EVALUASI K-MEANS DALAM ANALISIS STUNTING MELALUI INTEGRASI *PRINCIPAL COMPONENT ANALYSIS* DAN PERBANDINGAN *WARD-HIERARCHICAL CLUSTER*

Zakiyyah[1*], Ika Nur Laily Fitriana[2]
[1,2]Program Studi Statistika, Universitas Terbuka, Tangerang Selatan, Indonesia

**Penulis korespondensi: sheszakiyyah@gmail.com*

ABSTRAK

Stunting masih menjadi masalah serius di Indonesia dengan prevalensi 19,8% pada tahun 2024, disertai dengan masalah ketimpangan antardaerah. Hal ini mendorong pentingnya segmentasi antarwilayah menggunakan metode yang tepat. Penelitian ini berfokus pada evaluasi kinerja algoritma K-Means, yang sebelumnya diragukan efektivitasnya untuk analisis stunting, melalui modifikasi alur analisis. Tujuannya adalah menyajikan segmentasi provinsi, menganalisis performa K-Means yang diintegrasikan dengan *Principal Component Analysis* (PCA), dan membandingkan konsistensi hasilnya dengan metode *Ward-Hierarchical Cluster*. Untuk mencapai tujuan tersebut, analisis dilakukan terhadap data sekunder tujuh variabel determinan stunting dari 36 provinsi berdasarkan Survei Status Gizi Nasional Tahun 2024. Tahap awal mencakup standarisasi *Z-Score* dan uji asumsi PCA yang menunjukkan kelayakan data (KMO = 0,763, Sig. Bartlett = 0,000). Melalui reduksi dimensi, PCA berhasil mengekstraksi dua komponen utama yang merepresentasikan 71,742% total keragaman, memastikan informasi penting terwakili sebelum digunakan sebagai input klasterisasi. Adapun hasil evaluasi menggunakan *Cluster Silhouette* menunjukkan bahwa klasterisasi tanpa PCA memiliki kualitas rendah (K-Means: 0,258; Ward: 0,312). Integrasi PCA meningkatkan kualitas klaster secara signifikan dengan peningkatan koefisien *Cluster Silhouette* (PCA-K-Means: 0,420; PCA-Ward: 0,427). Temuan ini menegaskan bahwa integrasi PCA secara efektif meningkatkan performasi hasil klaster sekaligus memberikan pemahaman komprehensif mengenai perbaikan kinerja algoritma klasterisasi berbasis jarak saat multikolinearitas variabel diatasi.

Kata kunci: *Cluster Silhouette*, K-Means, *Principal Component Analysis*, Stunting, *Ward-Hierarchical Cluster*.

1 PENDAHULUAN

Kesehatan merupakan aspek fundamental dalam kehidupan manusia karena berperan terhadap keberlangsungan dan kualitas hidup di setiap tahap perkembangan (Azizah & Astuti, 2022). Pemenuhan gizi yang baik, terutama pada anak usia di bawah lima tahun, menjadi elemen penting untuk mendukung pertumbuhan dan perkembangan tersebut (Anandita & Gustina, 2022). Ketika terjadi kekurangan nutrisi, anak-anak cenderung mudah terjangkit penyakit karena tubuh tidak mampu melawan infeksi (Papotot et al., 2021). Salah satu bentuk nyata dari dampak malnutrisi kronis tersebut adalah stunting, yakni kondisi gagal tumbuh akibat kekurangan gizi dalam jangka waktu lama. Stunting menjadi masalah serius, terutama di wilayah pedesaan, karena berimplikasi pada rendahnya kualitas hidup, keterlambatan perkembangan kognitif, serta berkurangnya produktivitas di masa depan (Mustakim et al., 2022).

Hasil Survei Status Gizi Indonesia (SSGI) 2024 mencatat penurunan prevalensi stunting nasional menjadi 19,8%, sedikit lebih rendah dari proyeksi Bappenas sebesar 20,1% (Kementerian Kesehatan Republik Indonesia, 2025). Meski menunjukkan kemajuan, distribusi penurunan tersebut belum merata. Misalnya pada Provinsi Bali berhasil menurunkan angka stunting hingga 7,2%, sementara Papua Pegunungan mencatat prevalensi sebesar 40%, jauh di atas rata-rata nasional. Ketimpangan ini mengindikasikan disparitas penanganan stunting antardaerah dan menegaskan pentingnya strategi yang tidak hanya mempertahankan tren penurunan, tetapi juga mempercepat pencapaian target Rencana Pembangunan Jangka Panjang (RPJP) 2045, khususnya dalam aspek pemerataan. Salah satu upaya yang dapat dilakukan adalah memfokuskan intervensi pada wilayah yang memberikan kontribusi signifikan terhadap tingginya prevalensi stunting agar penggunaan anggaran dan sumber daya manusia lebih efisien serta menghasilkan dampak yang optimal (Nashriyah et al., 2023).

Kondisi tersebut menegaskan perlunya pendekatan berbasis data dalam segmentasi wilayah untuk penanganan stunting yang lebih terarah dan efektif (Mutawalli et al., 2025). Dalam konteks segmentasi, algoritma K-Means memiliki kelebihan sebagai metode klasterisasi yang banyak digunakan karena kesederhanaan, kecepatan, serta efektivitasnya dalam mengolah data (Khalish et al., 2025). Namun demikian, temuan Subayu (2022) menunjukan bahwa K-Means kurang efektif untuk analisis stunting karena menghasilkan akurasi yang rendah, sehingga klaster yang terbentuk dinilai belum merepresentasikan pengelompokan yang valid. Ketidakkonsistenan hasil penelitian terdahulu ini mengindikasikan adanya kebutuhan untuk mengevaluasi kembali performa K-Means melalui rancangan analisis yang lebih sesuai dengan karakteristik data stunting.

Sebagai respons terhadap isu tersebut, penelitian ini kemudian menyusun alur analisis yang dimodifikasi dengan menerapkan *Principal Component Analysis* (PCA) sebelum proses klasterisasi. Penerapan PCA ditujukan untuk mereduksi dimensi, meminimalkan redundansi antarvariabel, serta menonjolkan struktur informasi utama sehingga proses pengelompokan menjadi lebih stabil dan interpretatif. Pendekatan ini sejalan dengan temuan Saputri & Huda (2024) yang menekankan pentingnya segmentasi berbasis faktor penyebab stunting, dan diperkuat oleh Valentino et al. (2025) yang melaporkan bahwa kombinasi PCA dan K-Means dapat meningkatkan kejelasan struktur klaster dengan *Total Variance Explained* ≥ 80% serta *Silhouette Coefficient* sebesar 0,7 dalam analisis pengelompokan negara berdasarkan indikator pembangunan global.

Untuk memperkuat validitas hasil, penelitian ini juga membandingkan performa K-Means dengan metode klasterisasi lain yang bekerja dengan prinsip serupa, yakni *Ward-Hierarchical Cluster*, sehingga evaluasi kualitas klaster dapat dianalisis secara lebih menyeluruh. Dengan demikian, penelitian ini tidak hanya menyajikan segmentasi wilayah guna penanganan stunting di Indonesia, tetapi juga memberikan pemahaman komprehensif mengenai kinerja algoritma K-Means ketika diterapkan dengan alur analisis yang diperbaiki, serta menilai konsistensinya terhadap metode klasterisasi alternatif.

2 METODE

2.1 Data Penelitian

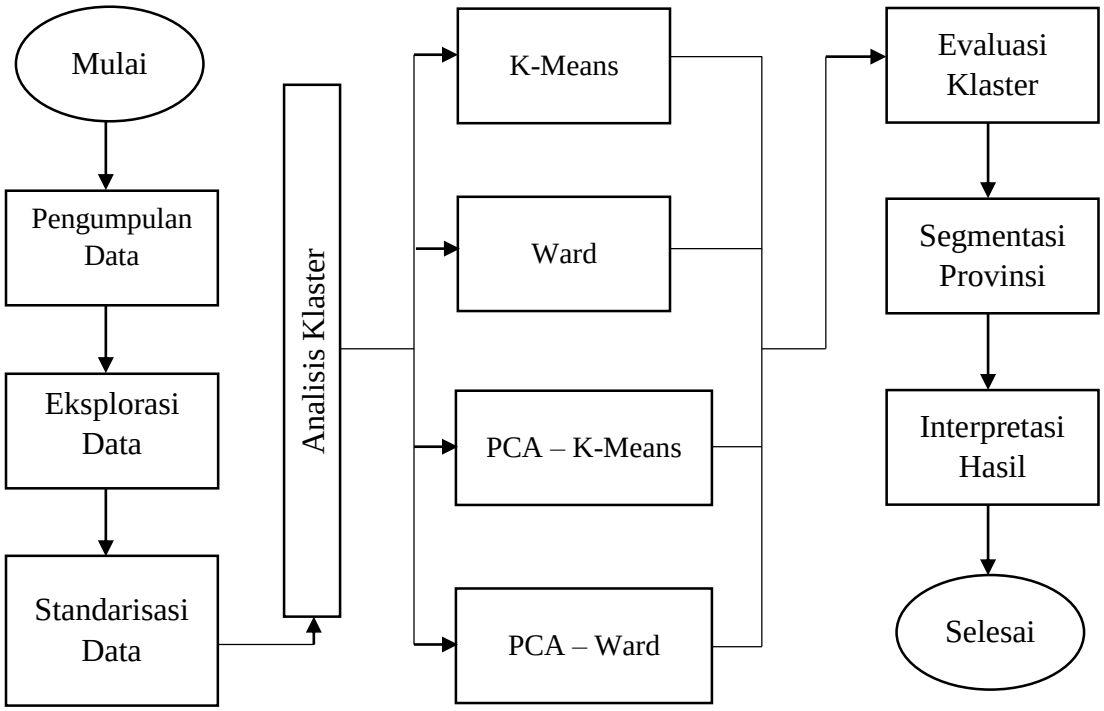
Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Survei Status Gizi Nasional tahun 2024 untuk 36 provinsi di Indonesia (Kementerian Kesehatan Republik Indonesia, 2025). Provinsi Papua Tengah dan Papua Pegunungan tidak diikutsertakan dalam analisis karena terdapat *missing value* yang berpotensi memberikan hasil kurang baik. Adapun variabel penelitian ditampilkan pada **Tabel 1.** berikut.

Tabel 1. Variabel Penelitian

Variabel	Keterangan
X_1	Proporsi bayi usia 0–5 bulan yang hanya menerima ASI tanpa tambahan makanan atau minuman lain, kecuali obat-obatan dan vitamin/mineral tetes (ASI Eksklusif)
X_2	Proporsi balita usia 6–23 bulan yang mengonsumsi makanan sumber protein hewani seperti daging, unggas, telur, dan ikan
X_3	Proporsi balita usia 6–23 bulan yang mengonsumsi makanan dengan kategori tidak sehat
X_4	Proporsi balita usia 6–23 bulan yang tidak mengonsumsi buah dan/atau sayuran
X_5	Proporsi balita yang mendapatkan obat cacing
X_6	Proporsi balita usia 6–59 bulan yang mendapatkan vitamin A dosis tinggi
X_7	Proporsi anak usia 12–23 bulan yang telah menerima imunisasi dasar lengkap (BCG, DPT-HB-Hib, Polio, dan Campak/MR)

2.2 Tahapan Penelitian

Tahapan penelitian ditunjukkan pada **Gambar 1.**



Gambar 1. *Flowchart* Tahapan Penelitian

2.3 Standarisasi Data

Data yang diperoleh menunjukkan tingkat variabilitas berbeda, sehingga perlu dilakukan standarisasi atau transformasi variabel tersebut ke bentuk *Z-score* (Juandi et al., 2023).

$$z_{ij} = \frac{\left(x_{ij} - \bar{x}_j\right)}{s_j} \quad (1)$$

Standarisasi ini penting agar tidak ada variabel yang memberi pengaruh berlebihan, sehingga seluruh variabel dapat berkontribusi secara seimbang.

2.4 *Principal Component Analysis* (PCA)

Principal Component Analysis (PCA) adalah teknik reduksi dimensi yang digunakan untuk merangkum informasi utama dari sekumpulan variabel yang saling berkorelasi ke dalam beberapa komponen utama yang tidak saling berkorelasi (Rosyada & Utari, 2024).

Secara matematis, setiap nilai variabel terstandarisasi z_{ij} digunakan untuk membentuk komponen utama melalui kombinasi linier, di mana skor komponen utama untuk observasi ke-i pada komponen ke-k dituliskan sebagai berikut.

$$PC_{ik} = \sum_{j=1}^{p} a_{jk} z_{ij} \tag{2}$$

dengan a_{jk} merupakan bobot variabel j pada komponen utama k.

Komponen utama ini dipilih berdasarkan urutan nilai eigen dari matriks korelasi variabel, yang menunjukkan besarnya varians yang dijelaskan oleh masing-masing komponen. Nilai eigen diperoleh dari penyelesaian persamaan determinan berikut.

$$|R - \lambda I| = 0 \tag{3}$$

dengan R adalah matriks korelasi dari data terstandarisasi dan I adalah matriks identitas.

Komponen utama yang memiliki nilai eigen terbesar mewakiliki proporsi varians data yang paling besar, sehingga hanya komponen dengan nilai terkini yang dipertahankan untuk analisis selanjutnya. Untuk memperkuat interpretasi struktur data, matriks *loading* dikalkulasikan dari bobot a_{jk}, yang menunjukkan kontribusi setiap variabel terhadap masing-masing komponen utama. Dengan demikian, setiap observasi berada dalam ruang dimensi yang tereduksi namun informatif.

Sebelum pembentukan komponen utama, dilakukan pengujian untuk memastikan bahwa struktur korelasi antarvariabel layak direduksi menggunakan PCA.

2.4.1 *Kaiser-Meyer-Olkin* (KMO) dan *Bartlett's Test*

Pengujian pertama dilakukan melalui ukuran *Kaiser–Meyer–Olkin* (KMO) dan *Bartlett's Test of Sphericity*. Ukuran KMO dihitung berdasarkan perbandingan antara jumlah kuadrat korelasi dan jumlah kuadrat korelasi parsial antarvariabel, yang dirumuskan sebagai berikut.

$$KMO = \frac{\sum_{j \neq k} r_{jk}^2}{\sum_{j \neq k} r_{jk}^2 + \sum_{j \neq k} q_{jk}^2} \tag{4}$$

dengan r_{jk} menyatakan koefisien antara variabel ke-j dan ke-k, serta q_{jk} menyatakan korelasi parsialnya. Nilai KMO yang tinggi $(> 0{,}5)$ menunjukkan kecukupan sampel dan kuatnya struktur korelasi untuk dilakukan PCA (El Haqq et al., 2025).

Selanjutnya, *Bartlett's Test* digunakan untuk menguji hipotesis nol bahwa matriks korelasi sama dengan matriks identitas. Statistik uji *Bartlett* dirumuskan sebagai berikut.

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

dengan n menyatakan jumlah observasi, p jumlah variabel, dan R matriks korelasi. Nilai signifikansi yang kecil $(< 0{,}05)$ mengindikasikan bahwa korelasi antarvariabel tidak bersifat acak, sehingga PCA layak untuk diterapkan (Permana & Guci, 2024).

2.4.2 *Measure of Sampling Adequacy* (MSA)

Selain ukuran KMO secara keseluruhan, kelayakan PCA juga dievaluasi pada tingkat variabel menggunakan *Measure of Sampling Adequacy* (MSA). MSA digunakan untuk menilai apakah setiap variabel memiliki korelasi yang memadai dengan variabel lain sehingga layak dipertahankan dalam analisis PCA (Damanik & Sitepu, 2024). Nilai MSA untuk variabel ke-j sama seperti pada persamaan (4). Begitupun interpretasi dalam menentukan skor kelayakannya.

2.4.3 *Communalities*

Setelah komponen utama terbentuk, kualitas representasi setiap variabel terhadap komponen yang dipertahankan dievaluasi melalui nilai *communalities*. *Communalities* menunjukkan proporsi varians suatu variabel yang dapat dijelaskan oleh komponen utama terpilih.

Nilai *communalities* untuk variabel ke-j dirumuskuan sebagai berikut.

$$h_i^2 = \sum_{j=1}^{m} a_{ij}^2 \tag{6}$$

dengan a_{jk} menyatakan *loading* variabel ke-j pada komponen utama ke-k, dan m adalah jumlah komponen utama yang dipertahankan. Nilai *communalities* yang tinggi $(> 0{,}5)$ menunjukkan bahwa variabel tersebut direpresentasikan dengan baik oleh struktur komponen utama (Arda & Andriany, 2019).

2.5 Analisis Klaster

Analisis klaster digunakan untuk mengelompokkan provinsi berdasarkan kemiripan karakteristik determinan stunting. Proses pengelompokan dilakukan menggunakan metode K-Means dan *Ward-Hierarchical Cluster* yang sama-sama berbasis jarak, sehingga hasil klaster yang terbentuk dapat dibandingkan secara seimbang.

2.5.1 K-Means

Metode K-Means mengelompokkan observasi ke dalam K klaster dengan meminimalkan variasi dalam klaster (Nugraha et al., 2022). Setiap observasi dialokasikan ke klaster dengan pusat terdekat, kemudian pusat klaster diperbarui secara iteratif hingga konvergen. Fungsi objektif K-Means dirumuskan sebagai berikut.

$$\min \sum_{k=1}^{K} \sum_{i \epsilon C_k} \|x_i - \mu_k\|^2 \tag{7}$$

dengan x_i menyatakan vektor observasi ke-i, μ_k adalah pusat klaster ke-k, dan C_k merupakan himpunan anggota klaster ke-k. Pembaruan pusat klaster dilakukan dengan menghitung rata-rata seluruh anggota klaster.

$$\mu_k = \frac{1}{|C_k|} \sum_{i \epsilon C_k} x_i \tag{8}$$

Proses ini diulang hingga tidak terjadi perubahan signifikan pada keanggotaan klaster.

2.5.2 *Ward-Hierarchical Cluster*

Sebagai pembanding, digunakan metode *Ward-Hierarchical Cluster* yang membentuk klaster secara bertahap melalui penggabungan pasangan klaster dengan peningkatan variasi dalam klaster yang paling kecil (Silalahi Sidebang, 2024). Kriteria penggabungan pada metode Ward didasarkan pada perubahan *Error Sum of Squares* (ESS) yang dirumuskan sebagai berikut

$$\Delta ESS = ESS_{(A \cup B)} - (ESS_A + ESS_B) \quad (9)$$

dengan ESS menyatakan jumlah kuadrat penyimpangan dalam klaster. Pada metode ini digunakan jarak *squared Euclidean*, dan jumlah klaster yang dibentuk disamakan dengan jumlah klaster pada metode K-Means agar hasil pengelompokan yang diperoleh dari kedua metode dapat dibandingkan secara konsisten dan seimbang. Klaster dengan nilai ΔESS terkecil dipilih untuk digabungkan pada setiap tahap, sehingga struktur klaster yang terbentuk cenderung lebih homogen.

2.6 Implementasi PCA dalam Analisis Klaster

2.6.1 Integrasi PCA – K-Means

Output PCA kemudian digunakan sebagai input baru untuk algoritma K-Means. Dengan mengurangi tumpang tindih informasi antarvariabel, K-Means bekerja lebih stabil dan mampu menghasilkan klaster yang lebih terpisah. Reduksi dimensi juga meningkatkan efisiensi komputasi dan mengurangi risiko pembentukan klaster yang bias.

2.6.2 Integrasi PCA – Ward-Hierarchical Cluster

Ward juga dijalankan kembali menggunakan komponen PCA agar proses penggabungan antarobjek lebih jelas dan tidak terdistorsi oleh korelasi antardeterminan. Dengan menggunakan komponen yang saling bebas, metode ini dapat menghasilkan struktur aglomerasi yang lebih terdefinisi. Perbandingan hasil PCA – Ward dan PCA – K-Means membantu menilai apakah pola klaster bersifat konsisten pada dua pendekatan yang berbeda.

2.7 Evaluasi Hasil Klaster (*Silhouette Coefficient)*

Kualitas hasil pengelompokan dievaluasi menggunakan *Silhouette Coefficient*, yang mengukur tingkat kesesuaian suatu observasi terhadap klaster yang diikutinya dibandingkan dengan klaster lain terdekat.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (10)$$

dengan $a(i)$ menyatakan jarak rata-rata observasi ke-i terhadap seluruh anggota dalam klaster yang sama, sedangkan $b(i)$ menyatakan jarak rata-rata minimum observasi ke-i terhadap klaster lain terdekat. Nilai *Silhouette* berada pada rentang $[-1,1]$, dimana nilai yang lebih tinggi menunjukkan bahwa observasi tersebut terklaster dengan baik (Salsabila & Ridwan, 2024).

3 HASIL DAN PEMBAHASAN

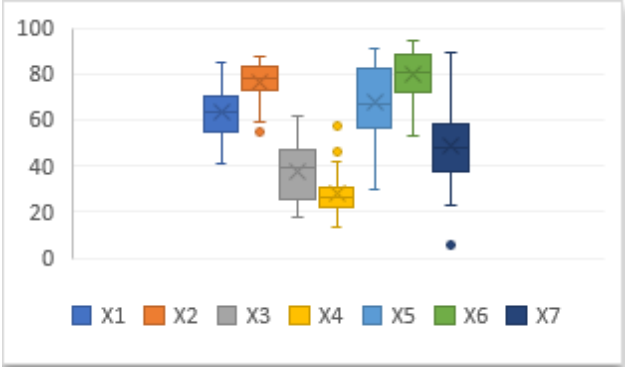
3.1 Eksplorasi Data

Gambaran awal mengenai karakteristik ketujuh variabel penyebab stunting dapat dilihat melalui statistika deskriptif. Informasi mengenai nilai minimum, maksimum, rata-rata, serta penyebaran masing-masing variabel memberikan petunjuk awal mengenai pola distribusi dan potensi ketimpangan antarprovinsi.

Tabel 2. Statistika Deskriptif

Variabel	*N*	*Minimum*	*Maximum*	*Mean*	*Standar Deviation*
X_1	36	41,400	85,200	64,139	10,966
X_2	36	55,200	88,100	77,114	8,448
X_3	36	17,500	62,000	38,122	12,064
X_4	36	13,200	57,400	28,231	9,844
X_5	36	29,600	91,800	68,261	17,215
X_6	36	53,400	94,800	80,087	10,298
X_7	36	5,400	89,500	48,564	17,752

Dari **Tabel 2.**, terlihat bahwa beberapa variabel menunjukkan rentang nilai yang cukup lebar, menandakan adanya perbedaan kondisi antarwilayah yang cukup tajam. Hal ini diperjelas melalui **Gambar 2.**, yang menyajikan *boxplot* distribusi masing-masing variabel determinan stunting.



Gambar 2. *Boxplot* Distribusi Variabel Determinan Stunting

Visualisasi ini menunjukkan variasi tingkat sebaran, nilai median, serta keberadaan pencilan pada beberapa variabel, yang mengindikasikan ketimpangan karakteristik antarprovinsi. Perbedaan sebaran tersebut menegaskan perlunya standarisasi data sebelum analisis lanjutan serta mendukung penggunaan metode reduksi dimensi dan klasterisasi untuk menangkap pola segmentasi wilayah secara lebih objektif.

3.2 Analisis PCA

Berdasarkan **Tabel 3.**, nilai KMO berada di atas batas minimal $(0{,}763 > 0{,}5)$ dan nilai signifikansi *Bartlett* sangat kecil $(0{,}000 < 0{,}05)$ mengindikasikan bahwa hubungan antarvariabel tidak bersifat acak, sehingga PCA layak digunakan dalam penelitian ini.

Tabel 3. KMO dan *Bartlett's Test*

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0,763
Bartlett's Test of Sphericity	*Approx. Chi-Square*	130,512
	Df	21
	Sig.	0,000

Temuan ini dipertegas oleh hasil *Anti-Image Correlation*, di mana. MSA yang konsisten di atas ambang batas mengisyaratkan bahwa masing-masing variabel memiliki kontribusi yang cukup dalam pembentukan komponen utama.

Tabel 4. *Anti-Image Correlation*

		X_1	X_2	X_3	X_4	X_5	X_6	X_7
Anti-image Correlation	X_1	0,844[a]	−0,064	0,061	0,295	−0,301	0,032	0,203
	X_2	−0,064	0,756[a]	−0,291	0,643	0,028	−0,098	0,095
	X_3	0,061	−0,291	0,810[a]	−0,014	−0,238	−0,181	0,313
	X_4	0,295	0,643	−0,014	0,745[a]	0,061	−0,101	0,193
	X_5	−0,301	0,028	−0,238	0,061	0,734[a]	−0,622	−0,459
	X_6	0,032	−0,098	−0,181	−0,101	−0,622	0,796[a]	−0,062
	X_7	0,203	0,095	0,313	0,193	−0,459	−0,062	0,653[a]

Nilai MSA pada **Tabel 4.** menunjukkan bahwa seluruh variabel memiliki kelayakan yang baik untuk dianalisis menggunakan PCA karena memiliki nilai MSA $> 0{,}5$ (0,844; 0,756; 0,810; 0,745; 0,734; 0,796; 0,653). Hal ini menandakan bahwa tidak ada variabel yang perlu dieliminasi dan seluruhnya dapat digunakan dalam PCA untuk mereduksi dimensi tanpa mengorbankan kualitas informasi.

Hubungan tersebut turut tergambar pada nilai *communalities* yang menunjukkan besarnya proporsi varians variabel yang berhasil dijelaskan oleh faktor yang terbentuk.

Tabel 5. *Communalities*

	Extraction
X_1	0,559
X_2	0,796
X_3	0,530
X_4	0,718
X_5	0,878
X_6	0,768
X_7	0,772

Nilai *communalities* pada **Tabel 5.** menunjukkan bahwa seluruh variabel memiliki proporsi varians yang cukup tinggi dan mampu dijelaskan dengan baik oleh komponen utama PCA.

Dengan demikian, seluruh uji asumsi telah terpenuhi, mulai dari KMO dan *Bartlett's Test*, MSA, hingga *communalities* menunjukkan bahwa data memiliki kualitas yang baik dan layak untuk dianalisis lebih lanjut menggunakan PCA sebagai langkah awal sebelum klasterisasi.

Setelah kelayakan PCA dipastikan, langkah berikutnya adalah mengevaluasi jumlah komponen yang layak dipertahankan melalui Tabel *Total Variance Explained*.

Tabel 6. *Total Variance Explained*

	Initial Eigenvalues			*Extraction Sum of Squared Loadings*			*Rotation Sums of Squared Loadings*		
Component	*Total*	*% of Var*	*Cum %*	*Total*	*% of Var*	*Cum %*	*Total*	*% of Var*	*Cum %*
1	3,834	54,776	54,776	3,834	54,776	54,776	2,871	41,013	41,013
2	1,188	16,966	71,742	1,188	16,966	71,742	2,151	30,729	71,742
3	0,784	11,196	82,938						
4	0,554	7,916	90,854						
5	0,303	4,334	95,187						
6	0,196	2,795	97,983						
7	0,141	2,017	100,000						

Tabel 6. menunjukkan bahwa dua komponen utama telah mampu menjelaskan proporsi keragaman data yang cukup besar, yaitu 71,742% total varians. Proporsi keragaman yang tinggi ini mengindikasikan bahwa informasi penting yang terkandung dalam tujuh variabel penyebab stunting dapat direpresentasikan secara efektif hanya dengan dua komponen.

Setelah diperoleh dua komponen utama yang mampu menjelaskan sebagian besar keragaman data, analisis dilanjutkan dengan menelaah kontribusi masing-masing variabel terhadap kedua komponen tersebut melalui nilai *loading*. Hasil awal menunjukkan bahwa komponen pertama didominasi oleh variabel X_1, X_2, X_3, X_4, X_5 dan X_6 yang memiliki nilai *loading* relatif tinggi, sedangkan variabel X_7 memiliki kontribusi paling kuat pada komponen kedua. Kondisi ini mengindikasikan bahwa komponen pertama berkaitan dengan aspek konsumsi gizi dan perilaku pangan, sementara komponen kedua mencerminkan dimensi intervensi kesehatan.

Namun, pada tahap awal struktur faktor masih menunjukkan tumpang tindih kontribusi antarvariabel, sehingga dilakukan rotasi Varimax untuk memperoleh pemisahan yang lebih jelas. Hasil rotasi memperlihatkan bahwa variabel X_1, X_2, X_3, dan X_4 memuat dominasi kuat pada komponen pertama, yang secara konsisten merepresentasikan faktor perilaku dan kualitas asupan gizi. Sementara itu, variabel X_5, X_6, dan X_7 memiliki *loading* tinggi pada komponen kedua, yang menggambarkan faktor intervensi kesehatan anak, seperti suplementasi dan imunisasi.

Rotasi ini menghasilkan struktur yang lebih bersih dan terpisah, sehingga dapat memberikan interpretasi yang jauh lebih kuat untuk tahap klasterisasi selanjutnya. Kemudian, hasil ini dirangkum pada tabel kesimpulan faktor yang menunjukkan pemisahan variabel ke dalam dua kelompok faktor utama.

Tabel 7. Kesimpulan Faktor

Faktor	Variabel
1	X_1, X_2, X_3, X_4
2	X_5, X_6, X_7

Struktur faktor yang dirangkum pada **Tabel 7.** kemudian dipertegas melalui **Tabel 8.**, yang menunjukkan bagaimana kedua komponen diputar untuk menghasilkan orientasi yang lebih jelas.

Tabel 8. *Component Transformation Matrix*

Component	1	2
1	0,798	0,603
2	−0,603	0,798

Tabel 8. menunjukkan nilai *Component Transformation Matrix* yang menggambarkan seberapa kuat masing-masing komponen berkontribusi setelah proses rotasi dilakukan. Nilai 0,798 dan 0,603 menunjukkan korelasi antara komponen awal dengan komponen hasil rotasi, di mana keduanya berada di atas batas kelayakan umum (> 0,50). Hal ini menandakan bahwa rotasi berhasil mempertahankan struktur komponen secara kuat dan tidak menyebabkan komponen kehilangan makna statistiknya. Dengan kontribusi yang masih tinggi dan stabil tersebut, kedua komponen layak digunakan sebagai dasar analisis lanjutan, termasuk proses klasterisasi, tanpa risiko distorsi atau kehilangan informasi penting.

3.3 Analisis Klaster

Pada tahap ini dilakukan simulasi untuk menentukan jumlah klaster yang paling optimal. Berdasarkan serangkaian percobaan, jumlah tiga klaster dipilih karena menghasilkan nilai evaluasi terbaik untuk K-Means. Jumlah klaster yang sama kemudian diterapkan untuk metode lainnya agar perbandingan kinerja dapat dilakukan secara seimbang dan konsisten. Perbandingan jumlah anggota pada masing-masing klaster antar metode ditampilkan pada **Tabel 9.** berikut.

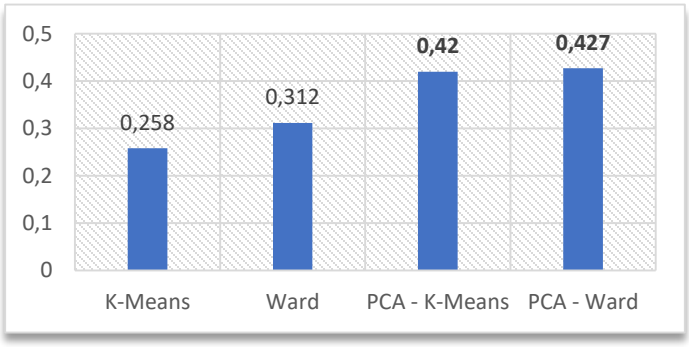
Tabel 9. Perbandingan Jumlah Anggota Klaster Antar-metode

Klaster	K-Means	PCA - K Means	Ward-Hierarchical Cluster	PCA - Ward-Hierarchical Cluster
1	13	12	21	24
2	15	13	12	7
3	8	11	3	5

Perbedaan jumlah anggota klaster pada **Tabel 9.** memperlihatkan bahwa masing-masing metode menghasilkan pola segmentasi wilayah yang tidak identik, sehingga tidak dapat langsung disimpulkan mana yang paling akurat dalam merepresentasikan kondisi determinan stunting antarprovinsi. Variasi ini menunjukkan pentingnya melakukan analisis kelayakan lanjutan, seperti evaluasi berdasarkan nilai *Silhouette*, untuk menilai seberapa baik tiap klaster terbentuk dan seberapa jelas pemisahan antarwilayahnya.

3.4 Evaluasi Hasil Klaster

Hasil evaluasi berdasarkan *Silhouette Cluster* menunjukkan tingkat pemisahan klaster yang berbeda pada setiap metode, sehingga memungkinkan penilaian objektif terhadap efektivitas masing-masing pendekatan.

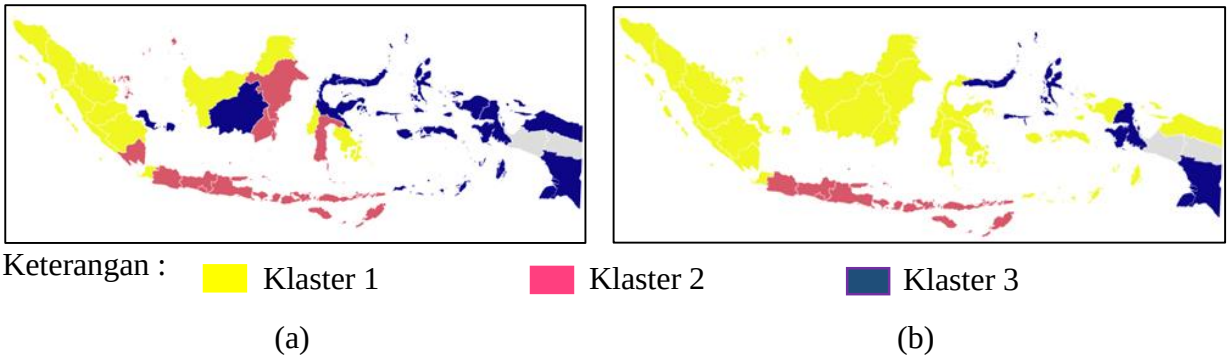


Gambar 3. Perbandingan Evaluasi Hasil Klaster Berdasarkan Nilai *Silhouette*

Grafik pada **Gambar 3.** menunjukkan bahwa nilai *Silhouette* untuk K-Means (0,258) dan *Ward-Hierarchical Cluster* (0,312) relatif rendah, menandakan klaster yang terbentuk tanpa PCA masih kurang optimal dan memiliki pemisahan antarklaster yang lemah. Setelah PCA diterapkan, kualitas klaster meningkat signifikan, terlihat dari naiknya nilai *Silhouette* pada PCA – K-Means (0,420) dan PCA – Ward (0,427). Kedua nilai tersebut berada di kategori lebih baik dan lebih stabil dibandingkan metode awal.

3.5 Segmentasi Provinsi Berdasarkan Klaster Terbaik

Berdasarkan hasil evaluasi menggunakan *Silhouette Coefficient*, metode PCA–Ward menghasilkan nilai koefisien tertinggi dibandingkan metode lainnya. Namun demikian, metode PCA–K-Means juga menunjukkan nilai *Silhouette* yang relatif mendekati serta menghasilkan distribusi anggota klaster yang lebih merata. Oleh karena itu, segmentasi provinsi dalam penelitian ini didasarkan pada kedua metode tersebut untuk memperoleh gambaran yang lebih komprehensif dan seimbang.



Gambar 4. Segmentasi Provinsi (a) Berdasarkan PCA – K-Means (b) Berdasarkan PCA – Ward

Visualisasi hasil segmentasi provinsi berdasarkan PCA–K-Means dan PCA–Ward ditunjukkan pada **Gambar 4.** Secara umum, kedua peta memperlihatkan pola segmentasi wilayah yang relatif serupa. Pulau Sumatera didominasi oleh satu klaster yang sama, Pulau Jawa membentuk kelompok tersendiri, sementara wilayah Papua cenderung terkelompok pada klaster yang berbeda dibandingkan wilayah lainnya. Kesamaan pola spasial ini mengindikasikan adanya konsistensi struktur klaster antara kedua metode.

Berdasarkan hasil PCA–K-Means pada **Gambar 4(a)**, klaster 1 mencakup provinsi-provinsi seperti Aceh, Sumatera Utara, Riau, Jambi, Sumatera Selatan, Bengkulu, Banten, Kalimantan Barat, Kalimantan Utara, Sulawesi Tenggara, dan Sulawesi Barat. Klaster 2 terdiri dari Lampung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Selatan, Kalimantan Timur, dan Sulawesi Selatan. Adapun klaster 3 mencakup Bangka Belitung, Kalimantan Tengah, Sulawesi Utara, Sulawesi Tengah, Gorontalo, Maluku, Maluku Utara, Papua Barat, Papua Barat Daya, Papua, dan Papua Selatan.

Sementara itu, hasil segmentasi berdasarkan PCA–Ward pada **Gambar 4(b)** menunjukkan komposisi klaster yang sedikit berbeda, meskipun pola besarnya tetap serupa. Klaster 1 mencakup sebagian besar provinsi di Sumatera dan Kalimantan serta beberapa provinsi di Sulawesi dan Papua. Klaster 2 didominasi oleh provinsi-provinsi di Pulau Jawa, Bali, Nusa Tenggara Barat, dan Nusa Tenggara Timur. Klaster 3 mencakup sejumlah provinsi di wilayah timur Indonesia, khususnya Sulawesi Utara, Gorontalo, Maluku Utara, Papua Barat, dan Papua Selatan.

Untuk memperjelas karakteristik masing-masing klaster, dilakukan *profiling* klaster berdasarkan nilai rata-rata variabel determinan stunting yang disajikan pada **Tabel 10.** dan **Tabel 11.**

Tabel 10. *Profiling* PCA – K-Means

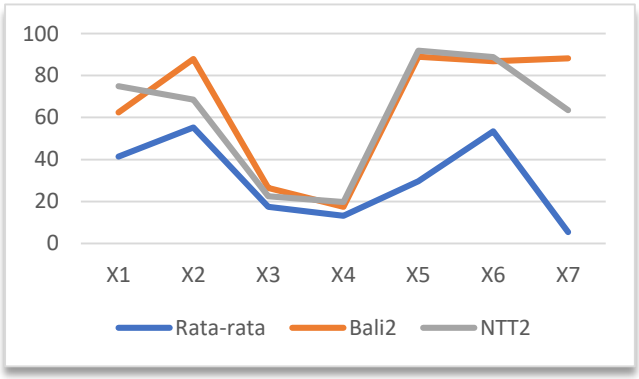
Klaster	X_1	X_2	X_3	X_4	X_5	X_6	X_7	Risiko Stunting
1	67,692	81,158	42,692	24,942	62,708	77,475	34,608	Sedang
2	70,438	80,908	43,662	22,515	85,146	88,362	64,285	Rendah
3	52,818	68,218	26,591	38,573	54,364	73,155	45,209	Tinggi
Rata-rata	64,139	77,114	38,122	28,231	68,261	80,086	48,564	

Tabel 11. *Profiling* PCA - Ward

Klaster	X_1	X_2	X_3	X_4	X_5	X_6	X_7	Risiko Stunting
1	64,979	79,608	39,083	26,733	66,008	78,679	41,388	Sedang
2	71,84	79,400	45,143	21,329	90,657	91,857	73,471	Rendah
3	49,360	61,940	23,680	45,080	47,720	70,360	48,209	Tinggi
Rata-rata	64,139	77,114	38,122	28,231	68,261	80,086	48,564	

Hasil *profiling* menunjukkan bahwa klaster 2 pada kedua metode cenderung memiliki nilai determinan yang lebih baik dan dikategorikan sebagai wilayah dengan risiko stunting rendah, sedangkan klaster 3 menunjukkan karakteristik sebaliknya dan dikategorikan sebagai wilayah dengan risiko stunting tinggi. Klaster 1 berada pada posisi antara keduanya dan merepresentasikan wilayah dengan risiko stunting sedang.

Temuan yang menarik dari penelitian ini adalah Bali sebagai provinsi dengan prevalensi stunting terendah, dan Nusa Tenggara Timur sebagai provinsi dengan prevalensi stunting tertinggi kedua setelah Papua Pegunungan secara konsisten berada di klaster yang sama, yaitu klaster 2. Pola ini mengindikasikan bahwa segmentasi klaster tidak semata-mata ditentukan oleh angka prevalensi stunting, melainkan oleh kemiripan struktur determinan yang mendasarinya.



Gambar 5. Perbandingan variabel determinan stunting pada Bali dan NTT

Berdasarkan grafik pada **Gambar 5.**, Bali dan Nusa Tenggara Timur memiliki pola yang serupa terutama pada variabel-variabel yang termasuk dalam faktor 2, yakni yang berhubungan dengan intervensi kesehatan. Dengan demikian, intervensi berupa pemberian obat cacing, vit A dosis tinggi, dan imunisasi dasar sudah cukup terpenuhi. Adapun pada variabel yang termasuk faktor 1, yaitu perilaku gizi, tidak ada pola yang dapat menyimpulkan kondisi stunting pada kedua provinsi tersebut. Provinsi Bali memiliki angka ASI eksklusif yang sedikit berada di bawah rata-rata dan di bawah angka capaian Nusa Tenggara Timur, sedangkan Nusa Tenggara Timur memiliki angka konsumsi sumber protein hewani yang berada di bawah angka rata-rata dan juga di bawah angka capaian Bali.

4 KESIMPULAN

Penelitian ini menunjukkan bahwa efektivitas klasterisasi determinan stunting sangat dipengaruhi oleh kualitas representasi data. Tujuan penelitian untuk mengevaluasi kinerja K-Means dan melihat perbaikan melalui integrasi PCA terbukti tercapai, di mana PCA berhasil mereduksi multikolinearitas dan menghasilkan struktur informasi yang lebih jelas. Perbandingan

dengan Ward memperlihatkan bahwa klasterisasi berbasis PCA memberikan pemisahan yang lebih stabil, sehingga lebih relevan digunakan untuk memetakan kondisi antarprovinsi. Temuan ini menegaskan pentingnya *preprocessing* dan reduksi dimensi dalam analisis klaster pada data kesehatan masyarakat yang kompleks.

Penelitian selanjutnya dapat mengembangkan pendekatan ini dengan menambahkan variabel determinan yang lebih beragam, menggunakan algoritma klasterisasi lain yang lebih adaptif, atau menguji model pada tingkat wilayah yang lebih kecil untuk memperoleh segmentasi kebijakan yang lebih presisi. Integrasi data temporal juga dapat dipertimbangkan untuk melihat dinamika perubahan risiko stunting dari waktu ke waktu.

DAFTAR PUSTAKA

Anandita, M. Y. R., & Gustina, I. (2022). Pencegahan Stunting pada Periode Golden Age Melalui Peningkatan Edukasi Pentingnya MPASI. *Al Ghafur: Jurnal Ilmiah Pengabdian Kepada Masyarakat, 1*(2), 79-86. https://doi.org/10.47647/alghafur.v1i2.917.

Arda, M., & Andriany, D. (2019). Analisis Faktor Stimuli Pemasaran dalam Keputusan Pembelian Online Produk Fashion Pada Generasi Z. *Proceeding FRIMA (Festival Riset Ilmiah Manajemen Dan Akuntansi), 2*(1), 434 - 440. https://doi.org/10.55916/frima.v0i2.66.

Azizah, S., & Astuti, E. T. (2022). Pengelompokan Provinsi di Indonesia Berdasarkan Determinan Kesehatan Balita dengan Menggunakan Analisis Cluster Tahun 2018. *Seminar Nasional Official Statistics, 2022*(1), 415-426. https://doi.org/10.34123/semnasoffstat.v2022i1.1484.

Damanik, S. S. U., & Sitepu, S. (2024). Principal Component Analysis Dalam Melihat Faktor-Faktor Yang Mempengaruhi Perkembangan Ternak Ayam (Studi Kasus: Peternakan Ayam Mutiara Sani, Mitra Pt Ciomas Adisatwa, Pekanbaru). *Leibniz: Jurnal Matematika, 4*(1), 47-65. https://doi.org/10.59632/leibniz.v4i1.394.

El Haqq, B. T., Antika, A., & Wulandari, S. P. (2025). Analisis Faktor-faktor Volume Ekspor Hasil Perikanan Menurut Provinsi di Indonesia Tahun 2021 menggunakan Analisis Faktor. Zoologi: Jurnal Ilmu Peternakan, Ilmu Perikanan, Ilmu Kedokteran Hewan, 3(1), 01-18. https://doi.org/10.62951/zoologi.v2i2.89.

Juandi, D., Amalia, R., Asmah, S. N., & Suryaningsih, Y. Analisis Klaster Karakteristik Mahasiswa Pendidikan Matematika Universitas Lambung Mangkurat. *EDU-MAT: Jurnal Pendidikan Matematika, 13*(2), 331-344. http://dx.doi.org/10.20527/edumat.v13i2.22555.

Kementerian Kesehatan Republik Indonesia. (2025). *Survei Status Gizi Indonesia (SSGI) 2024: Dalam angka*. Badan Kebijakan Pembangunan Kesehatan.

Khalish, F., Piranti, N. M., & Martadireja, O. (2025). Implementasi Data Mining Menggunakan Teknik Clustering dengan Metode K-Means. *JIIP - Jurnal Ilmiah Ilmu Pendidikan, 8*(5), 5392-5397. https://doi.org/10.54371/jiip.v8i5.7874.

Muhammad R. D. Mustakim, Irwanto, Roedi Irawan, Mira Irmawati, & Bagus Setyoboedi. (2022). Impact of Stunting on Development of Children between 1-3 Years of Age. *Ethiopian Journal of Health Sciences, 32*(3). https://doi.org/10.4314/ejhs.v32i3.13.

Mutawalli, L., Zaen, M. T. A., Tantoni, A., & Suhriani, I. F. (2025). Segmentasi Risiko Kesehatan Bayi dan Balita Menggunakan Algoritma K-Means. *Bulletin of Computer Science Research, 5*(4), 382-391. https://doi.org/10.47065/bulletincsr.v5i4.504.

Nashriyah, S. F., Makhful, M. R., & Devi, Y. P. (2023). Gambaran Spasial Hubungan antara Faktor Lingkungan dan Ekonomi dengan Stunting Balita di Provinsi Nusa Tenggara Timur. *Jurnal Spatial Wahana Komunikasi dan Informasi Geografi, 23*(2), 95–102. https://doi.org/10.21009/spatial.232.01.

Nugraha, A., Nurdiawan, O., & Dwilestari, G. (2022). Penerapan data mining metode K-Means clustering untuk analisa penjualan pada toko Yana Sport. *JATI (Jurnal Mahasiswa Teknik Informatika), 6*(2), 849-855. https://doi.org/10.36040/jati.v6i2.5755.

Papotot, G. S., Rompies, R., & Salendu, P. M. (2021). Pengaruh Kekurangan Nutrisi Terhadap Perkembangan Sistem Saraf Anak. *Jurnal Biomedik: JBM, 13*(3), 266-273. https://doi.org/10.35790/jbm.13.3.2021.31830.

Permana, J., & Guci, D. A. (2024). Analisis Faktor Eksploratori (EFA) dari Kinerja UKM di Medan. *Jurnal Ekonomi Bisnis, Manajemen dan Akuntansi (JEBMA), 4*(1), 463-467. https://doi.org/10.47709/jebma.v4i1.3682.

Rosyada, I. A., & Utari, D. T. (2024). Penerapan Principal Component Analysis untuk Reduksi Variabel pada Algoritma K-Means Clustering. *Jambura Journal of Probability and Statistics, 5*(1), 6-13. https://doi.org/10.37905/jjps.v5i1.18733.

Salsabila, F., & Ridwan, T. (2024). Analisa Volume Penyebaran Sampah di Karawang Menggunakan Algoritma K-Means Clustering. *Jurnal Informatika dan Teknik Elektro Terapan, 12*(2). https://doi.org/10.23960/jitet.v12i2.4226.

Saputri, A. L., & Huda, N. A. M. (2024). Pengelompokan Provinsi Berdasarkan Faktor Penyebab Stunting Menggunakan Finite Mixture Partial Least Square (FIMIX-PLS). *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya, 13*(4), 385–494. https://doi.org/10.26418/bbimst.v13i4.78039.

Silalahi Sidebang, A. M. A. (2024). Pemetaan Kemiskinan Digital Kabupaten/Kota di Sumatera Utara Menggunakan Ward Hierarchical Clustering. *Journal of Analytical Research, Statistics and Computation, 3*(2), 23–39. https://doi.org/10.4590/jarsic.v3i2.54.

Subayu, A. (2022). Penerapan Metode K-Means untuk Analisis Stunting Gizi pada Balita: Systematic Review. *Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi*, *2*(1), 42–50. https://doi.org/10.20885/snati.v2i1.18.

Valentino, M., Sipahutar, Y., & Farhan, M. (2025). Pengelompokan Negara Berdasarkan Indikator Pembangunan Global Menggunakan Metode PCA dan Clustering K-Means Tahun 2000-2020. *Jurnal Ilmu Komputer dan Sistem Informasi, 13*(2). https://doi.org/10.24912/jiksi.v13i2.35130.