

Penambahan Data Berbasis Pembelajaran Mesin untuk Pengelompokan Wilayah Beresiko *Stunting* di Kabupaten Bandung

Widhy Juliawan^{1*}, Haryoso Wicaksono²

^{1,2}STMIK Indonesia Mandiri
Jl. Belitung No.7, Merdeka, Sumur Bandung, Kota Bandung, Indonesia

e-mail korespondensi: juliawanwidhy@gmail.com

Submit: 25-08-2026 | Revisi: 22-09-2026 | Terima: 30-08-2026 | Terbit online: 02-10-2026

Abstrak - *Stunting* merupakan masalah pertumbuhan yang bersifat multidimensional sehingga penentuan prioritas wilayah memerlukan pembacaan serentak atas faktor kesehatan, lingkungan, layanan, dan demografi. Penelitian ini bertujuan mengelompokkan 31 kecamatan di Kabupaten Bandung berdasarkan kemiripan pola kerentanan terkait *stunting*. Data sekunder tahun 2024 dihimpun dari Profil Kesehatan Kabupaten Bandung dan Satu Data Kabupaten Bandung, sedangkan geometri kecamatan menggunakan batas administrasi Badan Informasi Geospasial edisi 2022. Tujuh fitur yang digunakan adalah prevalensi berat badan lahir rendah, sanitasi tidak aman, air minum tidak layak, rumah tangga non-PHBS, rasio Posyandu, rasio puskesmas, dan logaritma kepadatan penduduk. Seluruh fitur distandardisasi dan diarahkan agar nilai Z lebih tinggi menunjukkan kerentanan relatif lebih besar. Algoritma *K-Means* diuji untuk $k=2-6$. Pemilihan jumlah kluster mempertimbangkan *elbow*, *silhouette*, *Davies-Bouldin*, *Calinski-Harabasz*, keseimbangan ukuran, stabilitas 100 *seed*, dan *leave-one-out*. Solusi $k=2$ dipilih dengan *silhouette* 0,222, stabilitas *seed* ARI 1,000, dan ukuran kluster 15 serta 16 kecamatan. Kluster A menunjukkan pola kepadatan-air minum-keterbatasan layanan, sedangkan Kluster B menunjukkan pola BBLR-perilaku kesehatan. Prevalensi *stunting* tertimbang masing-masing kluster adalah 7,76% dan 10,19%. Namun, perbedaan distribusi prevalensi kecamatan tidak signifikan (*Mann-Whitney* $U=139,0$; $p=0,465$; *Cliff's delta*=0,158). Hasil ini mendukung penggunaan kluster sebagai tipologi eksploratif untuk merancang intervensi yang berbeda antar pola, bukan sebagai klasifikasi risiko tinggi dan rendah yang bersifat definitif.

Kata Kunci : *K-Means*; *Clustering*; Penambahan data; Kabupaten Bandung

Abstract - *Stunting* is a multidimensional growth problem; therefore, determining priority areas requires a simultaneous assessment of health, environmental, service, and demographic factors. This study aims to classify 31 sub-districts in Bandung Regency based on similarities in stunting-related vulnerability patterns. Secondary data for 2024 were collected from the Bandung Regency Health Profile and Satu Data Kabupaten Bandung, while sub-district geometries were obtained from the 2022 administrative boundaries provided by the Geospatial Information Agency. Seven features were employed, consisting of low birth weight prevalence, unsafe sanitation, inadequate drinking water access, non-PHBS households, Posyandu ratio, Puskesmas ratio, and the logarithm of population density. All features were standardized and transformed so that higher Z-scores represented greater relative vulnerability. The K-Means algorithm was evaluated for $k=2-6$ clusters. The optimal number of clusters was determined by considering the elbow method, silhouette score, Davies-Bouldin index, Calinski-Harabasz index, cluster size balance, 100-seed stability analysis, and leave-one-out validation. The $k=2$ solution was selected, achieving a silhouette score of 0.222, ARI seed stability of 1.000, and balanced cluster sizes of 15 and 16 sub-districts. Cluster A represents a density-drinking water-service limitation pattern, whereas Cluster B reflects a low birth weight-health behavior pattern. The weighted stunting prevalence of each cluster was 7.76% and 10.19%, respectively. However, the distribution of sub-district-level stunting prevalence between clusters was not statistically significant (*Mann-Whitney* $U=139.0$; $p=0.465$; *Cliff's delta*=0.158). These findings support the use of Clustering as an exploratory typology for designing differentiated interventions based on vulnerability patterns rather than as a definitive classification of high- and low-risk areas.

Keywords : *K-Means*; *Clustering*; Data mining; Bandung Regency.

1. Pendahuluan

Stunting adalah gangguan pertumbuhan dan perkembangan yang berkaitan dengan kekurangan gizi kronis, infeksi berulang, dan stimulasi psikososial yang tidak memadai. Seorang anak dikategorikan *stunting* ketika tinggi



badan menurut umurnya berada lebih dari dua simpangan baku di bawah median Standar Pertumbuhan Anak WHO. Dampaknya tidak berhenti pada tinggi badan, tetapi berkaitan dengan kemampuan kognitif, capaian pendidikan, produktivitas, dan risiko penyakit kronis pada tahap kehidupan berikutnya [1]. Karena penyebabnya terbentuk sejak masa prenatal hingga setelah kelahiran, penanganan *stunting* memerlukan intervensi lintas sektor dan tidak cukup hanya mengandalkan satu indikator gizi.

Indonesia telah mencatat penurunan prevalensi *stunting* nasional dari 21,5% pada 2023 menjadi 19,8% pada 2024. Meskipun demikian, variasi geografis dan sosial-ekonomi masih besar. Jawa Barat mencatat prevalensi 15,9% pada 2024, tetapi tetap menjadi salah satu provinsi dengan jumlah absolut balita *stunting* terbesar, yaitu sekitar 638.000 balita [2]. Kondisi tersebut menunjukkan bahwa angka rata-rata provinsi belum cukup untuk menentukan kebutuhan program pada tingkat kabupaten dan kecamatan. Pemerintah daerah membutuhkan informasi yang mampu membedakan karakteristik wilayah secara lebih rinci agar intervensi tidak diterapkan secara seragam pada wilayah dengan pola masalah yang berbeda.

Kabupaten Bandung terdiri atas 31 kecamatan dengan karakter demografi, lingkungan, dan kapasitas layanan yang beragam. Berdasarkan agregasi pencatatan layanan kesehatan tahun 2024 yang digunakan dalam penelitian ini, terdapat 240.796 balita yang diukur tinggi badan menurut umur dan 21.362 balita pendek, atau 8,87% secara agregat. Angka tersebut merupakan rasio dari data layanan yang dianalisis dan tidak disamakan secara langsung dengan estimasi survei SSGI karena sumber dan desain pengukurannya berbeda. Di balik nilai agregat tersebut terdapat variasi kecamatan yang lebar, sehingga penentuan prioritas berdasarkan prevalensi saja berisiko mengabaikan perbedaan mekanisme kerentanan yang melatarbelakangi setiap wilayah.

Literatur Indonesia menunjukkan bahwa *stunting* dipengaruhi oleh faktor yang bekerja pada beberapa tingkat. Berat badan lahir rendah merepresentasikan kerentanan awal kehidupan, sedangkan akses air minum, sanitasi, dan perilaku higiene berhubungan dengan paparan infeksi dan penyerapan zat gizi. Ketersediaan layanan kesehatan dan Posyandu memengaruhi jangkauan pemantauan, pencegahan, dan rujukan. Faktor rumah tangga dan konteks komunitas juga saling berinteraksi [3], [4], [5], [6]. Oleh sebab itu, tipologi wilayah seharusnya dibangun dari kombinasi indikator kerentanan, bukan hanya dari tinggi-rendahnya prevalensi *stunting*.

Penambangan data dan pembelajaran mesin menawarkan cara untuk mengidentifikasi struktur yang tidak langsung terlihat dalam data multivariat. *Clustering* bersifat tidak terawasi karena membentuk kelompok berdasarkan kemiripan tanpa memerlukan label kelas. *K-Means* relevan untuk penelitian ini karena efisien pada fitur numerik yang telah distandardisasi dan menghasilkan *centroid* yang dapat diterjemahkan menjadi profil kebijakan. Tinjauan sistematis terbaru menunjukkan bahwa penggunaan pembelajaran mesin dalam penelitian *stunting* di Indonesia meningkat, tetapi tantangan yang masih menonjol mencakup keterbatasan validasi eksternal, interpretabilitas, dan integrasi informasi spasial [7].

Sejumlah penelitian telah menerapkan *Clustering* untuk pemetaan *stunting*. Izza et al. (2019) mengelompokkan provinsi menggunakan prevalensi *stunting* bersama indikator pemberian ASI, penimbangan, serta kecukupan energi dan protein [8]. Handayani dan Sibuea (2023) membandingkan ukuran jarak *K-Means* untuk memetakan kasus *stunting* di Kabupaten Asahan [9]. Hasdyna et al. (2024) mengembangkan kerangka hibrida SVM, regresi, dan *K-Medoids* teroptimasi pada data Aceh [10]. Pada tingkat lokal, Mulyati et al. (2025) mengelompokkan kondisi kesehatan balita berdasarkan faktor perilaku dan lingkungan [11], sedangkan Ramadhani et al. (2025) menggabungkan kejadian *stunting*, faktor lingkungan, akses layanan, dan data spasial di Sumatera Utara [12]. Cahya et al. (2025) juga mengelompokkan 38 provinsi berdasarkan kasus *stunting* dan cakupan layanan esensial [13]. Posisi penelitian ini terhadap studi-studi tersebut diringkas pada Tabel 1.

Tabel 1. Posisi penelitian dan matriks kebaruan

Studi	Unit dan data	Metode/validasi	Perbedaan utama
Izza et al. (2019)[8]	Provinsi Indonesia; prevalensi dan indikator gizi	<i>K-Means</i> ; empat klaster	Prevalensi menjadi fitur; unit provinsi; validasi stabilitas tidak dilaporkan
Handayani & Sibuea (2023)[9]	Kabupaten Asahan; jumlah kasus <i>stunting</i>	<i>K-Means</i> ; <i>elbow</i> , DBI, <i>silhouette</i> ; perbandingan jarak	Berfokus pada kasus dan ukuran jarak; target eksternal terpisah tidak digunakan
Hasdyna et al. (2024)[10]	Kabupaten/kota Aceh 2019–2023	SVM, regresi, <i>K-Medoids</i> teroptimasi	Kerangka hibrida berorientasi prevalensi; bukan tipologi faktor kecamatan dengan target terpisah
Mulyati et al. (2025)[11]	231 catatan balita pada 25 kecamatan anonim	<i>K-Means</i> ; <i>elbow</i> dan <i>profiling</i>	Data individu diagregasi; validasi internal dan stabilitas masih terbatas

Studi	Unit dan data	Metode/validasi	Perbedaan utama
Ramadhani et al. (2025)[12]	Sumatera Utara; kejadian, lingkungan, layanan, koordinat	<i>K-Means; elbow</i> dan <i>silhouette</i>	Kejadian <i>stunting</i> masuk fitur dan klaster diberi tingkat risiko langsung
Cahya et al. (2025)[13]	38 provinsi; kasus dan layanan esensial 2022–2024	<i>K-Means; Spearman</i> dan <i>silhouette</i>	Unit provinsi; prevalensi/kasus menjadi pembentuk label risiko tinggi–rendah
Penelitian ini	31 kecamatan Kabupaten Bandung; tujuh faktor 2024	<i>K-Means; elbow, silhouette</i> , DBI, CHI, 100 <i>seed</i> , LOO-ARI, validasi target, GIS	Target <i>stunting</i> dipisahkan dari fitur; klaster ditafsirkan sebagai pola kerentanan, bukan peringkat risiko absolut

Tabel 1 menunjukkan bahwa penelitian terdahulu umumnya menggunakan prevalensi atau jumlah kasus *stunting* sebagai fitur pembentuk klaster dan banyak bekerja pada unit provinsi atau kabupaten. Berbeda dari pendekatan tersebut, penelitian ini membentuk klaster dari tujuh faktor kerentanan pada tingkat kecamatan, memisahkan prevalensi *stunting* sebagai target validasi eksternal, serta menambahkan pengujian stabilitas dan pemetaan GIS.

Celah penelitian terletak pada tiga hal. Pertama, sebagian studi memasukkan prevalensi atau jumlah kasus *stunting* ke dalam fitur *Clustering* dan kemudian menggunakan hasil yang sama untuk memberi label risiko. Prosedur ini bermanfaat untuk segmentasi prevalensi, tetapi kurang memadai untuk menilai apakah pola faktor kerentanan yang dibentuk tanpa target memiliki hubungan eksternal dengan *stunting*. Kedua, pemilihan jumlah klaster sering bertumpu pada satu atau dua ukuran internal tanpa menguji reproduktibilitas terhadap inisialisasi atau perubahan kecil pada sampel. Ketiga, bukti pada tingkat kecamatan di Kabupaten Bandung yang menggabungkan faktor kesehatan, WASH, perilaku, kapasitas layanan, demografi, dan pemetaan GIS belum tersedia dalam studi yang teridentifikasi.

Penelitian ini menawarkan empat kebaruan. Pertama, prevalensi *stunting* dipisahkan secara tegas dari tujuh fitur pembentuk klaster dan hanya digunakan setelah model selesai sebagai target validasi eksternal. Kedua, jumlah klaster dipilih melalui konsensus *elbow*, *silhouette*, DBI, CHI, keseimbangan ukuran, dan penalti *singleton*. Ketiga, reproduktibilitas diperiksa melalui 100 random *seed* dan leave-one-out ARI, disertai *silhouette* individu untuk mengenali kecamatan di batas klaster. Keempat, hasil diintegrasikan dengan batas GIS 31 kecamatan dan diterjemahkan menjadi pola intervensi yang berbeda tanpa memaksakan label tinggi–rendah ketika validasi target tidak mendukungnya. Dari perspektif teknologi dan sistem informasi, keluaran ini membentuk pipeline analitik yang transparan dan dapat dikembangkan menjadi komponen sistem pendukung keputusan kesehatan daerah.

Tujuan penelitian ini adalah mengidentifikasi dan menafsirkan pola kerentanan terkait *stunting* pada 31 kecamatan di Kabupaten Bandung menggunakan K-Means berbasis tujuh indikator tahun 2024. Secara khusus, penelitian ini bertujuan mengelompokkan kecamatan tanpa memasukkan prevalensi *stunting* sebagai fitur, menentukan jumlah klaster yang paling dapat dipertanggungjawabkan berdasarkan validasi internal, keseimbangan ukuran, dan stabilitas solusi, serta menganalisis profil, persebaran spasial, dan perbedaan prevalensi *stunting* eksternal pada klaster yang terbentuk.

2. Metode Penelitian

2.1 Desain, unit analisis, dan periode data

Penelitian ini menggunakan desain ekologis potong lintang berbasis data sekunder. Unit analisis adalah seluruh 31 kecamatan di Kabupaten Bandung. Atribut analitis merepresentasikan tahun 2024, sedangkan geometri digunakan hanya untuk visualisasi spasial dan berasal dari batas administrasi kecamatan edisi 2022 [14]. Karena analisis dilakukan pada tingkat wilayah, hasil tidak digunakan untuk menarik kesimpulan mengenai risiko individu.

2.2 Sumber data

Data kesehatan diperoleh dari Profil Kesehatan Kabupaten Bandung Tahun 2024 (edisi terbit 2025) [15], data Posyandu berasal dari portal Satu Data Kabupaten Bandung [16], dan batas administrasi berasal dari layanan *Geospasial* resmi Badan Informasi *Geospasial*. Tabel sumber dan penggunaan variabel disajikan pada Tabel 2.

Tabel 2 memperlihatkan integrasi empat sumber data yang memiliki fungsi berbeda. Data kesehatan dan Posyandu tahun 2024 membentuk atribut analitis, data Posyandu aktif digunakan untuk audit kualitas sebelum dikeluarkan karena variansnya nol, sedangkan geometri BIG tahun 2022 hanya digunakan untuk menampilkan persebaran spasial hasil klaster.

Tabel 2. Sumber data, tingkat asal, dan penggunaan

Dataset	Instansi	Tahun	Tingkat asal	Penggunaan
Profil Kesehatan Kabupaten Bandung Tahun 2024	Dinas Kesehatan	2024	Puskesmas/kecamatan	Penduduk, luas, puskesmas, BBLR, sanitasi, air minum, PHBS, <i>stunting</i>
Jumlah Posyandu (DPMD-075)	DPMD	2024	Desa/kelurahan	Jumlah Posyandu; diagregasi ke kecamatan
Jumlah Posyandu yang aktif (DPMD-069)[17]	DPMD	2024	Desa/kelurahan	Audit kualitas; dikeluarkan karena varians nol
Administrasi_AR_Kecamatan_10K	BIG	2022	Kecamatan	Geometri peta 31 kecamatan, EPSG:4326

2.3 Variabel penelitian

Prevalensi *stunting* dihitung sebagai $100 \times \text{jumlah balita pendek} / \text{jumlah balita diukur tinggi badan menurut umur}$. Variabel tersebut tidak dimasukkan ke proses *Clustering* dan hanya digunakan sebagai target eksternal setelah kluster terbentuk. Tujuh fitur pembentuk kluster dirinci pada Tabel 3. Pemilihan fitur didasarkan pada kerangka multidimensional *stunting* yang mencakup faktor awal kehidupan, WASH, perilaku kesehatan, kapasitas layanan, dan konteks demografi [1], [3], [4], [6].

Tabel 3. Definisi fitur dan arah standarisasi risiko

Fitur	Definisi operasional	Arah Z-risk
Prevalensi BBLR (%)	$100 \times \text{bayi BBLR} / \text{lahir hidup}$	+
Sanitasi tidak aman (%)	$100 - \text{persentase KK berakses sanitasi aman}$	+
Air minum tidak layak (%)	$100 - \text{persentase penduduk berakses air minum layak}$	+
Rumah tangga non-PHBS (%)	$100 - \text{persentase rumah tangga ber-PHBS dari yang dipantau}$	+
Posyandu per 1.000 balita diukur	Jumlah Posyandu/1.000 balita diukur TB/U	– (dibalik)
Puskesmas per 100.000 penduduk	Jumlah puskesmas/100.000 penduduk	– (dibalik)
Log kepadatan penduduk	$\log_{10} \text{penduduk per km}^2$	+

Catatan: tanda + berarti nilai mentah yang lebih tinggi diarahkan sebagai kerentanan relatif lebih tinggi; tanda – berarti skor Z dikalikan -1 sehingga rasio layanan yang lebih rendah menghasilkan skor risiko lebih tinggi.

Tabel 3 menunjukkan bahwa ketujuh fitur mencakup dimensi awal kehidupan, WASH, perilaku kesehatan, kapasitas layanan, dan demografi. Pembalikan arah pada rasio Posyandu dan puskesmas menyamakan interpretasi seluruh fitur, sehingga nilai Z-risk yang lebih tinggi selalu menunjukkan kerentanan relatif yang lebih besar.

2.4 Persiapan dan pemeriksaan data

Data puskesmas dan desa/kelurahan diagregasi ke kecamatan dengan mempertahankan *numerator* dan *denominator* sebelum rasio dihitung. Nama wilayah diseragamkan dan diverifikasi terhadap kode PUM. Dataset final berisi 31 kecamatan tanpa nilai kosong pada fitur maupun target. Kepadatan penduduk ditransformasi menggunakan logaritma basis 10. Variabel Posyandu aktif dikeluarkan karena seluruh kecamatan bernilai 100% sehingga tidak memiliki variasi. Lima observasi indikator yang terdeteksi sebagai *outlier* dengan aturan IQR dipertahankan karena tidak ditemukan bukti kesalahan entri dan merepresentasikan variasi wilayah.

Standarisasi dilakukan dengan *StandardScaler*. Untuk fitur kerugian, skor risiko dihitung sebagai $z = (x - \text{rata-rata}) / \text{simpangan baku}$. Untuk dua fitur kapasitas layanan, tanda skor Z dibalik. Dengan demikian, semua dimensi memiliki orientasi yang seragam: nilai Z-risk lebih tinggi menunjukkan kerentanan relatif yang lebih menonjol. Korelasi Spearman digunakan dalam eksplorasi karena ukuran sampel kecil dan dua variabel tidak memenuhi normalitas *Shapiro-Wilk*. Tidak ada pasangan fitur dengan $|\rho| \geq 0,85$; karena itu ketujuh fitur dipertahankan.

2.5 Pemodelan dan pemilihan jumlah kluster

K-Means digunakan untuk meminimalkan jumlah kuadrat jarak dalam kluster pada ruang tujuh fitur Z-risk. Kandidat $k=2$ sampai $k=6$ dievaluasi. Solusi dipilih berdasarkan gabungan *inertia/elbow*, *silhouette*, *Davies-Bouldin index (DBI)*, *Calinski-Harabasz index (CHI)*, keseimbangan ukuran kluster, keberadaan singleton, stabilitas terhadap 100 random seed, dan stabilitas *leave-one-out* menggunakan *adjusted Rand index (ARI)*. *Silhouette* yang lebih tinggi, *DBI* yang lebih rendah, dan *CHI* yang lebih tinggi menunjukkan struktur relatif lebih baik [18]. Pengujian stabilitas diperlukan karena tidak terdapat label benar pada *Clustering* tak terawasi [19].

Model final menggunakan $k=2$, $random\ state=42$, $n\ init=100$, dan $max\ iter=1.000$. Implementasi menggunakan Python dan scikit-learn [20]. Analisis komponen utama (PCA) hanya digunakan untuk visualisasi diagnostik dan tidak memengaruhi pembentukan kluster.

2.6 Profiling, validasi eksternal, dan pemetaan

Label substantif diberikan setelah model selesai dengan membaca *centroid Z-risk* dan statistik mentah setiap kluster. Validasi eksternal menggunakan prevalensi *stunting* yang sejak awal dipisahkan dari fitur. Prevalensi tertimbang dihitung dari total balita pendek dibagi total balita diukur pada setiap kluster. Perbedaan distribusi prevalensi antar kecamatan diuji dengan *Mann-Whitney U* dua arah; besaran efek dilaporkan menggunakan *Cliff's delta*. Pemetaan dilakukan dengan menggabungkan label kecamatan ke 31 geometri BIG melalui kode wilayah.

3. Hasil dan Pembahasan

3.1 Kesiapan dan karakteristik data

Seluruh 31 kecamatan memiliki data lengkap untuk tujuh fitur dan target validasi. Agregasi sumber mencakup 62 puskesmas dan 280 desa/kelurahan. Total 240.796 balita diukur TB/U dan 21.362 balita pendek, menghasilkan prevalensi kabupaten 8,87%. Dua variabel tidak normal pada $\alpha=0,05$, yaitu sanitasi tidak aman ($p=0,047$) dan air minum tidak layak ($p<0,001$). Korelasi antar fitur terbesar adalah rasio Posyandu dengan rasio puskesmas ($\rho=0,529$), masih di bawah ambang eksklusi 0,85.

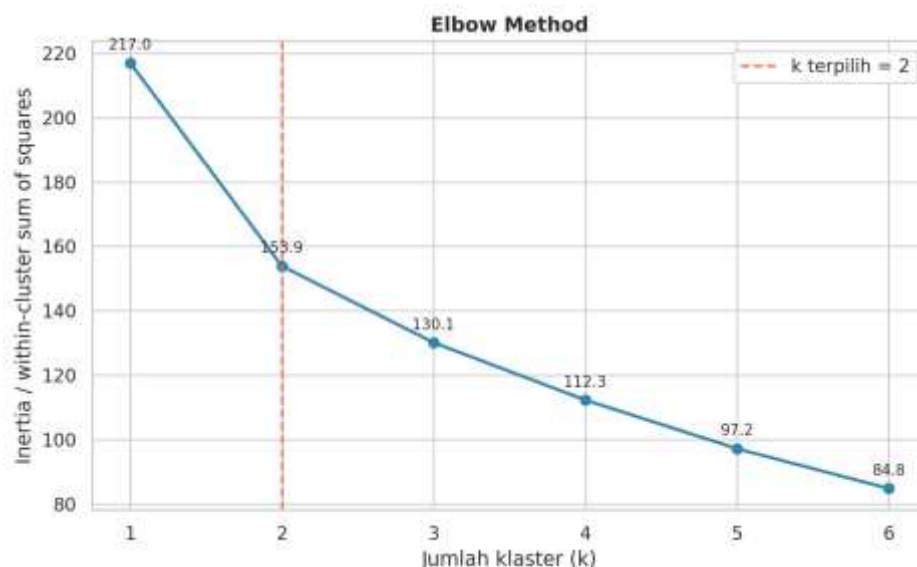
3.2 Pemilihan jumlah kluster

Penurunan *inertia* terbesar terjadi dari $k=1$ ke $k=2$, yaitu 29,09%, kemudian melandai menjadi 15,44% pada $k=3$ dan sekitar 12,75–13,71% pada kandidat berikutnya. Di antara $k=2-6$, $k=2$ memiliki *silhouette* tertinggi (0,222), CHI tertinggi (11,895), stabilitas *seed* sempurna ($mean\ ARI=1,000$), stabilitas *leave-one-out* tertinggi ($mean\ ARI=0,983$), ukuran seimbang 15–16, dan tidak memiliki *singleton*. Walaupun DBI $k=2$ bukan yang terendah, konsensus seluruh kriteria mendukung $k=2$.

Tabel 4. Evaluasi kandidat jumlah kluster

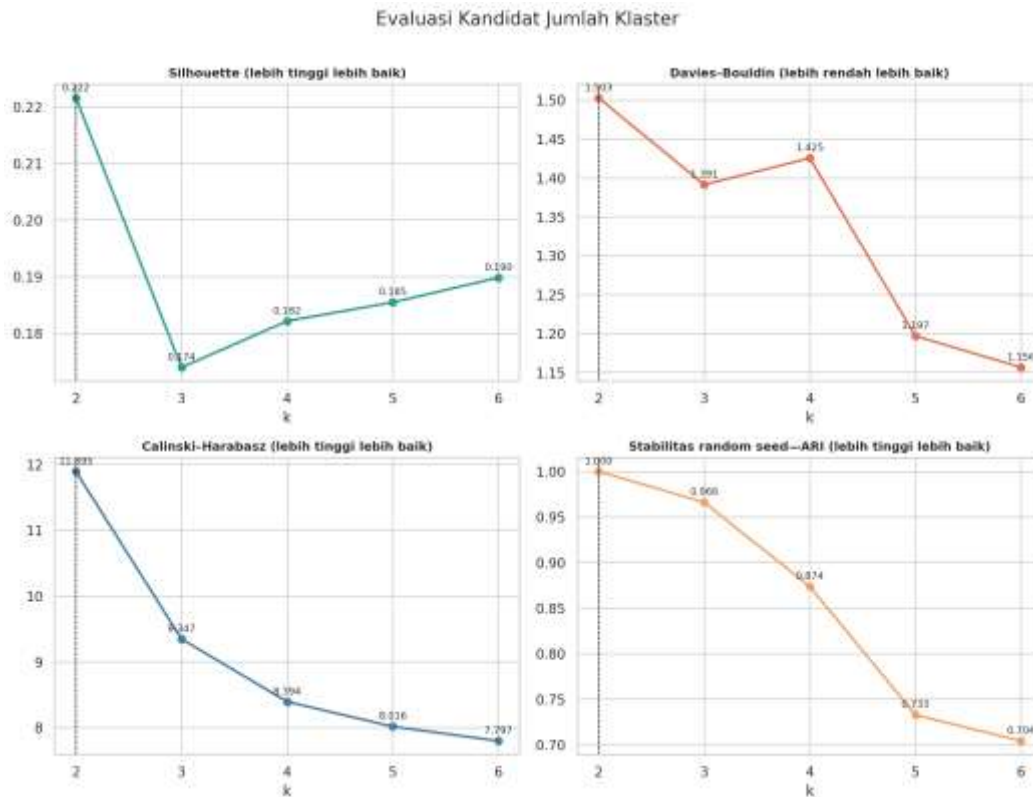
k	<i>Silhouette</i>	DBI	CHI	Ukuran	<i>Singleton</i>	<i>ARI seed</i>	<i>ARI LOO</i>
2	0.222	1.503	11.895	16, 15	Tidak	1.000	0.983
3	0.174	1.391	9.347	13, 11, 7	Tidak	0.966	0.965
4	0.182	1.425	8.394	10, 8, 7, 6	Tidak	0.874	0.841
5	0.185	1.197	8.016	8, 8, 7, 7, 1	Ya	0.733	0.754
6	0.190	1.156	7.797	7, 7, 7, 6, 3, 1	Ya	0.704	0.738

Tabel 4 memperlihatkan bahwa $k=2$ memberikan kompromi paling kuat: *silhouette* dan CHI tertinggi, *ARI seed* sebesar 1,000, *ARI leave-one-out* sebesar 0,983, ukuran kluster seimbang, serta tidak menghasilkan *singleton*. Nilai DBI yang lebih rendah pada $k=5$ dan $k=6$ tidak dipilih karena kedua solusi tersebut membentuk *singleton* dan memiliki stabilitas yang lebih rendah.



Gambar 1. Kurva *elbow* untuk $k=1-6$; penurunan *inertia* terbesar terjadi pada $k=2$.

Gambar 1 menunjukkan penurunan inertia paling tajam sebesar 29,09% dari k=1 ke k=2. Setelah titik tersebut, tambahan penurunan menjadi lebih kecil, yaitu 15,44% pada k=3 dan sekitar 12,75–13,71% pada kandidat berikutnya, sehingga siku kurva berada pada k=2.



Gambar 2. Perbandingan metrik validasi internal pada kandidat k=2–6.

Gambar 2 memperlihatkan bahwa k=2 unggul pada silhouette, CHI, dan stabilitas leave-one-out. Meskipun DBI cenderung membaik pada jumlah kluster yang lebih besar, k=5 dan k=6 menghasilkan singleton dan penurunan stabilitas; karena itu keputusan tidak didasarkan pada satu metrik saja.

3.3 Profil kluster final

Kluster A terdiri atas 15 kecamatan dan diberi label Pola Kepadatan–Air Minum–Keterbatasan Layanan. *Centroid*-nya menunjukkan kepadatan (+0,674), keterbatasan relatif Posyandu (+0,614), air minum tidak layak (+0,445), dan keterbatasan puskesmas (+0,441) lebih menonjol daripada rata-rata kabupaten. Sebaliknya, BBLR (−0,761) dan rumah tangga non-PHBS (−0,606) relatif lebih rendah.

Kluster B terdiri atas 16 kecamatan dan diberi label Pola BBLR–Perilaku Kesehatan. Dimensi yang paling menonjol adalah BBLR (+0,714) dan rumah tangga non-PHBS (+0,568). Kepadatan (−0,632), keterbatasan Posyandu (−0,576), air minum tidak layak (−0,417), dan keterbatasan puskesmas (−0,413) relatif lebih rendah. Sanitasi tidak aman berada dekat rata-rata pada kedua kluster (−0,041 dan +0,038), sehingga tidak menjadi pembeda utama.

Tabel 5. *Centroid Z-risk* kluster final

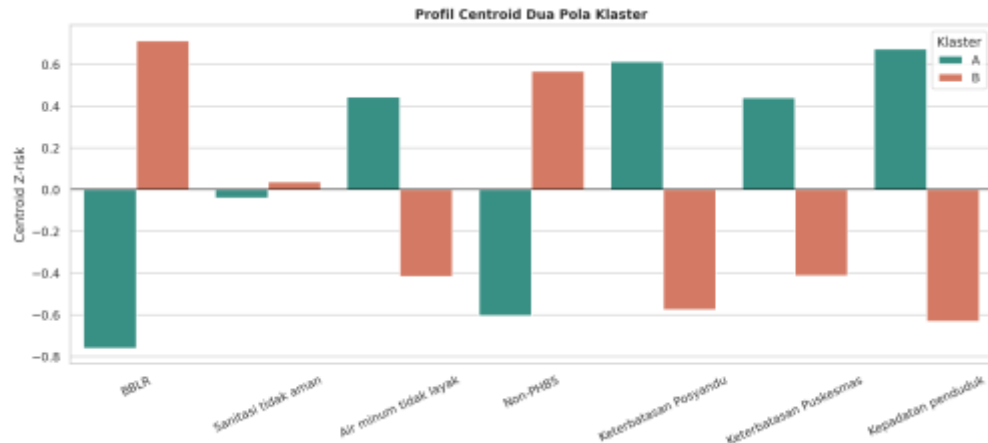
Kluster	BBLR	Sanitasi tidak aman	Air minum tidak layak	Rumah tangga non-PHBS	Keterbatasan Posyandu	Keterbatasan puskesmas	Log kepadatan penduduk
A	-0.761	-0.041	0.445	-0.606	0.614	0.441	0.674
B	0.714	0.038	-0.417	0.568	-0.576	-0.413	-0.632

Tabel 5 menegaskan arah pembeda kedua kluster. Kluster A memiliki centroid positif terbesar pada kepadatan, keterbatasan Posyandu, air minum tidak layak, dan keterbatasan puskesmas, sedangkan Kluster B memiliki centroid positif terbesar pada BBLR dan rumah tangga non-PHBS. Nilai sanitasi yang mendekati nol pada kedua kluster menunjukkan bahwa indikator ini bukan pembeda utama.

Tabel 6 menerjemahkan centroid terstandar ke ukuran yang lebih mudah dibaca. Dibandingkan Kluster B, Kluster A memiliki air minum tidak layak yang lebih tinggi (13,35% berbanding 3,27%), sedangkan Kluster B memiliki BBLR (4,47% berbanding 1,91%), rumah tangga non-PHBS (41,75% berbanding 28,49%), dan prevalensi stunting tertimbang (10,19% berbanding 7,76%) yang lebih tinggi.

Tabel 6. Ringkasan profil mentah dan target per klaster

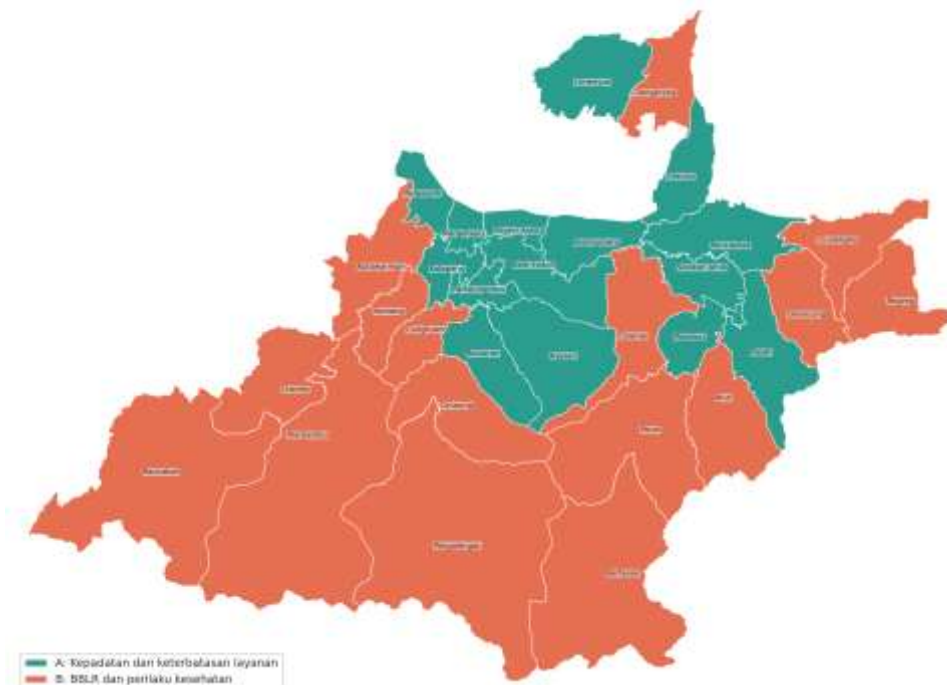
Klaster	Label	n	BBLR (%)	Air tak layak (%)	Non-PHBS (%)	Stunting tertimbang (%)
A	Pola Kepadatan–Air Minum–Keterbatasan Layanan	15	1.91	13.35	28.49	7.76
B	Pola BBLR–Perilaku Kesehatan	16	4.47	3.27	41.75	10.19



Gambar 3. Profil *centroid* Z-risk. Nilai positif menunjukkan dimensi yang relatif lebih menonjol dibanding rata-rata 31 kecamatan.

Gambar 3 memvisualisasikan arah centroid yang berlawanan pada dimensi utama. Batang Klaster A berada di atas rata-rata untuk kepadatan, keterbatasan layanan, dan air minum tidak layak, sedangkan batang Klaster B berada di atas rata-rata untuk BBLR dan non-PHBS; sanitasi berada dekat garis nol pada keduanya.

Peta Pola Klaster Risiko Stunting Kabupaten Bandung



Gambar 4. Persebaran dua pola kerentanan pada 31 kecamatan Kabupaten Bandung. Geometri: BIG edisi 2022; atribut analitis: 2024.

Gambar 4 menunjukkan bahwa Klaster A lebih banyak muncul pada koridor utara-tengah, sedangkan Klaster B mencakup banyak kecamatan di bagian selatan dan timur Kabupaten Bandung. Batas klaster tetap ditentukan oleh kemiripan tujuh indikator, bukan oleh kedekatan geografis antarkecamatan.

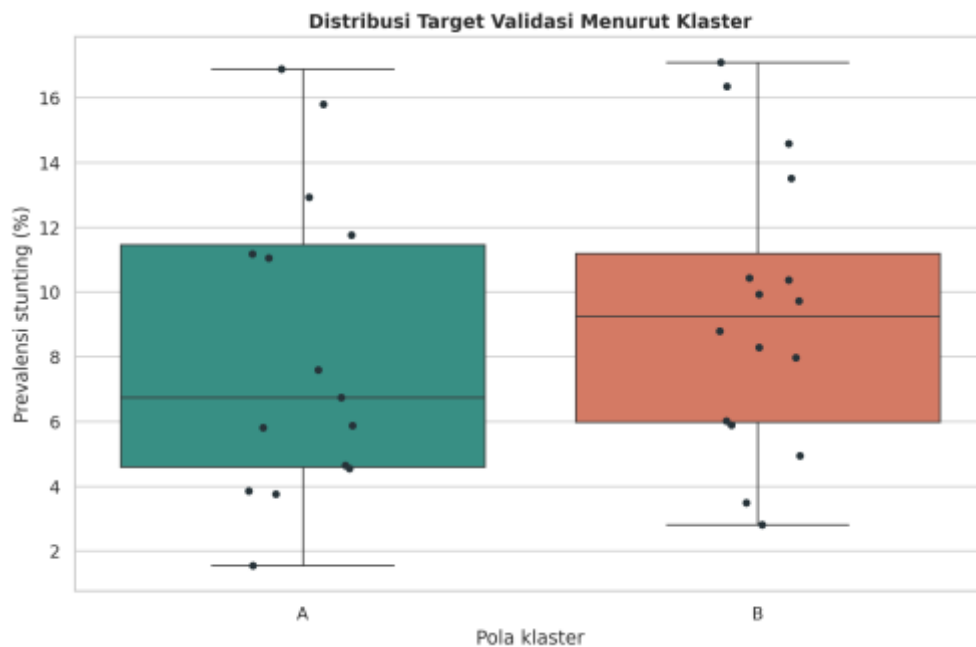
3.4 Validasi eksternal dengan prevalensi *stunting*

Klaster A memiliki 10.125 balita pendek dari 130.520 balita diukur, sedangkan Klaster B memiliki 11.237 dari 110.276. Prevalensi tertimbang Klaster B lebih tinggi 2,43 poin persentase daripada Klaster A. Namun, perbedaan distribusi prevalensi kecamatan tidak signifikan ($U=139,0$; $p=0,465$), dan Cliff's delta 0,158 menunjukkan efek kecil. Oleh karena itu, label klaster dipertahankan sebagai pola kerentanan, bukan kategori risiko tinggi dan rendah.

Tabel 7. Validasi eksternal target *stunting*

Kelompok	n	Mean (%)	Median (%)	Pembilang/penyebut atau uji	Tertimbang/efek
Klaster A	15	8,26	6,73	10.125/130.520	7,76
Klaster B	16	9,39	9,25	11.237/110.276	10,19
B-A / uji	–	–	–	$U=139,0$; $p=0,465$	+2,43 pp; delta=0,158

Tabel 7 menunjukkan bahwa median prevalensi *stunting* kecamatan pada Klaster B (9,25%) lebih tinggi daripada Klaster A (6,73%), dengan selisih prevalensi tertimbang 2,43 poin persentase. Namun, nilai $p=0,465$ dan Cliff's delta 0,158 menunjukkan bahwa perbedaan distribusi tersebut tidak signifikan dan besar efeknya kecil.



Gambar 5. Distribusi target prevalensi *stunting* menurut klaster. Target tidak digunakan saat pembentukan klaster.

Gambar 5 memperlihatkan median target yang lebih tinggi pada Klaster B, tetapi rentang dan sebaran titik kedua klaster saling tumpang tindih. Tumpang tindih tersebut konsisten dengan hasil Mann–Whitney yang tidak signifikan, sehingga grafik tidak mendukung pelabelan klaster sebagai risiko tinggi dan rendah secara definitif.

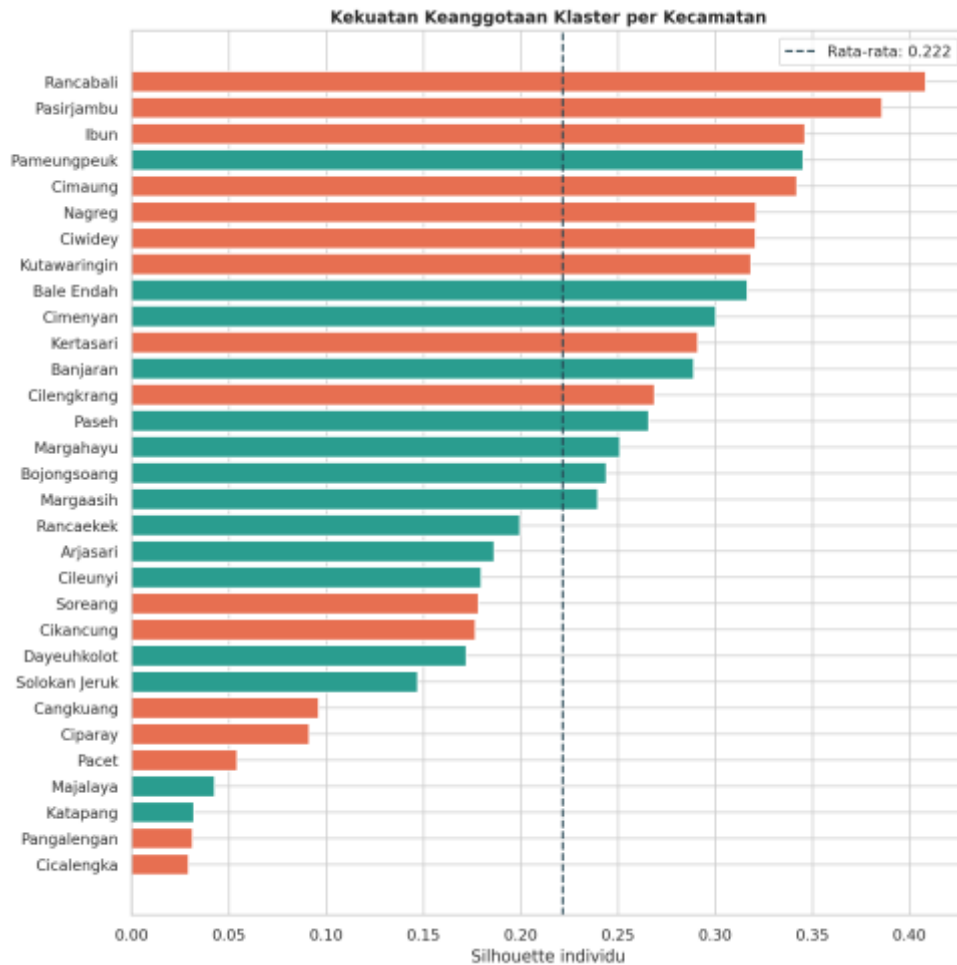
3.5 Kekuatan keanggotaan

Silhouette keseluruhan sebesar 0,222 menunjukkan pemisahan alami yang lemah. Tidak terdapat *silhouette* individu negatif, tetapi tujuh kecamatan memiliki nilai di bawah 0,10: Cicalengka, Pangalengan, Katapang, Majalaya, Pacet, Ciparay, dan Cangkuang. Kecamatan tersebut berada relatif dekat dengan batas antarklaster sehingga rekomendasi wilayah perlu dibaca bersama indikator mentahnya, bukan hanya label klaster.

Tabel 8. Kecamatan dengan keanggotaan relatif lemah

Kecamatan	Klaster	<i>Silhouette</i> individu	<i>Stunting</i> (%)
Cicalengka	B	0.029	13.51
Pangalengan	B	0.031	17.10
Katapang	A	0.032	11.05
Majalaya	A	0.043	6.73
Pacet	B	0.054	10.37
Ciparay	B	0.091	16.35
Cangkuang	B	0.096	5.88

Tabel 8 mengidentifikasi tujuh kecamatan dengan silhouette individu di bawah 0,10. Cicalengka, Pangalengan, dan Katapang memiliki nilai terendah, masing-masing 0,029, 0,031, dan 0,032, sehingga posisi ketiganya paling dekat dengan batas antarklaster dan perlu ditafsirkan bersama indikator mentah kecamatan.



Gambar 6. *Silhouette* individu pada solusi final $k=2$.

Gambar 6 menunjukkan variasi kekuatan keanggotaan antarkecamatan terhadap nilai rata-rata silhouette 0,222. Sebagian besar kecamatan memiliki nilai positif yang cukup jelas, sedangkan tujuh batang di bawah 0,10 menandai wilayah yang paling dekat dengan batas keputusan dan memerlukan interpretasi lebih hati-hati.

Hasil memperlihatkan bahwa kerentanan terkait *stunting* di Kabupaten Bandung tidak tersusun sebagai satu gradien sederhana. Klaster A lebih kuat dicirikan oleh konteks wilayah padat, ketidaklayakan air minum, dan keterbatasan relatif rasio layanan. Pola tersebut konsisten dengan literatur Indonesia yang menempatkan air, sanitasi, dan higiene sebagai bagian penting dari determinan pertumbuhan anak [4], [21]. Meta-analisis Indonesia juga melaporkan hubungan antara air minum tidak memadai dan sanitasi tidak memadai dengan *stunting* [6]. Akan tetapi, pada penelitian ini hubungan tersebut digunakan sebagai dasar interpretasi pola wilayah, bukan bukti kausal pada tingkat individu.

Klaster B lebih kuat dicirikan oleh BBLR dan rumah tangga non-PHBS. Temuan ini substantif karena BBLR berkaitan dengan kondisi prenatal dan pertumbuhan awal, sedangkan PHBS merepresentasikan praktik rumah tangga yang dapat memengaruhi paparan infeksi dan pemanfaatan layanan. Studi di Indonesia menunjukkan bahwa berat lahir rendah dan berbagai faktor prenatal/postnatal merupakan determinan penting *stunting* [22], sementara telaah nasional menegaskan sifat multidimensional determinan *stunting* [3]. Dengan demikian, Klaster B lebih tepat ditindaklanjuti melalui penguatan kesehatan ibu, pencegahan BBLR, pemantauan pertumbuhan, dan intervensi perilaku rumah tangga.

Centroid sanitasi tidak aman mendekati nol pada kedua klaster. Hal ini tidak berarti masalah sanitasi tidak penting. Nilai mendekati nol hanya menunjukkan bahwa indikator tersebut tidak membedakan dua kelompok secara kuat karena tingkatnya relatif serupa antar klaster. Oleh karena itu, peningkatan sanitasi aman sebaiknya diposisikan sebagai intervensi dasar lintas wilayah, sementara komponen lain disesuaikan dengan pola khas masing-masing klaster.

Prevalensi *stunting* tertimbang lebih tinggi pada Klaster B, tetapi hasil *Mann–Whitney* dan *Cliff's delta* tidak mendukung pemisahan risiko absolut yang kuat. Kondisi ini dapat terjadi karena *Clustering* mengelompokkan kombinasi faktor kerentanan, bukan memaksimalkan perbedaan target *stunting*. Selain itu, *stunting* memiliki banyak determinan yang tidak seluruhnya tersedia dalam *dataset*, termasuk konsumsi pangan, kemiskinan, pendidikan ibu, infeksi, dan faktor pengasuhan. Penelitian *multilevel* Indonesia menunjukkan bahwa faktor individual, rumah tangga, komunitas, dan wilayah dapat berkontribusi secara bersamaan [5].

Silhouette 0,222 menandakan tumpang tindih antar klaster sehingga hasil sebaiknya dipandang sebagai tipologi eksploratif. Meski demikian, $k=2$ menunjukkan kelebihan pragmatis: stabil terhadap 100 inisialisasi, hampir stabil sempurna pada *leave-one-out*, tidak menghasilkan *singleton*, dan mudah diterjemahkan menjadi dua paket intervensi. Dengan demikian, kestabilan solusi tidak boleh disamakan dengan pemisahan geometris yang kuat; keduanya memberikan informasi yang berbeda.

Pada Klaster A, prioritas dapat diarahkan pada pemetaan kantong akses air minum, penambahan atau redistribusi kapasitas layanan di wilayah padat, penguatan Posyandu, dan integrasi rujukan puskesmas. Pada Klaster B, prioritas dapat diarahkan pada pelayanan ibu hamil, pencegahan BBLR, pemantauan pertumbuhan, kunjungan rumah, dan penguatan PHBS. Sanitasi aman tetap menjadi paket minimum pada kedua kelompok. Untuk tujuh kecamatan dengan *silhouette* rendah, keputusan program sebaiknya menggunakan profil indikator kecamatan secara individual.

Dari sisi sistem informasi, hasil dapat diintegrasikan ke *dashboard* kecamatan dengan tiga lapis informasi: label pola kerentanan, nilai indikator mentah, dan tingkat kepastian keanggotaan. Pendekatan tersebut mengurangi risiko penggunaan label secara mekanis dan memudahkan pembaruan ketika data tahunan tersedia. Sedangkan keterbatasan penelitian ini adalah 1) Jumlah unit hanya 31 kecamatan sehingga daya pemisahan dan pengujian statistik terbatas. 2) Desain ekologis tidak mengizinkan inferensi mengenai risiko anak atau rumah tangga; menarik kesimpulan individual dari pola wilayah dapat menimbulkan *ecological fallacy* [23]. 3) Data sekunder dapat dipengaruhi perbedaan cakupan pencatatan; *denominator* PHBS hanya rumah tangga yang dipantau. 4) Indikator sanitasi merupakan komplemen terhadap akses sanitasi aman, sehingga tidak identik dengan ketiadaan seluruh bentuk sanitasi layak. 5) Nilai akses air minum yang sedikit melampaui *denominator* pada dua puskesmas dibatasi maksimum 100% setelah agregasi kecamatan dan dicatat dalam audit kualitas. 6) Data tidak mencakup seluruh determinan *stunting*, misalnya asupan pangan, status sosial ekonomi rinci, pendidikan ibu, dan riwayat infeksi. 7) Geometri administrasi beredisi 2022, sedangkan atribut analitis berperiode 2024; geometri hanya digunakan sebagai konteks visual. 8) Analisis satu tahun belum dapat menilai perubahan pola secara longitudinal atau kestabilan temporal.

5. Kesimpulan

Clustering K-Means terhadap tujuh indikator tahun 2024 menghasilkan dua pola kerentanan pada 31 kecamatan Kabupaten Bandung. Pola pertama menonjol pada kepadatan, air minum, dan keterbatasan relatif layanan; pola kedua menonjol pada BBLR dan perilaku kesehatan rumah tangga. Solusi $k=2$ stabil terhadap inisialisasi dan *leave-one-out*, tetapi *silhouette* keseluruhan rendah. Prevalensi *stunting* tertimbang lebih tinggi pada pola kedua, namun perbedaan distribusi antar klaster tidak signifikan dan besar efeknya kecil. Oleh karena itu, hasil paling tepat digunakan sebagai tipologi eksploratif untuk menyusun paket intervensi yang berbeda, bukan sebagai klasifikasi risiko tinggi dan rendah. Pemutakhiran data tahunan, penambahan determinan sosial-ekonomi, dan validasi dengan data individu atau desa direkomendasikan sebelum klaster digunakan untuk keputusan alokasi sumber daya yang berisiko tinggi.

Referensi

- [1] World Health Organization, "Stunting in a nutshell," 2015. [Online]. Available: <https://www.who.int/news/item/19-11-2015-stunting-in-a-nutshell>
- [2] Badan Kebijakan Pembangunan Kesehatan, "Survei Status Gizi Indonesia (SSGI) 2024 Dalam Angka," 2025.
- [3] T. Beal, A. Tumilowicz, A. Sutrisna, D. Izwardy, and L. M. Neufeld, "A review of child stunting determinants in Indonesia," *Matern. Child Nutr.*, vol. 14, no. 4, 2018, doi: 10.1111/mcn.12617.
- [4] H. Torlesse, A. A. Cronin, S. K. Sebayang, and R. Nandy, "Determinants of stunting in Indonesian children: Evidence from a cross-sectional survey indicate a prominent role for the water, sanitation and hygiene sector in stunting reduction," *BMC Public Health*, vol. 16, p. 669, 2016, doi: 10.1186/s12889-016-3339-8.
- [5] T. Mulyaningsih, "Beyond personal factors: Multilevel determinants of childhood stunting in Indonesia," 2021.
- [6] G. Gusnedi, R. D. Nindrea, I. Purnakarya, and H. B. Umar, "Risk factors associated with childhood stunting in Indonesia: A systematic review and meta-analysis," *Asia Pac. J. Clin. Nutr.*, vol. 32, no. 2, pp. 184–195, 2023, doi: 10.6133/apjcn.202306_32(2).0001.

- [7] E. Indrisari, "A systematic literature review on the application of machine learning for predicting stunting prevalence in Indonesia," 2025.
- [8] N. Izza, "Implementation of the K-Means clustering method on stunting case in Indonesia," 2019.
- [9] M. . S. Handayani M. F. L., "Performance analysis of clustering models based on machine learning in stunting data mapping," *JURTEKSI*, vol. 9, no. 4, pp. 715–720, 2023, doi: 10.33330/jurteks.v9i4.2770.
- [10] N. Hasdyna, "Hybrid machine learning for stunting prevalence," 2024.
- [11] S. Mulyati, "Stunting incidence segmentation: A cluster analysis approach and targeted intervention strategies," 2025.
- [12] F. Ramadhani, "Spatial clustering analysis of stunting in North Sumatra based on environmental factors using K-Means algorithm," 2025.
- [13] C. et al., "Clustering daerah rawan stunting di Indonesia menggunakan algoritma K-Means," 2025.
- [14] Badan Informasi Geospasial, "Administrasi AR Kecamatan 10K," 2022.
- [15] Dinas Kesehatan Kabupaten Bandung, "Profil Kesehatan Kabupaten Bandung Tahun 2024," 2025.
- [16] DPMD Kabupaten Bandung, "Jumlah Posyandu Dataset DPMD-075," 2024.
- [17] DPMD Kabupaten Bandung, "Jumlah Posyandu yang aktif Dataset DPMD-069," 2024.
- [18] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.
- [19] T. Liu, "Stability estimation for unsupervised clustering: A review," 2022.
- [20] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, and B. Thirion, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011, [Online]. Available: <https://jmlr.org/papers/v12/pedregosa11a.html>
- [21] J. H. Rah, "Improved sanitation is associated with reduced child stunting amongst Indonesian children under 3 years of age," 2020.
- [22] A. N. Sartika, "Prenatal and postnatal determinants of stunting at age 0–11 months: A cross-sectional study in Indonesia," 2021.
- [23] Y. C. T. Shih, "Ecological and individualistic fallacies in health disparities research," 2023.