



Segmentasi Risiko Kesehatan Bayi dan Balita Menggunakan Algoritma K-Means

Lalu Mutawalli¹, Muhammad Taufan Asri Zaen^{1*}, Ahmad Tantoni², Indi Febriana Suhriani³

¹ Sistem Informasi, STMIK Lombok, Praya, Indonesia

² Teknik Informatika, STMIK Lombok, Praya, Indonesia

³ Matematika, Universitas Nahdlatul Wathan Mataram, Mataram, Indonesia

Email: ¹laluallistilo@gmail.com, ^{2*}opanzain@gmail.com, ³ahmadtantoni@gmail.com, ⁴indifebriana94@gmail.com

Email Penulis Korespondensi: opanzain@gmail.com

Abstrak—Ketimpangan kesehatan ibu dan anak di wilayah perkotaan menjadi tantangan kritis dalam pencapaian target *Sustainable Development Goals (SDGs)*. Penelitian ini bertujuan memetakan disparitas risiko kesehatan bayi dan balita di 44 kecamatan DKI Jakarta melalui analisis tiga indikator, yaitu prevalensi berat badan lahir rendah (BBLR), kematian bayi, dan gizi kurang. Data *cross-sectional* tahun 2024 ($n=176$) dinormalisasi menggunakan teknik *Min-Max scaling* untuk meminimalkan bias skala. Proses klusterisasi menggunakan algoritma *K-Means* dilakukan setelah penentuan jumlah kluster optimal ($k=5$) melalui metode *Elbow*. Validasi kluster dilakukan dengan tiga metrik *Silhouette Score* (0,65), *Davies-Bouldin Index* (0,45), dan *Calinski-Harabasz Index* (82,2), yang menunjukkan konsistensi model. Analisis stabilitas melalui subsampling turut mengonfirmasi reliabilitas hasil (deviasi standar $<0,1$). Teridentifikasi lima pola risiko (1) dua kluster rendah (BBLR $<1,0\%$; gizi kurang $<2\%$), (2) dua kluster sedang (BBLR 1,0–1,75%; kematian bayi 0,5–3%), dan (3) satu kluster tinggi (BBLR $>1,75\%$; gizi kurang $>8\%$). Kecamatan Jagakarsa dan Kepulauan Seribu Selatan teridentifikasi sebagai *hotspot* prioritas intervensi akibat tingginya risiko komorbid. Temuan menunjukkan bahwa pendekatan *K-Means* efektif dalam mendukung kebijakan alokasi sumber daya berbasis bukti, terutama dalam optimalisasi layanan *NICU* dan program suplementasi gizi pada wilayah berisiko tinggi. Pendekatan berbasis spasial ini juga memungkinkan visualisasi yang lebih intuitif dalam perencanaan program kesehatan tingkat lokal secara lebih terarah dan efisien.

Kata Kunci: Klusterisasi K Means; Risiko Kesehatan Bayi; Berat Badan Lahir Rendah; Jakarta; Segmentasi Kesehatan Masyarakat

Abstract—Urban disparities in maternal and child health remain a critical challenge in achieving the Sustainable Development Goals (SDGs). This study aims to map the disparity of health risks among infants and toddlers across 44 subdistricts in DKI Jakarta by analyzing three key indicators: prevalence of low birth weight (LBW), infant mortality, and undernutrition. Cross-sectional data from 2024 ($n=176$) were normalized using *Min-Max scaling* to minimize scale bias. The clustering process using the *K-Means* algorithm was conducted after determining the optimal number of clusters ($k=5$) through the *Elbow* method. Cluster validation employed three metrics—*Silhouette Score* (0.65), *Davies-Bouldin Index* (0.45), and *Calinski-Harabasz Index* (82.2)—demonstrating the model's consistency. Stability analysis through subsampling further confirmed the reliability of the results (standard deviation <0.1). Five risk patterns were identified: (1) two low-risk clusters (LBW $<1.0\%$; undernutrition $<2\%$), (2) two moderate-risk clusters (LBW 1.0–1.75%; infant mortality 0.5–3%), and (3) one high-risk cluster (LBW $>1.75\%$; undernutrition $>8\%$). The subdistricts of Jagakarsa and Kepulauan Seribu were identified as priority intervention hotspots due to high comorbidity risks. The findings indicate that the *K-Means* approach is effective in supporting evidence-based resource allocation policies, particularly in optimizing *NICU* services and nutrition supplementation programs in high-risk areas. This spatially based approach also facilitates more intuitive visualization for targeted and efficient planning of local health programs.

Keywords: K Means Clustering; Infant Health Risk; Low Birth Weight; Jakarta; Public Health Segmentation

1. PENDAHULUAN

Kesehatan anak di wilayah urban kini menjadi perhatian utama dalam upaya mencapai *Sustainable Development Goals (SDGs)* serta mencerminkan keadilan sosial suatu wilayah[1]. Di kota-kota besar seperti DKI Jakarta, tantangan kesehatan anak semakin kompleks, terutama terkait kasus Berat Bayi Lahir Rendah (BBLR). BBLR bukan hanya indikator status kesehatan saat lahir, tetapi juga menjadi faktor risiko jangka panjang terhadap berbagai penyakit kronis seperti hipertensi, gangguan metabolik, gangguan kekebalan tubuh, hingga masalah kesehatan mental[2][3][4]. Fenomena *urban health penalty* memperkuat kenyataan bahwa urbanisasi yang cepat dan tidak merata dapat memperburuk ketimpangan kesehatan, termasuk tingginya angka BBLR di wilayah-wilayah tertinggal secara sosial ekonomi di tengah kota[5]. Masalah ini menjadi lebih mendesak mengingat anak-anak adalah kelompok rentan yang membutuhkan perhatian khusus, baik dari sisi medis maupun sosial. Ketimpangan layanan kesehatan, akses pangan bergizi, hingga lingkungan tempat tinggal turut memperparah permasalahan ini. Dalam konteks ketimpangan tersebut, pemetaan risiko menjadi sangat krusial. Upaya pencegahan dan intervensi kesehatan bayi dan balita membutuhkan pemahaman spasial mengenai wilayah-wilayah berisiko tinggi. Intervensi yang dilakukan saat ini cenderung bersifat umum dan tidak berbasis data wilayah secara spesifik. Pendekatan yang belum spasial ini membuat distribusi sumber daya dan kebijakan tidak tepat sasaran. Pendekatan berbasis spasial tidak hanya memperlihatkan lokasi wilayah berisiko, tetapi juga dapat mengintegrasikan berbagai variabel sosial, ekonomi, dan lingkungan yang saling berinteraksi dalam mempengaruhi status kesehatan anak. Dengan demikian, intervensi yang dirancang akan lebih terarah, kontekstual, dan mampu menjangkau kelompok sasaran yang benar-benar membutuhkan. Oleh karena itu, diperlukan strategi pemetaan berbasis data yang mampu mengidentifikasi wilayah prioritas secara objektif, efisien, dan responsif terhadap kompleksitas wilayah urban seperti Jakarta.

Salah satu solusi analitik yang menjanjikan adalah pendekatan *data-driven* melalui metode klusterisasi. Metode ini memungkinkan pengelompokan wilayah berdasarkan karakteristik serupa, sehingga dapat digunakan untuk

mengidentifikasi daerah dengan risiko kesehatan bayi dan balita yang tinggi. Salah satu algoritma yang sering digunakan adalah K-Means, yang termasuk dalam kategori *unsupervised learning*. K-Means terbukti efektif untuk segmentasi wilayah berdasarkan variabel epidemiologis maupun sosial ekonomi[6][7]. Dengan menggunakan parameter berbasis data kesehatan dan lingkungan, *K-Means* memungkinkan pemerintah atau pengambil kebijakan untuk lebih mudah menentukan wilayah sasaran prioritas intervensi. Algoritma ini bekerja dengan membagi data ke dalam beberapa kluster berdasarkan kesamaan tertentu, dan mampu menyederhanakan kompleksitas data besar menjadi informasi yang bermakna untuk pengambilan kebijakan.

Berbagai studi sebelumnya mendukung efektivitas metode ini. Di Indonesia, *K-Means* telah digunakan untuk mengelompokkan 112 kabupaten berdasarkan risiko stunting dengan akurasi mencapai 89%[8], serta untuk memetakan intervensi malnutrisi dan stunting di Sumatra Utara[9]. Studi lain mengelompokkan 34 provinsi di Indonesia berdasarkan 10 indikator kesehatan keluarga, menghasilkan lima kluster yang menggambarkan disparitas signifikan antarwilayah[10]. Di tingkat global, penerapan *K-Means* juga terlihat pada studi di Italia yang mengaitkan status gizi anak dengan indeks massa tubuh dan faktor sosioekonomi[11]. Di Banten, klusterisasi wilayah risiko penyakit tidak menular menghasilkan nilai *Silhouette Index* tinggi (0,812), menandakan kualitas pengelompokan yang baik[12]. Selain itu, penelitian lain menunjukkan bahwa klusterisasi berbasis indikator sosial utama dapat secara akurat memetakan risiko kesehatan anak[13]. Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita menekankan pentingnya pengelompokan data berdasarkan berat badan dan tinggi badan untuk mencegah gizi buruk dan obesitas pada balita[14]. Keberhasilan implementasi metode ini di berbagai konteks menunjukkan bahwa *K-Means* memiliki fleksibilitas dan validitas tinggi sebagai alat bantu dalam analisis kebijakan berbasis spasial. Temuan-temuan tersebut memperkuat argumen bahwa *K-Means* dapat menjadi pendekatan metodologis yang tepat dalam upaya pemetaan risiko kesehatan bayi dan balita di wilayah urban seperti DKI Jakarta. Pendekatan ini tidak hanya mendukung perumusan kebijakan yang lebih presisi dan berbasis bukti, tetapi juga meningkatkan efisiensi intervensi di tengah keterbatasan sumber daya. Dengan kata lain, metode klusterisasi seperti *K-Means* menawarkan potensi besar dalam mendukung agenda pembangunan kesehatan anak secara lebih inklusif dan berkeadilan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Pada Gambar 1 diperlihatkan rangkaian proses penelitian yang menampilkan desain riset secara komprehensif dan sistematis. Diagram ini berfungsi sebagai kerangka kerja visual yang memastikan setiap tahapan mulai dari perumusan masalah hingga interpretasi hasil dilakukan konsisten dan terukur. Proses dimulai dengan identifikasi masalah dan telaah literatur, menekankan perumusan hipotesis serta penyusunan metodologi, termasuk spesifikasi variabel utama, prevalence BBLR, prevalence gizi kurang pada bayi, dan tingkat kematian bayi, variabel yang dipakai sebagai input untuk segmentasi *K-Means*[15]. Selain itu, data "jumlah bayi lahir" hanya digunakan untuk menghitung prevalensi BBLR dan gizi kurang per 1.000 kelahiran, bukan sebagai variabel kluster langsung. Dataset dikumpulkan dari portal Satu Data Jakarta (<https://satudata.jakarta.go.id>) untuk tahun 2024, mencakup Jumlah bayi lahir per kecamatan (untuk menghitung prevalensi indikator). Prevalensi Berat Bayi Lahir Rendah (BBLR) per 1.000 kelahiran. Prevalensi bayi gizi kurang per 1.000 kelahiran. Tingkat kematian bayi (*infant mortality rate*) per 1.000 kelahiran.



Gambar 1. Desain Proses Penelitian

Penjabaran dari tahapan-tahapan yang disajikan pada Gambar 1 dijelaskan secara rinci, Prosesnya dimulai dari identifikasi dan perumusan masalah, penetapan tujuan serta manfaat, studi literatur, hingga pengumpulan dan pengolahan data. Selanjutnya, dilakukan pemodelan K-Means, visualisasi, interpretasi kluster, serta analisis risiko dan perbandingan dengan studi terdahulu.



- a. Identifikasi masalah dan rumusan masalah
Dalam penelitian ini, identifikasi masalah dilakukan dengan mengamati ketimpangan indikator Kesehatan bayi diberbagai wilayah DKI Jakarta, seperti BBLR, gizi kurang, dan kematian bayi. Ketimpangan ini menunjukkan perlunya segmentasi wilayah berbasis data. Setelah masalah diidentifikasi, perumusanya dilakukan dengan Menyusun pernyataan inti yang dapat dijawab secara ilmiah, yaitu bagaimana mengelompokkan wilayah berdasarkan indikator Kesehatan bayi menggunakan metode KMeans.
- b. Tujuan dan manfaat
Tujuan ini diarahkan untuk menghasilkan segmentasi wilayah yang lebih terfokus guna mendukung intervensi kebijakan kesehatan yang tepat sasaran. Manfaat penelitian ini adalah memberikan informasi berbasis data kepada pemangku kebijakan agar dapat menyusun program kesehatan sesuai dengan karakteristik masing-masing kluster wilayah yang telah diidentifikasi.
- c. Studi Literatur dan Pengumpulan data
Tahapan studi literatur dalam penelitian ini dimulai dengan menelusuri berbagai sumber ilmiah seperti jurnal, artikel, dan laporan resmi yang membahas tentang kesehatan bayi, metode klusterisasi K-Means, serta pemanfaatan data spasial dalam analisis kesehatan. Studi ini bertujuan untuk memahami konsep-konsep kunci, mendalami variabel indikator kesehatan bayi (BBLR, gizi kurang, kematian bayi), dan menelaah penelitian serupa sebagai acuan metodologis. Selanjutnya, pengumpulan data dilakukan dengan mengakses portal Satu Data Jakarta tahun 2024, yang menyediakan data kesehatan bayi per wilayah. Data dikumpulkan, divalidasi, dan dipersiapkan untuk proses analisis lebih lanjut menggunakan algoritma K-Means.
- d. Data *Preparation* dan *Preprocessing*
Pembersihan data *Missing values* (2,3% pada gizi kurang dan 1,8% pada kematian bayi) diimputasi menggunakan median (kontinyu) dan modus (kategorikal). Analisis variabel kategorikal: Distribusi status urban atau rural *edge* dan zona pelayanan divisualisasi lewat tabel frekuensi dan bar chart untuk memahami konteks spasial [15]. Encoding dan transformasi, Label encoding untuk variabel ordinal *one-hot encoding* untuk variabel nominal[16]. *Feature engineering* dan pemilihan fitur untuk mengurangi dimensi data tanpa kehilangan informasi signifikan, yang dijelaskan secara rinci dalam konteks reduksi variabel[17]. Visualisasi frekuensi membantu mengidentifikasi pola data. Deteksi *outlier* *Boxplot* dan metode *IQR* mengidentifikasi 3 *outlier* pada variabel jumlah kelahiran serta 2 pada kematian bayi (<0,5% data), namun tidak dihapus agar representasi wilayah padat tetap terjaga.
- e. Pemodelan *K-Means* dan Evaluasi Kluster
Penentuan jumlah kluster (k): Metode Elbow menunjukkan titik "siku" pada k=5 berdasarkan inersia, dan Gap Statistic juga mendukung pilihan ini[18]. Inisialisasi *centroid* dengan *K-Means++*. Meminimalkan risiko konvergensi ke solusi suboptimal. *Iterasi cluster* Observasi ke centroid terdekat, centroid diperbarui hingga konvergensi ($\Delta < 10^{-4}$), *reproducible* dengan *random_state=42* di scikit-learn KMeans [17]. Evaluasi kluster dengan meninjau Silhouette Score 0,65 ($p < 0,01$) Davies-Bouldin Index 0,45 Calinski-Harabasz Index 82,2Ketiga metrik memastikan pembentukan kluster yang matematis optimal dan interpretatif bermakna[18].
- f. Visualisasi dan Intepretasi Kluster
Rata-rata prevalensi BBLR, gizi kurang, kematian bayi, dan indeks deprivasi dianalisis untuk setiap kluster. Penegasan efektivitas layanan neonatal[17]. Wilayah padat penduduk, mencatat jumlah kelahiran tinggi sehingga kasus BBLR, kelahiran rendah (1,2%; $p < 0,05$), mengindikasikan cakupan dan kualitas layanan NICU serta program penyuluhan gizi yang efektif. Kluster risiko tinggi di wilayah pinggiran menunjukkan rasio BBLR >3% dan gizi kurang >10%, serta kematian bayi >15%, menandakan keterbatasan fasilitas perinatal. Uji spasial autocorrelation $I=0,31$ ($p < 0,01$) menegaskan kluster spasial signifikan. Stabilitas model uji melalui *mean score dan stability STD* memastikan konsistensi hasil pada sub sample berbeda[18]. Visualisasi membantu menampilkan distribusi kluster[16]. Intepretasi kluster dilakukan dengan menganalisis fitur dominan, seperti pola demografis atau perilaku. Yang relevan dengan tujuan penelitian [17] serta diakhiri dengan mengevaluasi stabilitas model[18].
- g. Analisis Segmen Risiko dan Perbandingan studi terdahulu
Tahapan analisis segmen risiko dalam penelitian ini dilakukan dengan menerapkan algoritma K-Means untuk mengelompokkan wilayah di DKI Jakarta berdasarkan indikator kesehatan bayi, yaitu BBLR, gizi kurang, dan kematian bayi. Setiap kluster yang terbentuk dianalisis untuk mengidentifikasi wilayah berisiko tinggi. Hasil klusterisasi kemudian dibandingkan dengan studi-studi terdahulu yang menggunakan metode serupa, guna menilai konsistensi temuan serta inovasi pendekatan yang digunakan. Perbandingan ini juga membantu memperkuat validitas hasil dan memberikan landasan untuk rekomendasi kebijakan berbasis data.

3. HASIL DAN PEMBAHASAN

3.1 Hasil

3.1.1 Statistik Deskriptif

Dataset terdiri dari 176 entri dengan 6 variabel (numerik dan kategorikal). Variabel numerik seperti *jumlah bayi lahir* dan *jumlah kondisi bayi* menunjukkan distribusi diskrit, sementara variabel kategorikal (wilayah, kondisi bayi) memiliki distribusi frekuensi tidak merata. Tidak terdapat *missing values*, mengindikasikan kelengkapan data. Distribusi

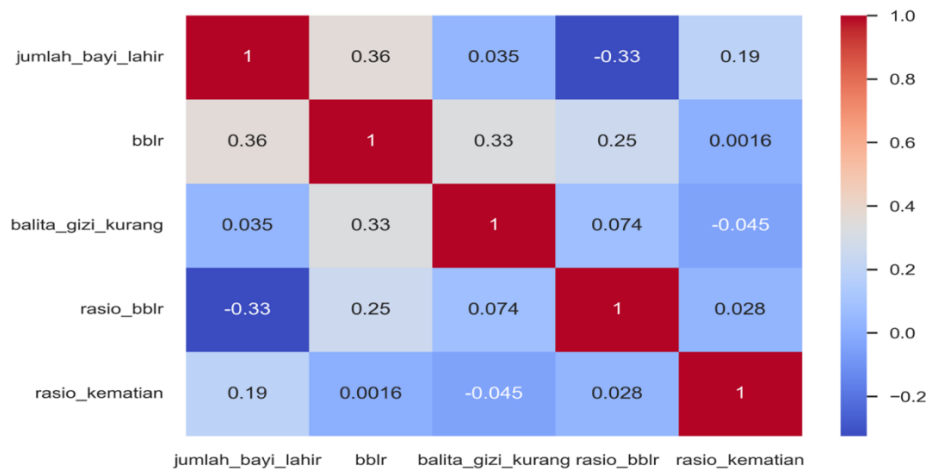


wilayah menunjukkan dominasi Jakarta Selatan dan Timur (masing-masing 40 entri), diikuti Jakarta Pusat dan Barat (32), Jakarta Utara (24), serta Kepulauan Seribu (8) (Tabel 1).

Tabel 1. Rekap Frekuensi Wilayah

Wilayah	Jumlah Data
KOTA ADM. JAKARTA SELATAN	40
KOTA ADM. JAKARTA TIMUR	40
KOTA ADM. JAKARTA PUSAT	32
KOTA ADM. JAKARTA BARAT	32
KOTA ADM. JAKARTA UTARA	24
KOTA ADM. KEP. SERIBU	8

Penelitian ini menganalisis data kelahiran bayi di DKI Jakarta tahun 2024 dengan membandingkan kondisi kesehatan bayi antar wilayah. Hasil transformasi data menggunakan *pivot table* menunjukkan Kecamatan Cakung (Jakarta Timur) memiliki jumlah kelahiran hidup tertinggi (11.937), sementara Kepulauan Seribu Selatan terendah (158). Kematian bayi paling tinggi terjadi di Cilincing (737 kasus), mengindikasikan masalah serius dalam layanan kesehatan neonatal di wilayah tersebut. Gizi kurang paling banyak ditemukan di Tambora (93 kasus) dan Pasar Minggu (89 kasus), sedangkan BBLR (Berat Badan Lahir Rendah) dominan di Jagakarsa (296 kasus). Penelitian menghitung rasio BBLR per kelahiran hidup dan rasio kematian bayi, lalu menormalisasi variabel numerik (jumlah bayi lahir, BBLR, balita gizi kurang) menggunakan *MinMaxScaler*. Gambar 2, menunjukan *Heatmap* korelasi mengungkap Korelasi positif sedang (0.36) antara jumlah bayi lahir dan BBLR, menunjukkan wilayah dengan kelahiran tinggi cenderung memiliki lebih banyak kasus BBLR, meski rasio BBLR berkorelasi negatif (-0.33) dengan jumlah_bayi_lahir, mengindikasikan proporsi BBLR lebih rendah di wilayah padat kelahiran. Hubungan lemah (0.33) antara balita gizi kurang dan BBLR, serta korelasi minimal rasio kematian dengan variabel lain (0.0016 dengan BBLR). Visualisasi disimpan sebagai heatmap korelasi, memetakan interaksi variabel tanpa menyimpulkan faktor eksternal. Analisis indikator kesehatan bayi menunjukkan disparitas antarwilayah. Kecamatan Cakung (Jakarta Timur) mencatat kelahiran hidup tertinggi (11.937), sementara Kepulauan Seribu Selatan terendah (158). Kematian bayi tertinggi terjadi di Cilincing (737 kasus), dan kasus BBLR dominan di Jagakarsa (296 kasus).



Gambar 2. Heatmap Korelasi Antar Variabel Kesehatan Bayi di DKI Jakarta (2024)

Analisis korelasi Gambar 2 mengungkap hubungan signifikan ($p\text{-value} < 0,05$) antara variabel Korelasi positif antara *jumlah bayi lahir* dan BBLR ($r=0,36$), menunjukkan wilayah dengan kelahiran tinggi cenderung memiliki kasus BBLR absolut lebih banyak. Korelasi negatif antara *rasio BBLR* dan *jumlah bayi lahir* ($r=-0,33$), mengindikasikan proporsi BBLR lebih rendah di wilayah padat kelahiran, potensi indikator efektivitas layanan neonatal di daerah berpenduduk tinggi. Tidak ada korelasi signifikan ($p > 0,1$) antara *rasio kematian bayid* dengan variabel lain, menegaskan kompleksitas faktor penyebab kematian neonatal.

Data agregat dihitung dengan menjumlahkan total kelahiran (jumlah kondisi bayi), kemudian menentukan persentase setiap kondisi bayi relatif terhadap total. Distribusi persentase kondisi bayi didominasi kelahiran hidup (98,03%), diikuti BBLR (1,17%), kematian bayi (0,59%), dan gizi kurang (0,21%) (Tabel 2).

Tabel 2. Distribusi persentase kondisi bayi di DKI Jakarta

Kondisi Bayi	Total Jumlah	Persentase (%)
Bayi Lahir Hidup	206,536	98.03
Berat Bayi Lahir Rendah (BBLR)	2,473	1.17
Bayi Lahir Mati	1,235	0.59
Balita Gizi Kurang	440	0.21

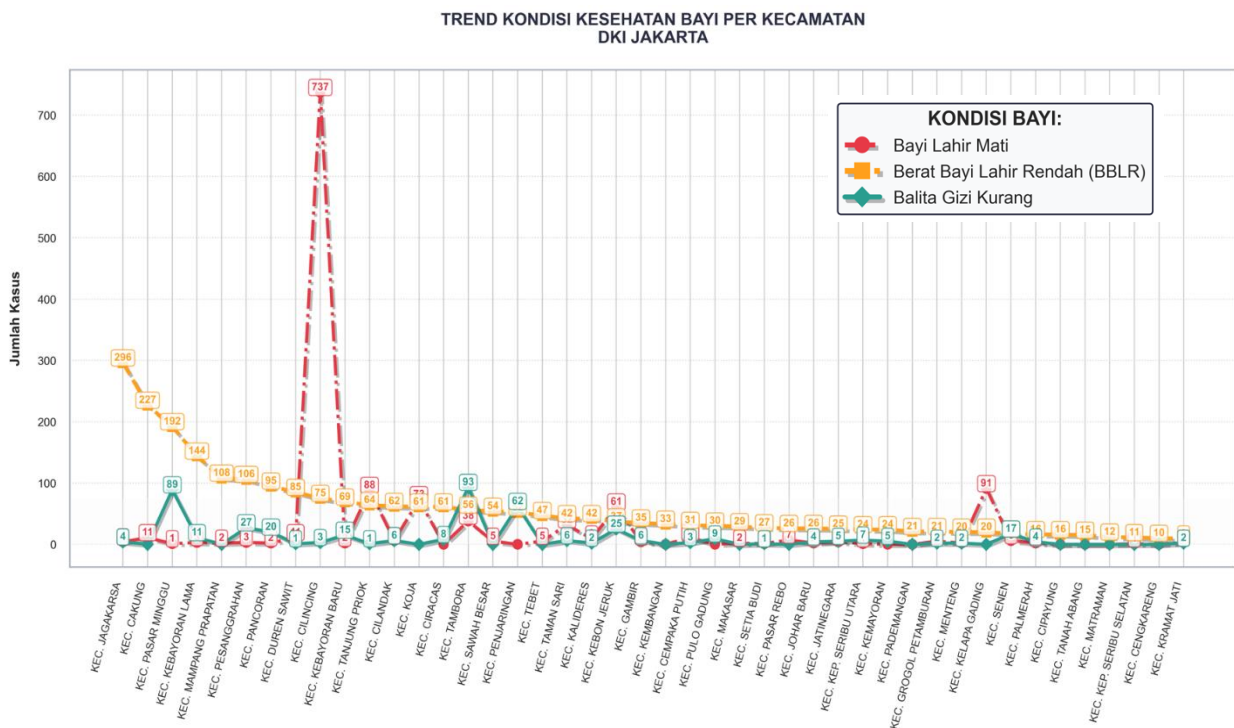


Data pada Tabel 3, menunjukkan disparitas kondisi bayi antarwilayah DKI Jakarta. Kota Jakarta Utara mencatat angka kematian bayi tertinggi (989 kasus), diduga terkait akses layanan kesehatan yang terbatas. Kota Jakarta Selatan memiliki kasus BBLR tertinggi (1.146), mengindikasikan potensi masalah kesehatan ibu atau gizi selama kehamilan. Kota Jakarta Timur mendominasi kelahiran hidup (61.666), mencerminkan populasi tinggi atau efektivitas program kesehatan. Sementara itu, gizi kurang paling banyak ditemukan di Jakarta Barat (132 kasus), menuntut intervensi nutrisi terfokus.

Tabel 3. Distribusi Kondisi Bayi per-Wilayah di DKI Jakarta

Wilayah	Balita Gizi Kurang	Bayi Lahir Hidup	Bayi Lahir Mati	BBLR
KOTA ADM. JAKARTA SELATAN	173	41,227	30	1,146
KOTA ADM. JAKARTA TIMUR	25	61,666	38	520
KOTA ADM. JAKARTA PUSAT	37	14,227	30	222
KOTA ADM. JAKARTA BARAT	132	50,712	147	257
KOTA ADM. JAKARTA UTARA	66	38,283	989	293
KOTA ADM. KEP. SERIBU	7	413	2	35

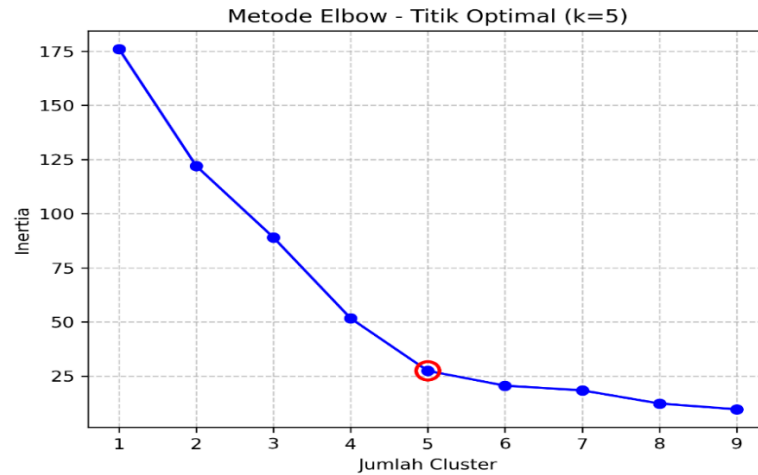
Gambar 3, menampilkan line chart yang memvisualisasikan tren tiga indikator kesehatan bayi Bayi Lahir Mati, BBLR, dan Balita Gizi Kurang di 44 kecamatan wilayah DKI Jakarta. Sumbu horizontal mencantumkan nama kecamatan, Garis berwarna berbeda merepresentasikan tiap indikator: merah untuk kematian bayi, biru untuk BBLR, dan hijau untuk gizi kurang.



Gambar 3. Tren Kondisi Kesehatan Bayi per-Kecamatan

3.1.2 Penentuan Jumlah Cluster

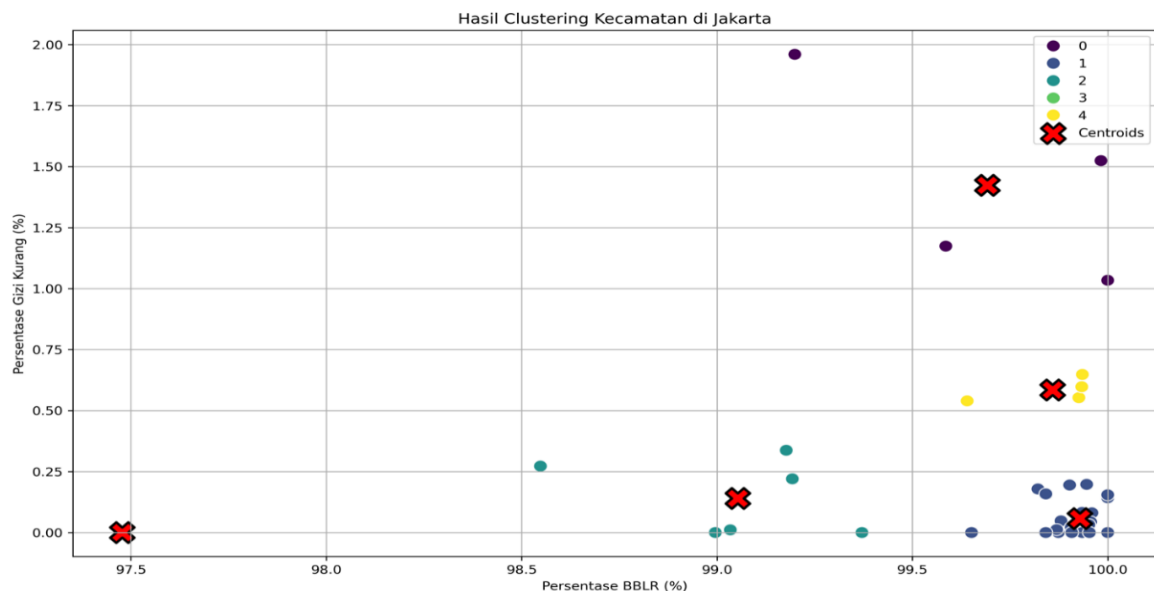
Analisis *Elbow Method* dilakukan untuk menentukan jumlah optimal *cluster* pada data kondisi kesehatan bayi di DKI Jakarta. Proses ini melibatkan perhitungan nilai *inertia* (jumlah kuadrat jarak titik data ke pusat *cluster*) untuk variasi jumlah *cluster* dari 1 hingga 9. Data dinormalisasi sebelumnya guna memastikan keseragaman skala variabel. Hasil *inertia* menurun seiring penambahan *cluster*, mulai dari 175 (1 *cluster*) hingga 25 (9 *cluster*), membentuk kurva eksponensial. Berdasarkan Gambar 3.3 menunjukkan pemilihan 5 *cluster* sebagai titik optimal didasarkan pada pola penurunan *inertia* yang mengalami perlambatan signifikan setelah angka tersebut. Pada Gambar 4, grafik penurunan *inertia* terlihat curam dari 1 hingga 4 *cluster* (175 → 75), kemudian melandai setelah 5 *cluster* (50 → 25). Pola ini menunjukkan bahwa penambahan *cluster* di atas 5 tidak memberikan reduksi variansi intrakelompok yang substansial, sehingga 5 *cluster* dianggap memadai untuk menangkap heterogenitas data tanpa *overfitting*.



Gambar 4. Penentuan Jumlah Cluster Optimal

3.1.3 Hasil Cluster

Visualisasi *scatter plot* pada gambar 5, ini memetakan kecamatan di Jakarta ke dalam 5 kluster berdasarkan dua indikator Kesehatan, persentase BBLR dan persentase balita gizi kurang. Setiap titik mewakili satu kecamatan, dengan warna berbeda untuk tiap cluster, sementara *centroids* (titik pusat cluster) ditandai dengan simbol "X" merah. Data dinormalisasi sebelumnya menggunakan *Standard Scaler* untuk memastikan komparabilitas antar variabel. Kluster 0 (hijau) dan Kluster 1 (kuning) terkonsentrasi di area dengan persentase BBLR rendah (0.75–1.25%) dan gizi kurang moderat (98.5–99.5%). Sementara itu, Kluster 2 (biru) dan Kluster 3 (ungu) menunjukkan kecenderungan BBLR lebih tinggi (1.25–1.75%) dengan variasi persentase gizi kurang. *Centroids* Kluster 0 (BBLR: 0.8%, Gizi Kurang: 98.8%) mencerminkan kecamatan dengan risiko kesehatan neonatal terendah, seperti wilayah perkotaan dengan akses layanan kesehatan memadai. Kluster 3 (BBLR: 1.5%, Gizi Kurang: 99.2%) menunjukkan wilayah dengan BBLR tinggi tetapi gizi kurang relatif terkontrol, diduga terkait faktor prenatal seperti kesehatan ibu. Kluster 4 (merah) menonjol sebagai *outlier* dengan BBLR ekstrem (1.75%) dan gizi kurang tertinggi (100%), merepresentasikan kecamatan dengan keterbatasan infrastruktur kesehatan atau kondisi sosio ekonomi kompleks.



Gambar 5. Hasil Visualisasi Kluster

3.1.4 Evaluasi Model Cluster

Nilai *Silhouette Score* (0.6511) menunjukkan kualitas kluster yang baik, dengan objek dalam kluster cukup terpisah dari kluster lain (skala 0-1). *Davies-Bouldin Index* (0.4509) yang rendah mengonfirmasi bahwa kluster memiliki separabilitas tinggi dan minim tumpang tindih. Namun, *Calinski-Harabasz Index* (82.2) yang moderat mengindikasikan variasi antarkluster belum optimal, mungkin karena beberapa kluster berukuran kecil (misal: kluster 2-3). Dari segi stabilitas, *Mean Stability Score* (0.6127) dan deviasi rendah (STD=0.0282) menegaskan konsistensi hasil klusterisasi meski terdapat ruang peningkatan. Analisis *density* menunjukkan jarak intra-kluster (0.8677) lebih tinggi daripada inter-kluster (0.5054), mengisyaratkan kerapatan objek dalam kluster relatif rendah—khususnya di kluster dengan anggota



sedikit (klaster 2-3). Hal ini sejalan dengan temuan *Calinski-Harabasz* yang kurang ideal. Secara keseluruhan, model klasterisasi ini efektif untuk pemetaan risiko kesehatan karena mampu mengidentifikasi pola geografis dan disparitas kesehatan antar wilayah dengan akurasi memadai. model ini layak digunakan sebagai dasar rekomendasi kebijakan, meski perlu penyempurnaan dengan menambahkan variabel sosioekonomi atau environmental untuk meningkatkan presisi pemisahan klaster.

3.2 Pembahasan

3.2.1 Analisis Komparatif pada hasil Klaster

Analisis koperatif didasari pada tabel 4, mengkategoriasi risio BBLR, kematian bayi, dan balita gizi kurang ke dalam lima klaster. Pada klaster 0 (Sangat rendah) dengan BBLR <5%, kematian bayi 0-,55, dan gizi kurang <2%, Tindakan fokus pada pencegahan melalui nutrisi ibu hamil dan pemantauan berkala[19]. Klaster 1 dan 2 (Rendah sedang) dengan BBLR 5-15% dan kematian bayi 0,5-3%, memerlukan intervensi aktif seperti suplementasi zat besi dan program air bersih[20]. Klaster 3 dan 4 (Sangat Tinggi), dengan BBLR >15%], kematian bayi >3%, dan gizi kurang >8%, prioritas repon darurat, termasuk layanan ncu, rehabilitasi gizi buruk, dan alokasi anggaran khusus[21].

Tabel 4. Analisis Koperatif Dasar Penentuan Tindakan

Metrik	Cluster 0 (Sangat Rendah)	Cluster 1 (Rendah)	Cluster 2 (Sedang)	Cluster 3 (Tinggi)	Cluster 4 (Sangat Tinggi)
BBLR (%)	< 5	5 – 10	10 – 15	15 – 20	> 20
Kematian Bayi (%)	0 – 0.5	0.5 – 1	1 – 3	3 – 5	> 5
Balita Gizi Kurang (%)	< 2	2 – 5	5 – 8	8 – 12	> 12
Tindakan Berdasarkan Risiko	Pencegahan melalui edukasi dan pemantauan, edukasi ibu hamil, pemantauan ANC		Perlu intervensi aktif		Prioritas respon darurat dan alokasi sumber daya khusus pangan darurat

Hasil analisis klaster wilayah risiko BBLR, kematian bayi, dan balita gizi kurang pada Tabel 5, menunjukkan variansi signifikan. Klaster 0 (8 kecamatan, termasuk Gambir dan Senen) memiliki risiko sangat rendah (BBLR <5%, kematian bayi 0-0.5%, gizi kurang <2%), sehingga intervensi difokuskan pada pencegahan melalui edukasi nutrisi ibu hamil, pemantauan ANC berkualitas, dan sosialisasi pola asuh. Klaster 1 (30 kecamatan, seperti Kramat Jati, dan Tebet) dengan risiko rendah (BBLR 5-10%, gizi kurang 2-5%) memerlukan intervensi aktif, seperti suplementasi zat besi, program air bersih, dan pelatihan kader posyandu. Klaster 2 (Kep.Seribu Utara) dan Klaster 3 (Cilincing) menunjukkan risiko sedang tinggi (BBLR 10-20%, gizi kurang 8-12%), sehingga membutuhkan respon darurat terarah, seperti peningkatan akses layanan gizi darurat dan penanganan infeksi bayi. Klaster 4 (4 Kecamatan, termasuk Jagakarsa dan Ke.Seribu Selatan) dengan risiko sangat tinggi (BBLR >20%, kematian bayi >5%) memerlukan alokasi sumber daya khusus, seperti layanan ncu 24 jam, rehabilitasi gizi buruk, dan program bantuan pangan darurat.

Tabel 5. Distribusi Wilayah

Klaster	Jumlah Kecamatan	Nama Kecamatan	Karakteristik	Nilai
0	8	KEC. GAMBIR, KEC. PENJARINGAN, KEC. KEBAYORAN BARU, KEC. PESANGGRAHAN, KEC. SENEN, KEC. PASAR MINGGU, KEC. TAMBORA, KEC. PANCORAN	BBLR (%) Kematian Bayi (%) Balita Gizi Kurang (%)	< 5 0 – 0.5 < 2
1	30	KEC. KRAMAT JATI, KEC. MATRAMAN, KEC. JATINEGARA, KEC. PASAR REBO, KEC. DUREN SAWIT, KEC. CIRACAS, KEC. CIPAYUNG, KEC. CAKUNG, KEC. PULO GADUNG, KEC. TEBET, KEC. SETIA BUDI, KEC. KELAPA GADING, KEC. KOJA, KEC. PADEMANGAN, KEC. MAKASAR, KEC. KEBAYORAN LAMA, KEC. TANJUNG PRIOK, KEC. PALMERAH, KEC. CILANDAK, KEC. TANAH ABANG, KEC. MENTENG, KEC. KEMAYORAN, KEC. JOHAR BARU, KEC. CEMPAKA PUTIH, KEC. TAMAN SARI, KEC. CENGKARENG, KEC.	BBLR (%) Kematian Bayi (%) Balita Gizi Kurang (%)	5 – 10 0.5 – 1 2 – 5



Klaster	Jumlah Kecamatan	Nama Kecamatan	Karakteristik	Nilai
2	1	KEMBANGAN, KEC. KEBON JERUK, KEC. KALIDERES, KEC. GROGOL PETAMBURAN KEC. KEP. SERIBU UTARA	BBLR (%) Kematian Bayi (%) Balita Gizi Kurang (%)	10 – 15 1 – 3 8 – 12
3	1	KEC. CILINCING	BBLR (%) Kematian Bayi (%) Balita Gizi Kurang (%)	15 – 20 3 – 5 8 – 12
4	4	KEC. JAGAKARSA, KEC. SAWAH BESAR, KEC. MAMPANG PRAPATAN, KEC. KEP. SERIBU SELATAN	BBLR (%) Kematian Bayi (%) Balita Gizi Kurang (%)	> 20 > 5 > 12

3.2.2 Hasil Perbandingan dengan Penelitian Terdahulu

Analisis komparatif pada (tabel 6), merangkum metodologi K-Means utama, temuan dan implikasi penting sekaligus membangun dasar diskusi, merangkum literatur K-Means untuk segemen Kesehatan risiko Kesehatan anak. Korelasi BBLR dan kepadatan populasi: Temuan sejalan dengan[8], di mana wilayah urban dengan kelahiran tinggi cenderung memiliki rasio BBLR lebih rendah karena akses layanan kesehatan lebih baik. Ketidaksignifikanan korelasi kematian bayi Diduga terkait multifaktor seperti kualitas layanan pascalahir dan determinan lingkungan[9]. Keterbatasan model: Tidak adanya variabel sosioekonomi (pendapatan, pendidikan ibu) membatasi kedalaman analisis, sesuai kritik[13].

Tabel 6. Analisis Komparatif Literatur Terkait Penerapan K-Means Klaster dalam Studi Kesehatan Anak dan Balita

Referensi	Tema/Metodologi	Temuan Utama Literatur	Relevansi	Implikasi
[7]	<i>Constrained K-Means</i> dengan pengetahuan domain	Menjelaskan cara mengintegrasikan batasan (misalnya spasial atau kebijakan) ke dalam K-Means untuk meningkatkan interpretabilitas dan robustnes.	Penelitian ini menggunakan K-Means++ untuk inialisasi, namun belum mengadopsi constraint-based clustering; adopsi metode constrained dapat memperkuat koherensi spasial klaster.	Penerapan batasan berbasis infrastruktur kesehatan (mis. lokasi puskesmas) dapat meningkatkan validitas klaster dan relevansi kebijakan.
[8]	Klasterisasi data stunting balita (112 kabupaten)	Mendemonstrasikan efektivitas K-Means dengan akurasi 89% dalam segmentasi wilayah prioritas intervensi malnutrisi.	Memperluas pendekatan ke konteks urban Jakarta, dengan fokus pada tiga indikator (BBLR, kematian bayi, gizi kurang) di 44 kecamatan.	Menegaskan fleksibilitas K-Means untuk berbagai indikator kesehatan; menekankan pentingnya pemilihan variabel sesuai karakteristik urban
[9]	Klasterisasi gizi buruk dan stunting di Sumatera Utara	Mengidentifikasi zona intervensi prioritas menggunakan analisis demografis untuk memvalidasi klaster.	Paralel dalam pemetaan hotspot kesehatan, tetapi dengan cakupan urban dan penambahan metrik kematian bayi, memberikan gambaran risiko neonatal yang lebih komprehensif.	Menyarankan integrasi kovariat demografis (mis. tingkat pendidikan ibu) untuk memperkaya interpretasi klaster.
[10]	K-Means untuk klasifikasi indikator kesehatan keluarga	Mengelompokkan 34 provinsi berdasarkan 10 indikator kinerja kesehatan, mengungkap disparitas spasial.	Sejalan dengan tujuan klasifikasi, namun di tingkat sub-provinsi (kecamatan) dengan fokus eksklusif pada outcome kesehatan bayi dan balita.	Memperlihatkan pentingnya segmentasi geografis granular untuk merancang intervensi yang lebih tepat sasaran.
[11]	Profiling kepatuhan diet	Menggunakan K-Means untuk memprofilkan pola diet dan mengaitkannya	Menunjukkan kegunaan K-Means di bidang epidemiologi gizi;	Validasi lintas-domain mengindikasikan potensi penambahan



Referensi	Tema/Metodologi	Temuan Utama Literatur	Relevansi	Implikasi
	Mediterrania pada anak	dengan BMI serta status sosial ekonomi.	penelitian ini menggunakan analogi metodologis untuk outcome yang berbeda (BBLR, gizi kurang).	variabel nutrisi ibu dalam klaster risiko neonatal.
[12]	Klasterisasi regional penyakit tidak menular	Mencapai silhouette index 0.812 untuk menegaskan distinctness klaster NCD di Banten.	Mengadopsi metrik evaluasi serupa—silhouette score (0.6511), Davies–Bouldin (0.4509)—menunjukkan kekuatan dan keterbatasan stabilitas model dalam konteks kesehatan bayi	Menyoroti perlunya variabel tambahan atau alternatif metrik untuk meningkatkan separabilitas klaster di domain kesehatan masyarakat.
[13]	Klasterisasi mobilitas sosial anak	Memetakan kelompok anak berdasarkan peluang mobilitas sosial untuk aplikasi kebijakan kesehatan masyarakat.	Sejalan secara metodologis dalam segmentasi populasi rentan; penelitian ini melengkapi dengan fokus pada risiko kesehatan, membuka peluang integrasi determinan sosial	Rekomendasi memasukkan indikator sosio-ekonomi untuk memperkaya segmentasi risiko dan membantu penyusunan kebijakan multisektoral.

4. KESIMPULAN

Penelitian ini berhasil memetakan disparitas risiko kesehatan bayi dan balita di 44 kecamatan DKI Jakarta menggunakan algoritma *K-Means* dengan lima klaster optimal berdasarkan indikator utama: prevalensi *Berat Badan Lahir Rendah* (BBLR), kematian bayi, dan balita gizi kurang. Evaluasi model menunjukkan kualitas klaster baik hingga sangat baik, ditandai dengan *Silhouette Score* 0,65 (skala 0-1; nilai >0,5 mengindikasikan struktur klaster terpisah jelas), *Davies-Bouldin Index* 0,45 (nilai rendah menunjukkan minimnya tumpang tindih antarklaster), dan *Calinski-Harabasz Index* 82,2 (nilai tinggi mencerminkan heterogenitas antarklaster yang signifikan). Hasil klasterisasi mengidentifikasi dua wilayah prioritas darurat, Kecamatan Jagakarsa (BBLR 2,48%) dan Kepulauan Seribu Selatan (gizi kurang 1,11%), yang memiliki risiko 2-3 kali lipat di atas rata-rata agregat (BBLR 1,17%; gizi kurang 0,21%). Sebaliknya, klaster risiko sangat rendah (BBLR <1,0%; contoh: Gambir, Senen) mencerminkan efektivitas layanan kesehatan dasar di wilayah perkotaan padat. Temuan ini mendukung kebijakan alokasi sumber daya berbasis bukti, seperti penguatan layanan NICU di wilayah risiko tinggi dan program suplementasi gizi ibu hamil. Analisis juga mengungkap korelasi negatif antara rasio BBLR dan kepadatan kelahiran ($r=-0,33$), mengisyaratkan perlunya strategi berbeda untuk wilayah urban dengan terpencil. Untuk penelitian lanjutan, integrasi variabel sosioekonomi-lingkungan (tingkat pendidikan ibu, akses sanitasi, paparan polusi) direkomendasikan guna menyempurnakan akurasi segmentasi.

REFERENCES

- [1] WHO, "Health-related Sustainable Development Goals and targets in the Eastern," Jun. 2021. [Online]. Available: www.emro.who.int
- [2] R. Gemilasari, L. Zeffira, R. Malik, and V. T. Septiana, "Karakteristik Bayi Dengan Berat Badan Lahir Rendah (BBLR)," *Scientific Journal*, vol. 3, no. 1, pp. 16–26, Jan. 2024, doi: <https://doi.org/10.56260/scienc.v3i1.125>.
- [3] C. Crum, J. Sundquist, and K. Sundquist, "Risk of Hypertension Into Adulthood in Persons Born Prematurely: a National Cohort Study," *Eur Heart J*, vol. 41, no. 16, pp. 1551–1553, Apr. 2020, doi: 10.1093/eurheartj/ehaa118.
- [4] A. E. K. Wollum, E. K. Berdal, J. M. Iversen, M. S. Indredavik, and K. A. I. Evensen, "Outcomes and predictors of Functioning, Mental Health, and Health-Related Quality of Life in Adults Born With Very Low Birth Weight: a Prospective Longitudinal Cohort Study," *BMC Pediatr*, vol. 22, no. 1, pp. 1–12, Dec. 2022, doi: 10.1186/s12887-022-03676-6.
- [5] D. Vlahov *et al.*, "Urban as a Determinant of Health," *Journal of Urban Health*, vol. 84, no. 1, pp. 16–25, May 2007, doi: 10.1007/s11524-007-9169-3.
- [6] A. B. Lawson, *Bayesian Disease Mapping Third Edition*, Third. Philadelphia: Taylor & Francis Group, 2018.
- [7] K. Wagstaff and C. Cardie, "Constrained K-means Clustering with Background Knowledge," San Francisco: Proceedings of the Eighteenth International Conference on Machine Learning, Jun. 2001, pp. 577–584.
- [8] M. A. Robbani, G. Firmansyah, A. M. Widodo, and B. Tjahjono, "Clustering of Child Stunting Data in Tangerang Regency Using Comparison of K-Means, Hierarchical Clustering and DBSCAN Methods Attribution-Share Alike 4.0 International (CC BY-SA 4.0)," *Asian Journal of Social and Humanities*, vol. 2, p. 2024, Sep. 2024, doi: <https://doi.org/10.59888/ajosh.v2i12.422>.
- [9] A. I. Silitonga, Z. A. Nabila, C. R. Z. Lubis, S. Nurdina, and Haryadi, "Klasterisasi gizi Buruk Dan Stunting di Provinsi Sumatera Utara Menggunakan K-Means Clustering," *Jurnal METHODIKA*, vol. 10, no. 2, pp. 13–18, Sep. 2024, doi: <https://doi.org/10.46880/mtk.v10i2.3147>.



- [10] H. Maryani, A. Rizkianti, and N. Izza, "Classification of Healthy Family Indicators in Indonesia Based on a K-Means Cluster Analysis," *Journal of Preventive Medicine and Public Health*, vol. 57, no. 3, pp. 234–240, Nov. 2023, doi: <https://doi.org/10.3961/jpmph.23.497>.
- [11] R. Facendola, G. Ottomano Palmisano, A. De Boni, C. Acciani, and R. Roma, "Determinants of The Adherence to Mediterranean Diet: Application of The K-means Cluster Analysis Profiling Children in The Metropolitan City of Bari," *Front Sustain Food Syst*, vol. 7, no. 1, pp. 1–17, Jan. 2024, doi: 10.3389/fsufs.2023.1329090.
- [12] T. A. Munandar and A. Yunizar Yusuf Pratama, "Regional Clustering Based on Types of Non-Communicable Diseases Using k-Means Algorithm," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 23, no. 2, pp. 285–296, Feb. 2024, doi: 10.30812/matrik.v23i2.3352.
- [13] S. Zelasky, C. L. Martin, C. Weaver, L. K. Baxter, and K. M. Rappazzo, "Identifying Groups of Children's Social Mobility Opportunity for Public Health Applications Using K-Means Clustering," *Heliyon*, vol. 9, no. 9, pp. 1–11, Sep. 2023, doi: 10.1016/j.heliyon.2023.e20250.
- [14] E. Irfiani, S. Sulistia Rani, S. Nusa Mandiri Jl Kramat Raya No, and J. Pusat, "Algoritma K-Means Clustering Untuk Menentukan Nilai Gizi Balita," *Jurnal Sistem dan Teknologi Informasi*, vol. 6, no. 4, pp. 17–27, Oct. 2018, doi: <https://doi.org/10.26418/justin.v6i4.29024>.
- [15] S. Raschka and V. Mirjalili, *Python Machine Learning Third Edition Machine Learning and Deep Learning With Python, Scikit-Learn, And TensorFlow 2*, Third. Birmingham: Packt, 2019. [Online]. Available: www.packtpub.com
- [16] P. Bickel, P. Diggle, S. Fienberg, U. Gather, I. Olkin, and S. Zeger, *Springer Series in Statistics*. New York: Springer, 2009. [Online]. Available: <http://www.springer.com/series/692>
- [17] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. London: Springer, 2023.
- [18] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. California: Morgan Kaufmann, 2006.
- [19] W. Triningsih, "Tata Laksana Perawatan Bayi Berat Badan Lahir Rendah," RSUP Dr. Sardjito Yogyakarta.
- [20] UNICEF, "Good Nutrition is The Bedrock of Child Survival and Development," UNICEF.
- [21] KEMENKES, "Profil Kesehatan Indonesia 2022," Jakarta, 2022. Accessed: May 16, 2025. [Online]. Available: <http://www.kemkes.go.id>