

Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pekerjaan Layak dengan Menggunakan K-Medoids Clustering

Edy Widodo*, Shifa Qonita, Safira Feri Amalina, Sifa Nurul Aoliya

Jurusan Statistika, Universitas Islam Indonesia, Sleman, Indonesia

Corresponding Author. E-mail: edywidodo@uii.ac.id

Abstrak

Pekerjaan layak memainkan peran krusial dalam mengurangi kemiskinan dan menjadi salah satu tujuan utama dalam pembangunan berkelanjutan. Studi ini dilakukan untuk mengkaji berbagai indikator yang berperan dalam menentukan pekerjaan layak di Indonesia dan mengklasifikasikan tingkat kemampuan kerja di 34 provinsi berdasarkan indikator yang terkait dengan *sustainable development goals (SDGs)*, khususnya Tujuan 8 tentang pekerjaan layak dan pertumbuhan ekonomi. Data yang digunakan meliputi indikator-indikator seperti tingkat pengangguran terbuka, proporsi pekerja informal, persentase anak usia 10-17 tahun yang bekerja, perusahaan yang menerapkan norma K3, dan persentase pemuda usia 15-24 tahun yang saat ini tidak terdaftar dalam pendidikan, pekerjaan, atau pelatihan (NEET). Teknik analisis yang digunakan adalah metode *principal component analysis (PCA)* dan metode *k-medoids clustering*. Berdasarkan analisis PCA, diperoleh 2 faktor utama, yaitu faktor kesejahteraan dan ketenagakerjaan serta faktor pengangguran dan keselamatan kerja. Secara bersama-sama, kedua faktor ini menyumbang 73,9516% dari keseluruhan varians dalam data. Analisis *cluster* dilakukan dengan menggunakan 2 faktor utama ini. Berdasarkan analisis pengelompokan menggunakan metode *k-medoids*, dihasilkan 3 *cluster*. *Cluster 1* yang terdiri dari 20 provinsi merupakan wilayah dengan indikator pekerjaan layak yang moderat, *cluster 2* yang terdiri dari 5 provinsi merupakan wilayah dengan indikator pekerjaan layak yang tinggi, dan *cluster 3* yang terdiri dari 9 provinsi merupakan wilayah dengan indikator pekerjaan layak yang tinggi.

Decent work plays a crucial role in reducing poverty and serves as one of the key objectives in sustainable development. This study was conducted to examine various indicators that play a role in determining decent work in Indonesia and to classify the level of workability in 34 provinces based on indicators related to Sustainable Development Goals (SDGs), particularly Goal 8 on decent work and economic growth. The data used includes indicators such as the open unemployment rate, proportion of informal workers, percentage of children aged 10-17 years who work, companies that apply OSH norms, and the percentage of youth aged 15-24 who are not currently enrolled in education, work, or training (NEET). The analysis techniques used are Principal Component Analysis (PCA) method and k-medoids clustering method. Based on PCA analysis, 2 main factors were obtained, namely the welfare and employment factor and the unemployment and job safety factor. Together, these two factors account for 73.9516% of the overall variance in the data. Cluster analysis was conducted using these 2 main factors. Based on clustering analysis using the k-medoids method, 3 clusters were generated. Cluster 1 consisting of 20 provinces is a region with moderate decent work indicators, cluster 2 consisting of 5 provinces is a region with high decent work indicators, and cluster 3 consisting of 9 provinces is a region with high decent work indicators.

Kata kunci: Pekerjaan layak; sustainable development goals (SDGs); k-medoids; principal component analysis (PCA).

Keywords: decent work; sustainable development goals (SDGs); k-medoids; principal component analysis (PCA).

This is an open access article under
the [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)



How to Cite:

E. Widodo, S. Qonita, S. F. Amalina, and S. N. Aoliya, "Pengelompokan provinsi di Indonesia berdasarkan indikator pekerjaan layak dengan menggunakan k-medoids clustering," *Indonesian Journal of Applied Statistics*, vol. 8, no. 1, pp. 88-101, 2025, doi: 10.13057/ijas.v8i1.97996.

1. PENDAHULUAN

Pekerjaan yang layak memegang peran krusial dalam mendukung program pengurangan kemiskinan dan salah satu faktor keberhasilan pembangunan berkelanjutan. Selain itu, pekerjaan layak juga menjadi salah satu isu penting dalam ketenagakerjaan. Hal ini tercermin dalam berbagai program pembangunan yang dilaksanakan pemerintah untuk menunjang pekerjaan layak. Pentingnya pekerjaan layak dalam mendukung pencapaian pembangunan berkelanjutan tertuang pada tujuan ke-8 dari *sustainable development goals* (SDGs) yaitu berusaha meningkatkan pertumbuhan ekonomi yang inklusif dan berkelanjutan, kesempatan kerja yang produktif dan menyebluruh serta pekerjaan yang layak untuk semua [1].

Menurut *International Labour Conference* (ILO), indikator pekerjaan layak dapat dilihat berdasarkan upah rata-rata per jam, tingkat pengangguran terbuka (TPT), persentase dan jumlah anak usia 10-17 tahun, jumlah perusahaan yang menerapkan norma K3 (Keselamatan [2] dan Kesehatan Kerja), persentase usia muda serta tingkat setengah pengangguran. Pada tahun 2003, upah rata-rata per jam di Indonesia tercatat sebesar Rp. 19.027, yang masih tergolong sangat rendah. Selain itu, TPT juga cukup tinggi, yaitu sebesar 5,32%. Proporsi anak berusia 10-17 tahun tercatat sebesar 2,39%. Sementara itu, pada sektor ketenagakerjaan, jumlah perusahaan yang menerapkan standar keselamatan dan kesehatan kerja (K3) masih relatif sedikit, sebanyak 1.744 perusahaan, dimana masih banyak perusahaan yang tidak menerapkan norma K3. Di sisi lain, *not in employment, education, or training* (NEET) menunjukkan nilai yang masih tergolong tinggi, yaitu mencapai 22,25%. Pada tahun 2023, proporsi pekerja di sektor informal di Indonesia mengalami peningkatan, mencapai 84,13 juta orang atau sekitar 59,17% dari seluruh penduduk yang bekerja. Selain itu, angka pendapatan bersih (APB) Masyarakat Indonesia masih tergolong rendah, karena banyak potongan pajak, BPJS dan potongan lain yang cukup tinggi [3]. Kondisi ini menggambarkan bahwa banyak pekerjaan terutama di sektor informal yang belum mendapatkan pekerjaan yang layak. Selain itu, ketidakmerataan distribusi kesempatan pekerjaan antar provinsi di Indonesia juga menjadi perhatian bagi pemerintah, dimana beberapa daerah masih kekurangan lapangan kerja yang memenuhi standar kualitas.

Pekerjaan layak menjadi salah satu unsur kunci yang berperan dalam meningkatkan taraf kesejahteraan masyarakat dan mengurangi berbagai dampak social yang ditimbulkan oleh pengangguran. Ketidadaan pekerjaan layak dapat meningkatkan angka kriminalitas di Indonesia [4]. Karena itu, hal ini memerlukan tindak lanjut yang lebih serius terhadap penyediaan pekerjaan yang layak untuk mencegah terjadinya masalah sosial lainnya. Salah satu upaya yang bisa ditempuh adalah dengan melakukan identifikasi wilayah-wilayah pada kelompok tertentu.

Pengelompokan wilayah dengan karakteristik serupa dapat dilakukan menggunakan analisis *k-medoids clustering*. Dalam penelitian ini, metode analisis yang digunakan adalah metode *K-Medoids Clustering* dengan penerapan algoritma reduksi *principal component analysis* (PCA). PCA berguna dalam meningkatkan efisiensi model dan menghindari dominasi atribut dengan dominan yang lebih besar. Selain itu, PCA mampu mengatasi masalah multikolinearitas antar variabel dengan membentuk komponen utama yang saling bebas korelasi, sekaligus tetap mempertahankan sebagian besar informasi penting dalam data, serta memudahkan proses visualisasi. Penerapan PCA dalam penelitian ini bertujuan untuk mereduksi dimensi variabel pekerjaan layak di Indonesia yang kemudian akan dilanjutkan dengan mengelompokkan pekerjaan layak di Indonesia dengan menggunakan *k-medoids clustering* [5].

Berdasarkan latar belakang tersebut, tujuan peneliti melakukan penelitian lebih lanjut adalah untuk mengetahui pengelompokan pekerjaan layak di Indonesia. Selain itu, penelitian ini juga bertujuan untuk menilai sejauh mana kelayakan pekerjaan pada 34 provinsi di Indonesia berdasarkan indikator pekerjaan layak dan pertumbuhan ekonomi. Penelitian ini diharapkan mampu memberikan pemahaman yang lebih mendalam mengenai distribusi dan karakteristik indikator pekerjaan layak. Penelitian sebelumnya dilakukan oleh Muningsi dkk. [6], Yafi dkk. [7], Azra [8], dan Zebua dkk. [9] dengan melakukan analisis pengelompokan menggunakan metode PCA, namun masih jarang penelitian yang

menggunakan indikator pekerjaan layak. Oleh karena itu, peneliti tertarik untuk melakukan pengelompokan provinsi di Indonesia berdasarkan indikator pekerjaan layak menggunakan metode *k-medoids clustering*.

2. METODE

2.1. Sumber Data

Data dalam penelitian ini diperoleh melalui *website* Badan Pusat Statistik pada tahun (BPS) tahun 2023. Data yang digunakan mencakup data rata-rata upah per jam, TPT, persentase serta jumlah anak usia 10-17 tahun yang bekerja, jumlah Perusahaan yang menerapkan norma K3, persentas pemuda usia 15-24 tahun yang tidak bersekolah, tidak bekerja, dan tidak mengikuti penelitian (NEET), tingkat setengah pengangguran, proporsi penduduk yang bekerja pada sektor informal, dan angka pendapatan bersih (APB) [10].

2.2. Populasi dan Variabel Penelitian

Populasi dalam penelitian ini mencakup seluruh provinsi yang ada di Indonesia sedangkan variabel yang digunakan dalam analisis ini disajikan pada Tabel 1 berikut:

Tabel 1. Variabel penelitian

Variabel	Nama Variabel	Definisi	Satuan
X_1	Upah rata-rata per jam	Jumlah uang yang dibayarkan kepada pekerja dalam satu jam kerja	Rupiah (Rp)
X_2	Tingkat pengangguran terbuka (TPT)	Proporsi penduduk yang menganggur dibandingkan dengan total angkatan kerja	Persen (%)
X_3	Persentase anak usia 10-17 tahun yang bekerja	Persentase jumlah anak yang berusia antara 10 hingga 17 tahun yang terlibat dalam pekerjaan terhadap jumlah anak yang berusia antara 10 hingga 17	Persen (%)
X_4	Jumlah perusahaan yang menerapkan K3	Jumlah perusahaan yang telah menerapkan K3 (Kesehatan keselamatan kerja)	Rasio
X_5	Persentase usia muda (15-24) yang sedang tidak bersekolah, bekerja, atau mengikuti pelatihan (NEET)	Persentase penduduk usia 15-24 tahun yang tidak sedang bersekolah, tidak bekerja, dan tidak mengikuti pelatihan dibandingkan dengan total penduduk pada kelompok usia tersebut	Persen (%)
X_6	Tingkat setengah pengangguran	Persentase jumlah pekerja yang sudah bekerja tidak bekerja secara penuh terhadap jumlah angkatan kerja	Persen (%)

Variabel	Nama Variabel	Definisi	Satuan
X_7	Penduduk yang bekerja pada kegiatan informal	Persentase pekerja di sektor informal dibandingkan dengan total penduduk yang bekerja.	Persen (%)
X_8	Angka pendapatan bersih (APB)	Jumlah imbalan yang diterima individu, berupa uang atau barang, yang diberikan oleh perusahaan, kantor, atau majikan di masing-masing provinsi.	Rupiah (Rp)

2.3. Analisis Deskriptif

Analisis deskriptif Adalah suatu pendekatan dalam melakukan analisis data dalam penelitian dengan tujuan menilai sejauh mana hasil penelitian pada suatu sampel dapat digeneralisasikan. Statistik deskriptif berkaitan dengan penyajian informasi yang menggambarkan suatu fenomena atau kondisi. Secara umum, statistik deskriptif digunakan untuk menggambarkan suatu kondisi, fenomena, atau masalah [11].

2.4. Principal Component Analysis (PCA)

Principal component analysis (PCA) merupakan pendekatan statistik yang melibatkan banyak variabel yang berfungsi untuk mereduksi jumlah variabel, mengidentifikasi multikolinieritas, serta membantu proses pengelompokan. Prinsip dari PCA adalah mengubah variable semula yang kemungkinan menunjukka korelasi tinggi antar dimensi menjadi sekumpulan variable baru yang saling bebas atau tidak saling berhubungan [12]. Penerapan PCA dilakukan dengan mengikuti langkah-langkah berikut.

Standarisasi Data

Proses standarisasi data sangat diperlukan ketika dataset memiliki variabel dengan satuan pengukuran yang berbeda-beda. Perbedaan skala pada data yang dipakai bisa berdampak pada hasil pengelompokan yang terbentuk dalam analisis indikator ketenagakerjaan. Untuk mengatasi perbedaan satuan pada data, dapat dilakukan proses standarisasi dengan menghitung nilai *z-score* [8].

$$Z = \frac{x - \mu}{\sigma} \quad (1)$$

dengan Z = nilai standarisasi, x = nilai data, μ = rata-rata, dan σ = standar deviasi

Matriks Kovariansi

Varians ($\text{Var}(x)$) digunakan untuk mengukur seberapa jauh data menyebar sebagai dasar untuk melihat tingkat penyimpangannya. Matriks kovarian $\text{Cov}(x,y)$ merupakan matriks yang berisi nilai kovariansi pada setiap elemennya yang dihitung berdasarkan data sampel. Misalkan x dan y merupakan variabel acak.

$$\text{Var}(x) = \sigma^2 = \frac{1}{n} \sum_{i=1}^n (z_{ij} - \mu_j)^2 \quad (2)$$

dengan z_{ij} = nilai standarisasi pada sampel ke- i , μ_j = rata-rata data pada variabel ke- j , dan n = jumlah sampel.

$$\text{Cov}(x, y) = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \mu_{xj})(y_{ij} - \mu_{yi}) \quad (3)$$

dengan μ_x = rata-rata sampel variabel x , μ_y = rata-rata sampel variabel y , x_{ij} = nilai observasi ke i dari variabel x , y_{ij} = nilai observasi ke i dari variabel y .

Uji Kelayakan Model

Kaiser-Meyer-Olkin (KMO) merupakan suatu indeks untuk membandingkan besarnya korelasi antar variabel dengan besarnya korelasi parsialnya secara keseluruhan yang digunakan untuk melihat sampel yang ada mencukupi untuk dilakukan metode analisis komponen utama. Nilai KMO yang rendah mengindikasikan bahwa penggunaan analisis faktor kurang sesuai untuk data tersebut. Suatu nilai KMO dinilai memadai apabila mencapai minimal 0,5 atau lebih [12].

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2} \quad (3)$$

dengan r_{ij} = koefisien korelasi kuadrat antara i dan j dan a_{ij} = koefisien korelasi kuadrat antara i dan j

Uji Kelayakan Data

Penilaian kelayakan setiap variabel untuk analisis komponen utama digunakan karakteristik *measure of sampling adequacy* (MSA) Jika suatu variabel memiliki nilai MSA berada di bawah 0,5, maka variabel tersebut harus dikeluarkan dari tahap seleksi variabel. Setelah itu, langkah pemilihan variabel diulang kembali tanpa menyertakan variabel tersebut.

$$MSA = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2} \quad (4)$$

dengan r_{ij} = koefisien korelasi kuadrat antara i dan j dan a_{ij} = koefisien korelasi kuadrat antara i dan j

Uji Korelasi Antar Variabel

Uji *Barlett's test of sphericity* digunakan untuk menilai apakah terdapat hubungan atau korelasi yang signifikan antar variabel yang dianalisis. Nilai *Barlett's test of sphericity* dikatakan signifikan jika χ^2 lebih dari $\chi^2_{(\alpha, df)}$.

$$\chi^2 = - \left| (N - 1) - \frac{(2p + 5)}{6} |R| \right| \quad (5)$$

dengan $df = \frac{p(p-1)}{2}$, N = jumlah pengamatan, p = jumlah variabel, dan $|R|$ = determinan matriks korelasi.

Analisis Komponen Utama

Proses PCA memanfaatkan sekumpulan vektor yang dikenal sebagai *eigenvector* serta nilai-nilai yang disebut *eigenvalue* untuk mengidentifikasi fitur-fitur paling berpengaruh dalam suatu dataset [13]. Penentuan jumlah komponen dilakukan berdasarkan nilai eigen lebih dari 1. Adapun persentase kontribusi varians dari setiap komponen utama terhadap total varians dalam data didapatkan dari:

$$PC(\%) = \frac{\text{NilaiEigen}}{\text{VarianceCovarian}} \times 100\% \quad (6)$$

Selain bermanfaat dalam menangani dimensi yang tinggi dan/atau kolinearitas antar variabel, penggunaan PCA dalam pembentukan *cluster* dapat meningkatkan ukuran kebaikan pengklasteran [14].

2.5. K-Medoids Clustering

Algoritma *k-medoids*, yang juga dikenal sebagai *partitioning around medoid* (PAM) merupakan teknik pengelompokan data yang membagi objek ke dalam beberapa *cluster* berdasarkan kedekatan jarak, karakteristik, atau kriteria tertentu. Metode ini cukup sensitif terhadap keberadaan *outlier* karena objek dengan karakteristik yang jauh berbeda dari mayoritas data dapat memengaruhi struktur *cluster* dan berpotensi mengganggu representasi *medoid* [15]. Proses penerapan algoritma *k-medoids* umumnya dilakukan melalui beberapa tahapan sebagai berikut:

1. Menentukan sejumlah pusat *cluster* awal sebanyak k sebagai titik awal pembentukan kelompok.
2. Mengelompokkan setiap objek ke dalam *cluster* terdekat dengan menghitung jarak antar objek dan pusat *cluster* menggunakan metode jarak *Euclidean*.
3. Memilih calon *medoid* baru secara acak dari objek-objek yang berada di dalam masing-masing *cluster* sebagai kandidat pengganti *medoid* sebelumnya.
4. Gunakan calon *medoid* tersebut untuk menghitung kembali jarak seluruh objek dalam masing-masing *cluster*.
5. Menghitung nilai total simpangan (S) dengan membandingkan selisih antara total jarak yang diperoleh pada iterasi baru dengan total jarak pada iterasi sebelumnya.
6. Mengulangi proses pada langkah 3 sampai 5 sampai tidak terjadi perubahan pada *medoid*, sehingga terbentuk *cluster* yang stabil dengan anggota yang tetap.

3. HASIL DAN PEMBAHASAN

3.1. Ringkasan Data

Ringkasan data merupakan penyajian informasi secara singkat yang menggambarkan karakteristik dari suatu kumpulan data. Pada Tabel 2, Disajikan ringkasan data terkait indikator-indikator pekerjaan layak di Indonesia tahun 2023 yang mencakup rata-rata upah per jam, tingkat pengangguran terbuka, persentase serta jumlah anak usia 10-17 tahun yang bekerja, jumlah perusahaan yang menerapkan norma K3, persentase usia muda (15-24) NEET, tingkat Setengah pengangguran, penduduk yang bekerja pada kegiatan informal, dan APB.

Tabel 2. Ringkasan data

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Mean	19662,029	4,614	2,818	22	22,318	7,264	59,352	2971855,971
Standar Deviasi	5664,102	1,419	1,256	44,434	5,213	2,430	10,409	695112,673

3.2. Standarisasi Data

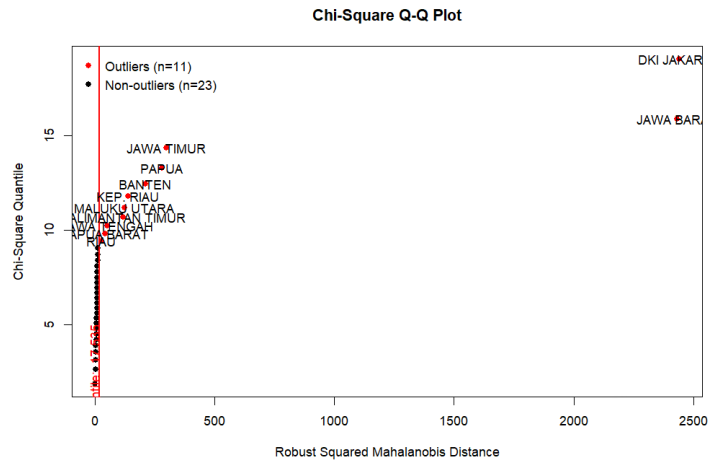
Standarisasi diperlukan ketika variansi antar variabel memiliki perbedaan yang cukup besar, sehingga pada penelitian ini proses standarisasi dilakukan. Metode yang digunakan adalah *z-score normalization*. Data setelah dilakukan standarisasi ditampilkan dalam Tabel 3.

Tabel 3. Nilai *z-score* hasil standarisasi data

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Aceh	-0,366	0,997	-1,422	-0,427	1,285	1,327	0,048	-0,926
Sumatera	-0,362	0,899	0,917	-0,090	-0,429	-0,219	-0,079	-0,568
Utara								
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Papua	1,065	0,539	0,200	-0,472	1,086	0,224	-0,077	0,447
Barat								
Papua	1,572	-1,369	0,805	-0,495	0,526	-1,182	2,409	1,413

3.3. Pengecekan Outlier

Pengecekan data *outlier* dilakukan dengan metode “*adj*”, yaitu metode kuantil berdasarkan jarak Mahalanobis. Suatu observasi dikategorikan sebagai *outlier* apabila posisinya jauh berada di luar batas *whisker* data yang ada.



Gambar 1. Q-Q plot

Berdasarkan Gambar 1 dengan menggunakan metode kuantil berdasarkan jarak Mahalanobis, terdapat 11 data *outlier* dan 23 data bukan *outlier*. pada penelitian ini data *outlier* tetap dipertahankan karena memberikan informasi penting tentang karakteristik data.

3.4. Pengecekan Korelasi Parsial

Pengecekan korelasi parsial dilakukan untuk mengetahui hubungan antar variabel. Besar korelasi masing-masing variabel ditampilkan dalam Tabel 4.

Tabel 4. Matrix korelasi

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8
Z_1	1							
Z_2	0,376	1						
Z_3	-0,341	-0,638	1					
Z_4	0,462	0,557	-0,436	1				
Z_5	-0,039	0,320	-0,072	-0,140	1			
Z_6	-0,572	-0,264	0,316	-0,407	0,387	1		
Z_7	-0,500	-0,577	0,517	-0,445	0,320	0,607	1	
Z_8	0,897	0,478	-0,372	0,549	-0,065	-0,690	-0,624	1

Tabel 4 menggambarkan hubungan linear antar variabel. Nilai korelasi berada pada rentang -1 hingga 1 , di mana nilai positif menandakan hubungan yang searah, sedangkan nilai negatif menunjukkan hubungan yang berlawanan arah. Hubungan yang kuat terlihat antara variabel X_5 dan X_8 yaitu sebesar $0,8977$. Hubungan yang lemah terlihat pada pasangan seperti X_5 dan X_4 yaitu sebesar $-0,1407$.

3.5. Uji Kecukupan Data dan Uji Kelayakan Model

Pengujian kecukupan data dilakukan dengan melihat nilai uji *Kaiser-Meyer-Olkin* (KMO) dan uji kelayakan setiap variabel dilakukan dengan menggunakan kriteria *measure of sampling adequacy* (MSA).

Tabel 5. Nilai KMO dan MSA

Variabel	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8
KMO	0,67							
MSA	0,71	0,60	0,84	0,70	0,26	0,84	0,68	0,68

Berdasarkan Tabel 5 dapat dilihat bahwa nilai KMO atau *overall* MSA sebesar 0,67. Karena nilai KMO > 0,5, maka sampel yang ada mencukupi untuk diolah dengan metode analisis komponen utama. Didapatkan juga nilai MSA untuk variabel $z_1, z_2, z_3, z_4, z_6, z_7, z_8 > 0,5$ sedangkan variabel $z_5 < 0,5$. Artinya, dari 8 variabel terdapat 1 variabel yang tidak layak untuk dilakukan analisis komponen utama dan akan dilakukan penghapusan variabel tersebut untuk langkah selanjutnya.

3.6. Uji Multikolinearitas

Uji Multikolinearitas dilakukan menggunakan metode *Bartlett's test of sphericity*. Pengujian *Bartlett's test* digunakan untuk menilai apakah variabel-variabel yang terlibat dalam dataset memiliki korelasi yang signifikan satu sama lain. Hasil pengujian *Bartlett's test* pada data Indikator Pekerjaan Layak di Indonesia tahun 2023 disajikan sebagai berikut.

I. Hipotesis

$H_0 : \rho_{xy} = 0$ (Tidak terdapat korelasi dari semua variabel)

$H_1 : \rho_{xy} \neq 0$ (Paling sedikit terdapat satu korelasi antar variabel)

II. Tingkat Signifikansi

$\alpha = 5\% = 0,05$

III. Daerah Kritis

Tolak H_0 jika $P - value < \alpha$ atau $\chi^2_{hitung} > \chi^2_{tabel}$

IV. Statistik Uji

$$\chi^2 = - \left| (N - 1) - \frac{(2p+5)}{6} |R| \right| = 137,997$$

V. Keputusan

Karena $P - value = 0,05 < 2,22 \times 10^{-16}$ dan

$\chi^2_{hitung} = 137,997 > \chi^2_{(0,05;21)} = 20,337$, maka tolak H_0

VI. Kesimpulan

Dengan menggunakan tingkat kepercayaan 95%, maka diperoleh kesimpulan paling sedikit terdapat satu korelasi antar variabel. Dengan demikian, menandakan bahwa pada data tersebut terjadi gejala multikolinearitas antar variabel yang akan diselesaikan dengan melakukan analisis komponen utama.

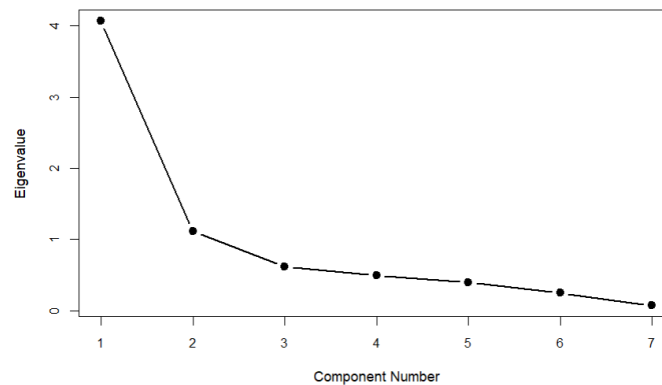
3.7. Analisis Komponen Utama

Nilai *eigen* menunjukkan jumlah varian yang mampu dijelaskan komponen utama. Nilai *eigen* pada data indikator pekerjaan layak ditunjukkan pada Tabel 6.

Tabel 6. Nilai *eigen values* data indikator pekerjaan layak di Indonesia tahun 2023

Komponen	1	2	3	4	6	7	8
<i>Eigen Value</i>							
Total	4,066	1,110	0,621	0,489	0,393	0,245	0,073
Varians (%)	58,085	15,865	8,873	6,995	5,621	3,501	1,056
Kumulative (%)	58,086	73,951	82,825	89,820	95,442	98,943	100

Tabel 6 menunjukkan bahwa komponen utama pertama dan kedua memiliki nilai *eigen* yang lebih besar dari 1. Dengan demikian, jumlah komponen utama yang dihasilkan adalah dua. Komponen pertama memiliki *eigen value* sebesar 4,0667 dengan kontribusi varians 58,085%, yang berarti komponen ini dapat menjelaskan 58,085% dari total variabilitas data. Komponen kedua memiliki *eigen value* sebesar 1,1105 dengan kontribusi varians sebesar 15,8656%, yang menunjukkan bahwa komponen ini mampu menjelaskan 15,8656% dari total keragaman data. Secara keseluruhan, faktor yang dihasilkan mampu menggambarkan 73,9516% variasi dari seluruh variabel pada indikator pekerjaan layak di Indonesia tahun 2023. Visualisasi nilai *eigen value* dapat dilihat melalui *scree plot* yang ditampilkan pada Gambar 2.



Gambar 2. Plot eigen value

Gambar 2 menunjukkan adanya penurunan yang sangat tajam dari komponen 1 menuju komponen 2. Namun, setelah melewati komponen 2, kurva mulai menurun dengan lebih landai dan terus melandai hingga mencapai komponen ke 8. Faktor *eigen* yang akan digunakan untuk menentukan kombinasi linier dari setiap komponen utama, yang ditampilkan pada Tabel 7.

Tabel 7. Nilai eigen faktor

Variabel		z_1	z_2	z_3	z_4	z_6	z_7	z_8
Komponen	1	0,395	0,356	-0,326	0,356	-0,363	-0,399	0,438
	2	0,387	-0,519	0,525	-0,178	-0,392	0,061	0,341

Berdasarkan Tabel 7 dapat dilihat kombinasi linier sebagai berikut:

$$y_1 = 0,395z_1 + 0,356z_2 - 0,326z_3 + 0,356z_4 - 0,363z_6 - 0,399z_7 + 0,438z_8$$

$$y_2 = 0,387z_1 - 0,519z_2 + 0,525z_3 - 0,178z_4 - 0,392z_6 + 0,061z_7 + 0,341z_8$$

dengan z_1 = Upah rata-rata per jam, z_2 = Tingkat Pengangguran Terbuka (TPT), z_3 = Proporsi anak berusia 10-17 tahun yang memasuki dunia kerja, z_4 = Banyaknya perusahaan yang menerapkan K3, z_5 = Proporsi penduduk usia muda (15-24) yang tidak sedang menampuh pendidikan, tidak memiliki pekerjaan dan tidak sedang menjalani pelatihan (NEET), z_6 = Tingkat setengah pengangguran, z_7 = penduduk yang bekerja pada kegiatan informal, z_8 = Angka Pendapatan Bersih (APB).

Pengelompokan komponen pada indikator-indikator pekerjaan layak di Indonesia tahun 2023 berdasarkan matriks komponen di rotasi dengan rotasi *varimax* ditampilkan pada Tabel 8.

Tabel 8. Pengelompokan komponen data indikator pekerjaan layak

Variabel		X_1	X_2	X_3	X_4	X_6	X_7	X_8
Komponen	1	19758,319	4,661	3,530	13,602	7,371	66,409	2678518,423
	2	17900,493	4,969	2,957	45,328	8,450	56,417	2636811,63

Pada Tabel 8 menunjukkan pengelompokan dilakukan berdasarkan nilai terbesar komponen matriks setiap variabel. Proses ini menggabungkan variabel-variabel awal menjadi kombinasi baru yang saling tidak berkorelasi, sehingga setiap komponen utama mewakili dimensi data yang unik dan memaksimalkan variansi data. Komponen 1 terdiri dari variabel X_1 , X_3 , X_7 , dan X_8 , sedangkan komponen 2 memuat variabel X_2 , X_4 , dan X_6 . Rincian variabel yang membentuk masing-masing faktor tersebut dapat dilihat pada Tabel 9.

Tabel 9. Faktor baru pada data indikator pekerjaan layak di Indonesia tahun 2023

Faktor	Variabel	Keterangan Variabel
Kesejahteraan dan ketenagakerjaan	X_1	Upah Rata-rata per Jam
	X_3	Persentase Anak Usia 10-17 Tahun
	X_7	Penduduk yang Bekerja pada Kegiatan Informal
	X_8	Pendapatan Bersih Satu Bulan
Pengangguran dan keselamatan kerja	X_2	Tingkat Pengangguran Terbuka (TPT)
	X_4	Jumlah Perusahaan yang Menerapkan K3
	X_6	Tingkat Setengah Pengangguran

Tabel 9 merupakan faktor baru yang terbentuk. Penentuan nama faktor dengan memperhatikan variabel-variabel yang membentuk masing-masing faktor. Faktor pertama diberi nama kesejahteraan dan ketenagakerjaan sedangkan faktor kedua dinamakan pengangguran dan keselamatan kerja.

3.8. Perhitungan Jarak

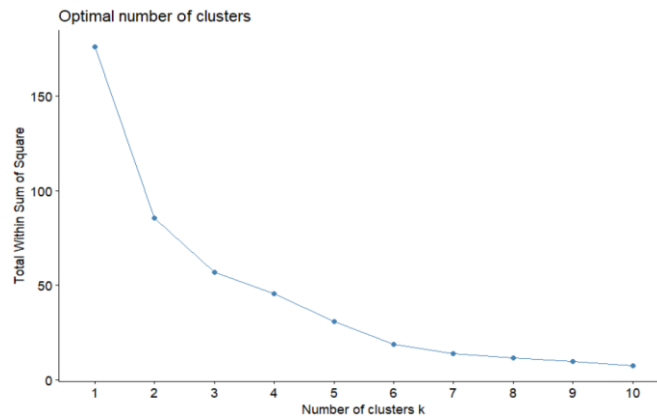
Perhitungan jarak antar observasi atau provinsi dilakukan untuk menentukan kesamaan antara titik-titik data. Jarak antar observasi dapat dilihat pada Tabel 10.

Tabel 10. Matrix jarak

	Aceh	Sumatera Utara	Sumatera Barat	...	Maluku Utara	Papua Barat	Papua
Aceh	0	1,973	1,035	...	2,862	2,724	5,165
Sumatera Utara	1,973	0	1,105	...	0,921	1,030	3,192
⋮	⋮	⋮	⋮	...	⋮	⋮	⋮
Papua Barat	2,724	1,030	2,068	...	0,553	0	2,688
Papua	5,165	3,192	4,229	...	2,347	2,688	0

3.9. Penentuan Jumlah Cluster

Jumlah *cluster* (k) dapat ditentukan langsung oleh peneliti ataupun dengan memanfaatkan beberapa metode, salah satunya adalah metode *within-cluster sum of squares* atau yang lebih dikenal sebagai metode *elbow*. Metode *elbow* digunakan untuk menentukan jumlah *cluster* yang paling optimal dengan mengamati titik terbentuknya “siku” pada grafik, yaitu perubahan signifikan pada persentase penurunan nilai ketika jumlah *cluster* ditambah. Perbandingan tersebut diperoleh dengan menghitung *error sum of squares* (ESS) untuk pilihan jumlah *cluster*.

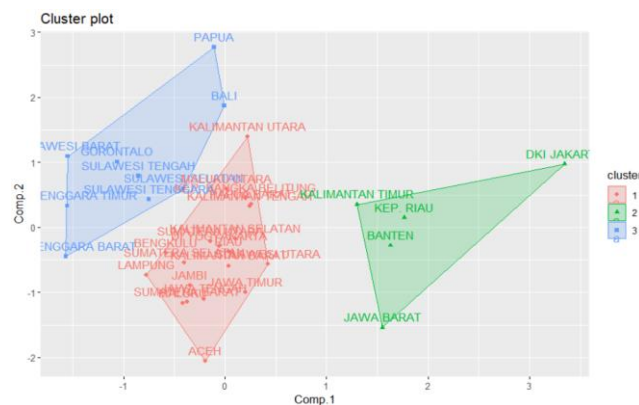


Gambar 3. K optimal elbow

Berdasarkan Gambar 3, terlihat adanya penurunan yang tajam dari komponen pertama hingga komponen ketiga. Setelah komponen ke-3, kurva tampak menurun dengan lebih landai hingga komponen ke-10. Dengan demikian, metode *elbow* menunjukkan bahwa nilai k yang optimal adalah 3.

3.10. Hasil Analisis Cluster

Melalui proses perhitungan jarak antar objek menggunakan metode *k-medoids*, diperoleh tiga kelompok provinsi di Indonesia yang memiliki karakteristik berbeda berdasarkan indikator pekerjaan layak pada tahun 2023.



Gambar 4. K-medoids clustering

Berdasarkan Gambar 4 diperoleh *plot* menggunakan metode *k-medoids* dengan terbentuknya 3 *cluster* dan hasilnya dapat dilihat pada Tabel 11.

Tabel 11. Anggota *k-medoids* clustering

Cluster	Provinsi	Jumlah
1	Kalimantan Utara, Maluku Utara, Kep. Bangka Belitung, Papua Barat, Kalimantan Tengah, Kalimantan Selatan, DI Yogyakarta, Aceh, Sumatera Utara, Riau, Sulawesi Utara, Kalimantan Barat, Bengkulu, Sumatera Selatan, Lampung, Jambi, Jawa Tengah, Sumatera Barat, Maluku, dan Jawa Timur	20

Cluster	Provinsi	Jumlah
2	Kep. Riau, Banten, Jawa Barat, DKI Jakarta, dan Kalimantan Timur	5
3	NTB, NTT, Sulawesi Barat, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Selatan, Gorontalo, Bali, dan Papua	9

Berdasarkan hasil *cluster*, peneliti melakukan pemetaan profil data indikator pekerjaan layak berdasarkan nilai rerata setiap variabel pada masing-masing *cluster*.

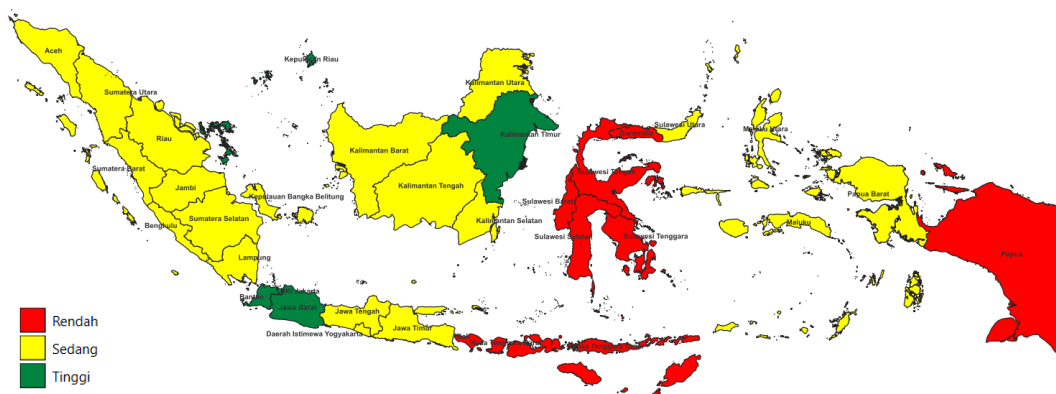
Tabel 12. Profilisasi *k-medoids clustering*

Cluster	Kesejahteraan dan ketenagakerjaan	Pengangguran dan Keselamatan Kerja
1	-0,178	-0,434
2	3,926	-0,073
3	-1,786	1,004

Tabel 12 menunjukkan nilai rata-rata setiap variabel dari tiap *cluster*. *Cluster* 1 dicirikan oleh pengangguran dan keselamatan kerja paling rendah dibandingkan dengan *cluster* lain. *Cluster* 2 dicirikan oleh kesejahteraan dan ketenagakerjaan paling tinggi dibandingkan *cluster* lain. *Cluster* 3 dicirikan oleh kesejahteraan dan ketenagakerjaan paling rendah serta pengangguran dan keselamatan kerja paling tinggi.

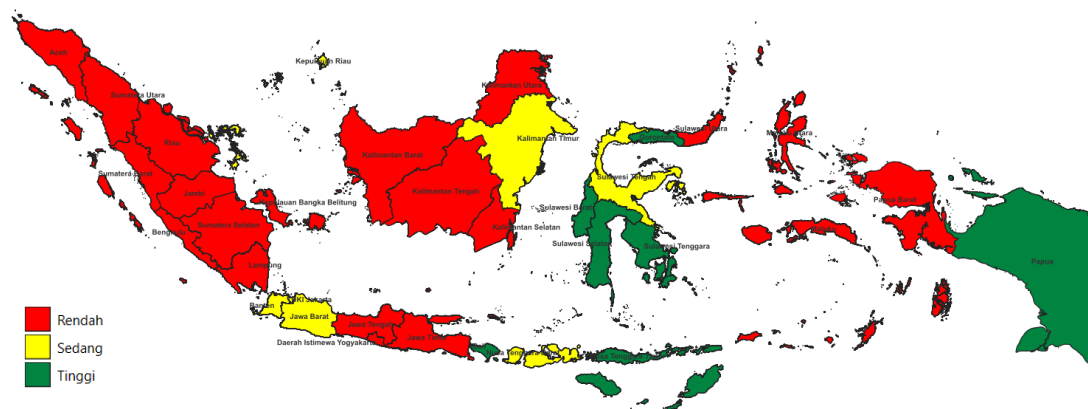
3.11. Pemetaan Hasil Analisis Cluster

Berdasarkan hasil *clustering* dengan menggunakan metode *k-medoids* dan menetapkan 3 cluster, pemetaan provinsi di Indonesia berdasarkan indikator pekerjaan layak dapat dilakukan, sebagaimana ditampilkan pada Gambar 5 dan Gambar 6.



Gambar 5. Peta penyebaran provinsi di Indonesia berdasarkan kesejahteraan dan ketenagakerjaan

Gambar 5 menunjukkan *cluster* berdasarkan faktor kesejahteraan dan ketenagakerjaan. Pada peta tersebut ditampilkan tiga variasi warna, di mana merah mempresentasikan provinsi dengan tingkat kesejahteraan dan ketenagakerjaan tinggi, warna kuning yang menunjukkan provinsi dengan kesejahteraan dan ketenagakerjaan sedang, dan warna hijau yang menunjukkan provinsi dengan kesejahteraan dan ketenagakerjaan rendah.



Gambar 6. Peta penyebaran provinsi di Indonesia berdasarkan pengangguran dan keselamatan kerja

Gambar 6 menunjukkan *cluster* berdasarkan faktor pengangguran dan keselamatan kerja. Pada peta tersebut ditampilkan tiga variasi warna, di mana merah mempresentasikan provinsi dengan pengangguran dan keselamatan kerja tinggi, warna kuning yang menunjukkan provinsi dengan pengangguran dan keselamatan kerja sedang, dan warna hijau yang menunjukkan provinsi dengan pengangguran dan keselamatan kerja rendah.

4. SIMPULAN

Hasil analisis menunjukkan bahwa *principal component analysis* (PCA) menghasilkan 2 faktor utama yang secara keseluruhan mampu menjelaskan 73.95% variansi data. Faktor pertama kemudian diinterpretasikan sebagai faktor kesejahteraan dan ketenagakerjaan. Faktor kedua adalah faktor pengangguran dan keselamatan kerja. Hasil dari analisis PCA ini selanjutnya digunakan dalam analisis *clustering* dengan menggunakan metode *k-medoids*, yang menghasilkan 3 *cluster*. *Cluster* 1 beranggotakan 20 provinsi yang dicirikan dengan pengangguran dan keselamatan kerja paling rendah dibandingkan dengan *cluster* lain. *Cluster* 2 beranggotakan 5 provinsi yang dicirikan dengan kesejahteraan dan ketenagakerjaan paling tinggi dibandingkan *cluster* lain. *Cluster* 3 beranggotakan 9 provinsi yang dicirikan dengan kesejahteraan dan ketenagakerjaan paling rendah serta pengangguran dan keselamatan kerja paling tinggi. Dengan diketahui kelompok-kelompok wilayah berdasarkan karakteristik tersebut, maka diharapkan dapat digunakan oleh pihak terkait dalam mengambil kebijakan untuk menetapkan skala prioritas dalam menciptakan lapangan kerja yang layak.

DAFTAR PUSTAKA

- [1] Jogja Dataku, "Sustainable Development Goals Pekerjaan yang Layak dan Pertumbuhan Ekonomi." [Online]. Available: <https://bappeda.jogjaprovo.go.id/dataku/sdgs/detail/8-pekerjaan-yang-layak-dan-pertumbuhan-ekonomi>.
- [2] E. Muningsih, N. Hasan dan G. B. Sulisty, "Penerapan metode principle component analysis (PCA) untuk clustering," *Bianglala Informatika*, vol. 8, no.1, pp. 58-62, 2020, <https://doi.org/10.31294/bi.v8i1.8470>.
- [3] Organisasi Perburuhan Internasional, "Menilai Pekerjaan Layak di Indonesia Sekilas tentang Profil Nasional untuk Pekerjaan Layak," [Online].
- [4] D. R. Dona, A. S. Effendi, dan M. Muliati, "Analisis faktor-faktor yang mempengaruhi kesempatan kerja," *Forum Ekonomi*, vol. 20, no.1, pp. 12-19, 2019, <https://doi.org/10.29264/jfor.v20i1.3308>.

- [5] E. Buulolo, *Data Mining untuk Perguruan Tinggi*. Yogyakarta, Indonesia: Deepublish, 2020.
- [6] E. Muningsih, N. Hasan, dan G. B. Sulisty, "Penerapan metode principle component analysis (PCA) untuk clustering data kunjungan wisatawan mancanegara ke Indonesia," *Bianglala Informatika*, vol. 8, No. 2, pp. 58-62, 2020, <https://doi.org/10.31294/bi.v8i1.8470>.
- [7] M. Yafi, R. Goejantoro, dan A. T. Rian Dani, "Pengelompokan algoritma k-medoids dengan principal component analysis (PCA) (studi kasus: kabupaten/kota di Pulau Kalimantan berdasarkan indikator kemiskinan tahun 2021)," *Prosiding Seminar Nasional Matematika, Statistika, dan Aplikasinya*, vol. 3, 2023.
- [8] M. A. Azra, "Penggunaan Davies Bouldin index dalam perbandingan algoritma k-means dan k-medoids untuk klaterisasi provinsi di Indonesia berdasarkan tingkat indikator ketenagakerjaan," 2023, <https://repository.unja.ac.id/50141/>.
- [9] Y. A. Zebua, F. Restuhadi, dan S. Hadi, "Analisis faktor-faktor kemiskinan kabupaten di Sumatera," *Berkala Perikanan Terubuk*, vol. 45, no. 3, pp. 57-66, 2023, <https://terubuk.ejournal.unri.ac.id/index.php/JT/article/view/5168>.
- [10] Badan Pusat Statistik, *Indikator Pekerjaan Layak di Indonesia 2023*. Jakarta: BPS, 2024. [Online]. Available: <https://www.bps.go.id/id/publication/2024/04/30/8fe6c4f34d5f1b9953903873/indikator-pekerjaan-layak-di-indonesia-2023.html>.
- [11] S. Sugiyono, *Statistik untuk Penelitian*. Bandung: Alfabeta, 2017, pp. 261-265.
- [12] A. Y. Sitompul, D. Leonidas, dan E. Sanggala, "Analisis faktor pengendalian kualitas teh hitam pada PT. ABC dengan menggunakan metode principal component analysis (PCA)," *Jurnal Cahaya Mandalika*, vol. 4, no. 3, pp. 1161-1172, 2023, <https://doi.org/https://doi.org/10.36312/jcm.v4i3.2074>.
- [13] B. M. S. Hasan dan A. M. Abdulazeez, "A review of principal component analysis algorithm for dimensionality reduction," *Journal of Soft Computing and Data Mining*, vol. 2, no. 1, pp. 20-30, 2021, <https://publisher.uthm.edu.my/ojs/index.php/jscdm/article/view/8032>.
- [14] Y. Nurohmah, R. Mayasari, dan B. N. Sari, "Optimalisasi performa k-means clustering dengan PCA dalam analisis tingkat kemiskinan di Jawa Barat," *Journal Teknik Informatika*, vol. 3, 2023, <https://doi.org/10.36040/jati.v7i3.6884>.
- [15] N. Puspitasari, G. Lempas, H. Hamdani, H. Haviluddin, dan A. Septiarini, "Perbandingan algoritma k-means dan algoritma k-medoids pada kasus covid-19 di Indonesia," *Building of Informatics Technology and Science (BITS)*, vol. 4, no. 4, pp. 2015-2017, 2023, <https://doi.org/10.47065/bits.v4i4.2994>.