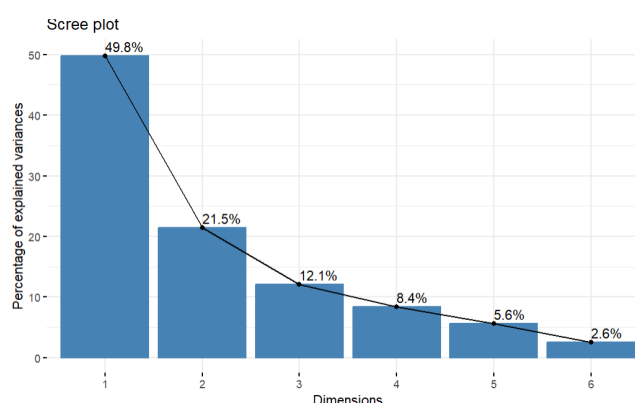
3.1. *Principal Component Analysis* (PCA)

Analisis *Principal Component Analysis* (PCA) dilakukan untuk mengidentifikasi variabel-variabel yang paling berpengaruh dalam menjelaskan variasi data terkait kemiskinan dan akses energi di Indonesia.

Tabel 3. Hasil Analisis PCA

	PC1	PC2	PC3	PC4	PC5	PC6
<i>Eigen value</i>	2.9866	1.2870	0.7277	0.5066	0.3367	0.1554
<i>Variability (%)</i>	49.78	21.45	12.13	8.44	5.61	2.59
<i>Cumulative Proportion (%)</i>	49.78	71.23	83.35	91.80	97.41	100.00

Berdasarkan hasil analisis *Principal Component Analysis* (PCA) pada **Tabel 3**, diketahui bahwa komponen utama pertama (PC1) memiliki nilai *eigen value* sebesar 2.9866 dengan kontribusi sebesar 49,78% terhadap total variasi data. Komponen utama kedua (PC2) memberikan kontribusi tambahan sebesar 21,45%, sehingga dua komponen utama pertama secara kumulatif mampu menjelaskan 71,23% dari total variasi data. Nilai ini menunjukkan bahwa sebagian besar informasi dalam enam variabel asal telah berhasil diringkas secara efektif ke dalam dua komponen utama tanpa kehilangan makna yang signifikan.



Gambar 1. Grafik *Scree Plot*

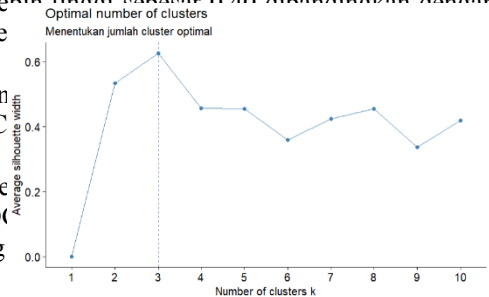
Hasil tersebut juga didukung oleh grafik *Scree Plot*, dimana terjadi penurunan tajam (*elbow point*) setelah komponen kedua. Hal ini menunjukkan bahwa dua komponen utama sudah cukup mewakili struktur utama data. Komponen utama pertama (PC1) merepresentasikan dimensi sosial ekonomi yang menggambarkan tingkat kesejahteraan dan pembangunan manusia, terlihat dari kontribusi besar variabel Indeks Pembangunan Manusia (IPM) yang berkorelasi positif terhadap PC1. Sedangkan persentase penduduk miskin dan tingkat pengangguran terbuka berkorelasi negatif. Artinya, semakin tinggi nilai PC1, semakin baik kondisi sosial ekonomi suatu provinsi. Sementara itu, komponen utama kedua (PC2) menggambarkan dimensi ketenagalistrikan yang berkaitan dengan rasio elektrifikasi, produksi tenaga listrik, dan daya tersambung pelanggan, yang mencerminkan kapasitas dan ketersediaan energi modern antarprovinsi.

Kedua komponen ini secara bersamaan mampu menjelaskan lebih dari 70% variasi total data, sehingga sudah cukup representatif untuk menggambarkan hubungan antara faktor kemiskinan dan akses energi di Indonesia. Temuan ini menunjukkan adanya keterkaitan erat antara pembangunan manusia dan pemerataan akses energi, di mana provinsi dengan nilai IPM tinggi dan rasio elektrifikasi besar umumnya memiliki tingkat kemiskinan dan pengangguran yang lebih rendah. Hasil PCA ini menjadi dasar bagi analisis kluster selanjutnya untuk mengelompokkan provinsi berdasarkan kesamaan karakteristik sosial-ekonomi dan energi dalam kerangka pencapaian SDGs 1 (Tanpa Kemiskinan) dan SDGs 7 (Energi Bersih dan Terjangkau).

Tabel 4. Matriks *Loading* PCA antar Variabel

	PC1	PC2
Persentase Penduduk Miskin	-0.47	-0.30
IPM	0.50	0.24
Tingkat Pengangguran Terbuka	0.35	0.01
Daya Tersambung	0.37	-0.55
Rasio Elektrifikasi	0.41	0.31
Produksi Tenaga Listrik	0.29	-0.66
Rata-rata	0.40	0.34

Tabel 4 menunjukkan bahwa tingkat dukungan terhadap pencapaian SDGs 1 (Tanpa Kemiskinan) yang direpresentasikan oleh PC1 memiliki rerata loading lebih tinggi sebesar 0.40 dibandingkan dengan tingkat dukungan terhadap pencapaian SDGs 7 (Energi Bersih dan Terjangkau). Variabel dengan loading tertinggi pada PC1 adalah menunjukkan peran besar kesejahteraan terhadap penurunan Listrik 0.29 menjadi yang terendah. Sementara itu, pada PC2 Daya Tersambung -0.55 memiliki kontribusi terbesar, infrastruktur energi dalam mendukung SDGs 7. Hasil ini menunjukkan kesejahteraan dan ketimpangan kemiskinan dan energi (SDGs 1), infrastruktur dan kapasitas energi (SDGs 7), yang saling berkelanjutan di Indonesia.

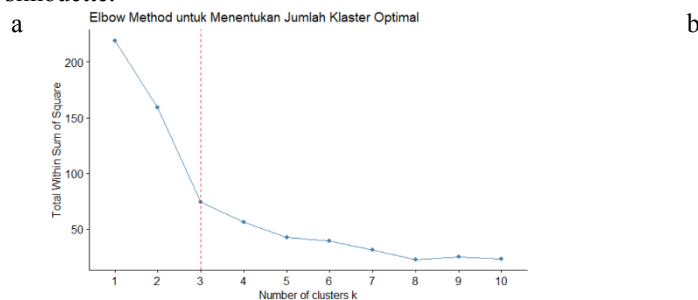


3.2. *K-Means Clustering*

Setelah melakukan analisis PCA dan didapat dua komponen utama yang menjelaskan 71,2% variasi total data dilanjutkan dengan analisis *K-Means Clustering*. Algoritma ini membutuhkan jumlah kluster yang sesuai karena centroid awal dapat berubah, sehingga berdampak pada hasil pengelompokan data yang tidak konsisten (Setiawan et al., 2022).

3.2.1. *Penentuan Jumlah Kluster*

Tahap pertama dalam *K-Means clustering* yaitu menentukan jumlah kluster (k) yang akan digunakan. Pada penelitian ini penentuan jumlah kluster (k) yang optimum menggunakan pendekatan metode elbow dan silhouette.



Gambar 2. (a) Metode *Elbow* ; (b) Metode *Silhouette*

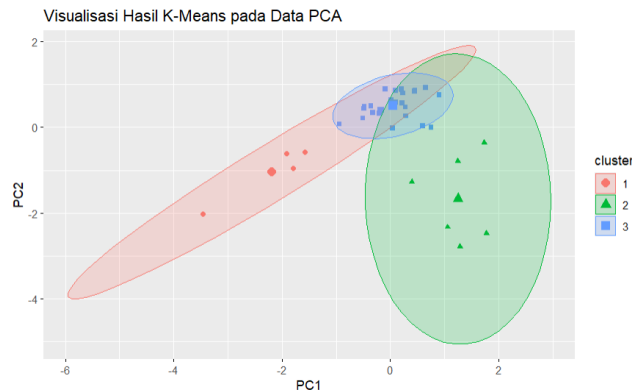
Gambar 2 merupakan hasil dari pendekatan Elbow untuk mendapatkan k yang optimal, berdasarkan plot yang didapat terlihat bahwa patahan gradien terbesar terdapat ketika kluster berjumlah 3. Plot metode *Silhouette* menunjukkan garis putus-putus yang tegak lurus dengan angka 3. Berdasarkan kedua analisis tersebut, dapat disimpulkan bahwa jumlah kluster yang optimal adalah 3.

3.2.2. *Pembentukan Kluster*

Hasil penentuan jumlah kluster dengan menggunakan kombinasi metode PCA dan *K-Means* menghasilkan 3 kluster yaitu kluster dengan ketimpangan energi dan kemiskinan rendah, sedang, dan tinggi. **Gambar 2** merupakan hasil output dari

clustering wilayah berdasarkan klaster dengan warna dan bentuk yang berbeda.

3.2.3. Analisis Hasil Klaster



Gambar 3. K-Means pada Data PCA

Gambar 3 merupakan hasil klasterisasi provinsi di Indonesia berdasarkan dua komponen utama (PC1 dan PC2) hasil analisis *Principal Component Analysis* (PCA). Setiap titik merepresentasikan satu provinsi, sedangkan warna dan bentuk titik menunjukkan klaster yang terbentuk melalui metode K-Means Clustering. Berdasarkan visualisasi tersebut, klaster dengan warna merah menunjukkan kelompok wilayah yang masuk dalam klaster 1, warna hijau menunjukkan kelompok wilayah yang masuk dalam cluster 2, dan warna kuning merupakan kelompok wilayah yang masuk ke dalam cluster 3. Jumlah anggota dari masing-masing klaster terdiri dari 4, 6, dan 28 provinsi di Indonesia. Berdasarkan gambar tersebut, terbentuk tiga klaster utama di mana setiap klaster memiliki karakteristiknya untuk membedakannya dengan klaster yang lain (Dewi & Pakereng, 2023).

Tabel 5. Hasil Analisis PCA per Klaster

Klaster	PC 1	PC 2
1	-3.78	-1.18
2	2.16	-1.89
3	0.08	0.57

Berdasarkan hasil pengelompokan, diperoleh tiga klaster utama yaitu sebagai berikut.

Klaster pertama memiliki nilai PC 1 dan PC 2 yang rendah, dengan nilai PC 1 sebesar -3.78 dan PC 2 sebesar -1.18. Hal ini menunjukkan bahwa provinsi yang masuk ke dalam klaster 1 memiliki tingkat kesejahteraan sosial-ekonomi yang relatif rendah dan kapasitas listrik yang terbatas.

Pada klaster kedua memiliki nilai PC 1 yang tinggi dan PC 2 yang rendah, yaitu sebesar 2.16 dan -1.89. Hal ini menunjukkan bahwa provinsi yang masuk ke dalam klaster 2 mengindikasikan wilayah dengan kondisi sosial-ekonomi yang baik dan kapasitas listriknya juga besar namun persebaran konsumsi energinya belum tentu efisien.

Nilai P1 dan P2 pada klaster ketiga bernilai positif dengan PC 1 sebesar 0.08 dan PC 2 sebesar 0.57. Kedua nilai PC yang positif tersebut merepresentasikan bahwa provinsi yang masuk ke dalam klaster ketiga memiliki kesejahteraan sosial-ekonomi menengah dan ketersediaan energi listrik sedang namun relatif efisien.

Gambar 4. Peta Indonesia berdasarkan Klaster

Gambar 4 menunjukkan hasil visualisasi spasial distribusi provinsi di Indonesia berdasarkan hasil klasterisasi menggunakan metode K-Means terhadap dua komponen utama hasil PCA. Peta ini memperlihatkan pembagian wilayah ke dalam tiga klaster utama yang merepresentasikan kondisi sosial-ekonomi dan kapasitas energi yang berbeda-beda.

Daerah berwarna merah merupakan klaster 1 yang meliputi provinsi-provinsi di kawasan Indonesia Timur, seperti Papua Pegunungan, Papua Tengah, Papua Selatan, dan Nusa Tenggara Timur. Wilayah ini ditandai dengan nilai PC1 dan PC2 yang rendah, menunjukkan karakteristik berupa tingkat kemiskinan yang tinggi, IPM yang relatif rendah, serta keterbatasan kapasitas infrastruktur energi listrik. Akses listrik di provinsi ini masih belum merata, baik dari sisi jangkauan maupun kualitas pasokan energi, sehingga menyebabkan keterlambatan pembangunan sosial dan ekonomi. Klaster ini dapat dikategorikan sebagai wilayah dengan prioritas tinggi dalam

Daerah
daerah klaster

Pulau Jawa
Jakarta, Jawa
Jawa Timur,
memiliki nilai
PC2 rendah,
tingkat
dan kapasitas

namun persebaran konsumsi energi listriknya belum tentu efisien. Rasio elektrifikasi dan produksi tenaga listrik di wilayah ini jauh di atas rata-rata nasional, serta didukung oleh infrastruktur dan perekonomian lebih mapan yang merepresentasikan wilayah maju.

Daerah berwarna hijau adalah daerah yang masuk dalam kategori klaster 3 yang mencakup mayoritas provinsi di luar Jawa dan Papua, termasuk wilayah Sumatera, Kalimantan, dan sebagian Sulawesi. Provinsi-provinsi pada klaster ini memiliki nilai PC1 dan PC2 yang sedang, menandakan tingkat kemiskinan menengah dengan akses energi yang cukup baik dan persebarannya sudah cukup efisien. Wilayah ini memiliki potensi besar untuk pengembangan energi terbarukan dan peningkatan produktivitas ekonomi melalui pemerataan infrastruktur listrik. Oleh karena itu, klaster ini dapat dikategorikan sebagai wilayah transisi menuju keseimbangan antara pertumbuhan ekonomi dan pemerataan energi.



Daerah berwarna biru adalah klaster 2 yang terdiri dari provinsi-provinsi di Jawa Barat, Jawa Tengah, dan Bali. Wilayah ini memiliki nilai PC1 yang tinggi dan PC2 yang rendah, yang mencerminkan kesejahteraan sosial yang tinggi, akses energi yang tinggi, namun persebaran konsumsi energi listriknya belum tentu efisien.

4. Simpulan

Penelitian ini memetakan ketimpangan antarprovinsi di Indonesia dalam konteks kesejahteraan sosial dan akses energi listrik dengan menggunakan pendekatan *Principal Component Analysis (PCA)* dan *K-Means Clustering*. Dua komponen utama yang diperoleh dari PCA mampu menjelaskan **71,2% variasi data**, yang menunjukkan bahwa sebagian besar informasi dari variabel-variabel asal telah terwakili dengan baik. Komponen pertama (PC1) menggambarkan dimensi kesejahteraan sosial dan akses energi. Komponen kedua (PC2) merepresentasikan kapasitas serta pemerataan energi, dengan beban negatif yang besar menandakan adanya perbedaan distribusi energi antarwilayah. Berdasarkan hasil

K-Means Clustering, diperoleh tiga kluster utama dengan ukuran masing-masing 4, 6, dan 28 provinsi, serta proporsi variasi antar-kluster sebesar 76,9%. Kluster 1 menunjukkan wilayah dengan tingkat kemiskinan tinggi, IPM rendah, dan akses energi terbatas. Kluster 2 merepresentasikan provinsi dengan tingkat kesejahteraan tinggi dan kapasitas energi besar, namun masih menghadapi tantangan efisiensi dan pemerataan energi karena sebagian besar konsumsi listrik terpusat pada sektor industri. Sementara itu, Kluster 3 mencerminkan wilayah dengan kondisi sosial-ekonomi menengah dan akses energi yang relatif merata. Secara keseluruhan, hasil analisis menunjukkan adanya kesenjangan spasial yang cukup signifikan antarprovinsi di Indonesia dalam hal kesejahteraan sosial dan ketersediaan energi listrik. Oleh karena itu, diperlukan kebijakan yang berfokus pada **pemerataan energi dan pengentasan kemiskinan**, terutama di wilayah yang masih tertinggal, guna mendukung pencapaian **Tujuan Pembangunan Berkelanjutan (SDG) 1 dan SDG 7** di Indonesia.

Daftar Pustaka

- Alirio, C., González, D., Manuel, Y., Calderón, M., Alberto, N., Cruz, M., Esteban, L., & Sandoval, P. (2022). *Typologies of Colombian off-grid localities using PCA and clustering analysis for a better understanding of their situation to meet SDG-7*. 3(April).
- Asghar, N., Asif, M., Munir, M., & Alhaji, R. (2022). Achieving sustainable development resilience : Poverty reduction through affordable access to electricity in developing economies. *Journal of Cleaner Production*, 376(August), 134040. <https://doi.org/10.1016/j.jclepro.2022.134040>
- Dewi, S., & Pakereng, M. A. I. (2023). Implementasi Principal Component Analysis Pada K Means Untuk Klasterisasi Tingkat Pendidikan Penduduk Kabupaten Semarang. *JIPi (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 8(4), 1186–1195. <https://doi.org/10.29100/jipi.v8i4.4101>
- Direktorat Jenderal Ketenagalistrikan. (2025). *Statistik 2024*. 38.
- Fourar, Y. O., Djebabra, M., & Boubaker, L. (2025). *Contribution of PCA / K-means methods to the mixed assessment of patient safety culture*. October. <https://doi.org/10.1108/IJHG-05-2020-0052>
- Hasiholan, D., Hamzah, M. Z., Sofilda, E., & Hariyanti, D. (2022). Relationship Analysis Between Unemployment and Poverty in 33 Provinces In Indonesia. *OIDA International Journal of Sustainable Development*, 15(2), 27–36.
- Hassanpour, H., Hamed, A. H., Mhaskar, P., House, J. M., & Salsbury, T. I. (2024). *A hybrid clustering approach integrating first-principles knowledge with data for fault detection in HVAC systems*. 187(February).
- IEA, IRENA, UNSD, World Bank, W. (2024). *THE ENERGY PROGRESS REPORT*.
- Ikhsan, I., & Amri, K. (2022). Does electrification affect rural poverty and households ' non-food spending ? Empirical evidence from western Indonesia Does electrification affect rural poverty and households ' non-food spending ? Empirical evidence from western Indonesia. *Cogent Economics & Finance*, 10(1). <https://doi.org/10.1080/23322039.2022.2095768>
- Mohammed, A. A., & Sumari, P. (2024). *Hybrid K-means and Principal Component Analysis (PCA) for Diabetes Hybrid K-means and Principal Component Analysis (PCA) for Diabetes Prediction : April*. <https://doi.org/10.12785/ijcds/1501121>
- Nur, R., Hartono, D., Dartanto, T., & Gultom, Y. M. L. (2024). Heliyon Multidimensional energy poverty : A study of its measurement , decomposition , and determinants in Indonesia. *Heliyon*, 10(3), e24135. <https://doi.org/10.1016/j.heliyon.2024.e24135>
- Pardita, D., Arjawa, I. G., & Setena, I. M. (2025). *The Development of Human and Poverty in Indonesia*. <https://doi.org/10.4108/eai.27-10-2023.2352808>

- Pokhriyal, N., Letouzé, E., & Vosoughi, S. (2022). Accurate intercensal estimates of energy access to track Sustainable Development Goal 7. *EPJ Data Sci.* <https://doi.org/10.1140/epjds/s13688-022-00371-5>
- Ria Jayanthi. (2021). *The Effect of Electricity Development in Indonesia on Poverty and Income Inequality*. 22(1), 104–116. <https://doi.org/10.23917/jep.v22i1>.
- Setiawan, K. E., Kurniawan, A., Chowanda, A., & Suhartono, D. (2022). Clustering models for hospitals in Jakarta using fuzzy c-means and k-means. *Procedia Computer Science*, 216(2022), 356–363. <https://doi.org/10.1016/j.procs.2022.12.146>
- Ultsch, A. (2024). *Informatics in Medicine Unlocked Comparative assessment of projection and clustering method combinations in the analysis of biomedical data*. 50(August). <https://doi.org/10.1016/j.imu.2024.101573>