

ANALISIS DATA EKSPLORATORI DAN CLUSTERING K-MODES UNTUK PEMETAAN STATUS GIZI BALITA PADA KASUS STUNTING DI KABUPATEN ASAHAN

William Ramdhan¹, Nurwati², Elly Rahayu³

Universitas Royal, Kisaran

e-mail: ¹william.ramdhan052@gmail.com, ²nurwati763@gmail.com,

³ellyrahayu68@gmail.com

Abstract: *Stunting is a chronic nutritional problem that impacts children's physical growth and development and is a public health challenge in Indonesia. This study integrates Exploratory Data Analysis (EDA) and K-Modes Clustering to map the nutritional status of toddlers in stunting cases in Asahan Regency. EDA was used to explore data distribution, identify relationships between nutritional status indicators, and identify risk factor patterns, while K-Modes was used to group toddlers based on shared categorical characteristics. The dataset used included sociodemographic variables (gender, age category, parental education and occupation) and nutritional status indicators (weight/age, height/age, weight/height). The analysis results showed a moderate positive correlation between weight/age and height/age (0.32) and a strong negative correlation between weight/age and weight/height (-0.50), indicating a link between stunting and wasting. The application of K-Modes resulted in three main clusters: Cluster 0, dominated by female toddlers with normal nutritional status but low parental education; Cluster 1 consists of infant girls with low weight for age and short height for age, despite most parents having a high school education; Cluster 2 contains infant boys with very low weight for age and very short height for age, and relatively low maternal education. The profile of each cluster was analyzed to identify dominant characteristics relevant for intervention. This integrative approach demonstrates that the combination of EDA and K-Modes is able to provide a comprehensive picture of variations in toddler nutritional status, thus serving as a basis for planning more targeted promotive and stunting prevention strategies at the regional level.*

Keyword: *Stunting, Exploratory Data Analysis, Clustering, K-Modes*

Abstrak: Stunting merupakan masalah gizi kronis yang berdampak pada pertumbuhan fisik dan perkembangan anak, serta menjadi salah satu tantangan kesehatan masyarakat di Indonesia. Penelitian ini mengintegrasikan Analisis Data Eksploratori (EDA) dan Clustering K-Modes untuk memetakan status gizi balita pada kasus stunting di Kabupaten Asahan. EDA digunakan untuk mengeksplorasi distribusi data, mengidentifikasi hubungan antarindikator status gizi, dan menemukan pola faktor risiko, sedangkan K-Modes dimanfaatkan untuk mengelompokkan balita berdasarkan kesamaan karakteristik kategorikal. Dataset yang digunakan mencakup variabel sosiodemografis (jenis kelamin, kategori umur, pendidikan dan pekerjaan orang tua) serta indikator status gizi (BB/U, TB/U, BB/TB). Hasil analisis menunjukkan adanya korelasi positif sedang antara BB/U dan TB/U (0,32) serta korelasi negatif kuat antara BB/U dan BB/TB (-0,50), yang mengindikasikan keterkaitan antara stunting dan wasting. Penerapan K-Modes menghasilkan tiga klaster utama: Klaster 0 didominasi balita perempuan dengan status gizi normal, namun tingkat pendidikan orang tua rendah; Klaster 1 terdiri dari bayi perempuan dengan BB/U kurang dan TB/U pendek, meskipun sebagian besar orang tua berpendidikan SLTA; Klaster 2 berisi bayi laki-laki dengan BB/U sangat kurang dan TB/U sangat pendek, serta pendidikan ibu relatif rendah. Profil setiap klaster dianalisis untuk mengidentifikasi karakteristik dominan yang relevan untuk intervensi. Pendekatan integratif ini membuktikan bahwa kombinasi EDA dan K-Modes mampu memberikan gambaran

komprehensif tentang variasi status gizi balita, sehingga dapat menjadi landasan perencanaan strategi promotif dan pencegahan stunting yang lebih tepat sasaran di tingkat daerah.

Kata kunci: Stunting, Analisis Data Eksploratori, Clustering, K-Modes

PENDAHULUAN

Stunting merupakan gangguan pertumbuhan kronis yang berdampak serius terhadap perkembangan fisik dan kognitif anak, serta menjadi indikator utama kualitas kesehatan masyarakat. Dampaknya tidak hanya bersifat individual, tetapi juga memengaruhi pertumbuhan ekonomi dan pembangunan nasional secara luas (Pertwi & Hendrati, 2023)(Kresna et al., 2024). Meskipun prevalensinya menurun, angka stunting di Indonesia masih melampaui batas maksimal WHO sebesar 20%, termasuk di Kabupaten Asahan yang mencatat angka di atas rata-rata nasional pada 2022(Fauzan Adzim et al., 2023) (Handayani et al., 2024).

Perkembangan teknologi big data dan data mining menghadirkan peluang besar dalam mengidentifikasi pola dan faktor risiko stunting melalui analisis data demografi yang kompleks. Pemanfaatan pendekatan ini dapat menghasilkan pengetahuan baru (*knowledge discovery*) yang mendalam dan bermanfaat untuk menyusun strategi promotif yang lebih akurat dan berbasis data (Weston et al., 2022). Terutama dengan penggunaan model matematis dan algoritma yang tepat, analisis dapat dikembangkan untuk menghasilkan visualisasi kluster risiko dan memberikan dukungan terhadap pengambilan keputusan berbasis bukti (*evidence-based policy*)

Berdasarkan penelitian dan intervensi selama ini berfokus pada faktor internal balita seperti umur, jenis kelamin, berat dan tinggi badan(Wargiyo et al., 2025)(Reza & Muhammad Syaifur Rohman, 2024)(Pratistha & Kristianto, 2024), sehingga pendekatan promotif yang diterapkan cenderung belum menyentuh akar penyebab yang lebih kompleks.

Padahal, berbagai studi terkini menunjukkan bahwa stunting juga dipengaruhi oleh faktor eksternal (Ramdhan et al., 2025), seperti kondisi demografi keluarga (pendidikan orang tua, pekerjaan orang tua dan ekonomi keluarga). Kondisi ini menunjukkan pentingnya pemanfaatan pendekatan analitik yang mampu mengintegrasikan faktor internal dan eksternal secara komprehensif melalui pengolahan data demografi yang kompleks.

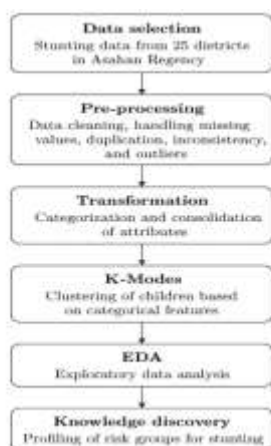
Salah satu algoritma yang relevan dalam konteks data demografi yang bersifat kategorikal adalah K-Modes, yang merupakan pengembangan dari K-Means untuk menangani atribut nominal. Dalam penelitian ini, K-Modes digunakan dalam pendekatan multi-view clustering, yang memungkinkan pengelompokan berdasarkan berbagai dimensi data seperti sosial, lingkungan, dan kesehatan(Ramdhan et al., 2021). Penggunaan teknik ini terbukti menghasilkan kluster yang lebih representatif dan berkualitas tinggi dengan akurasi di atas 98% dalam studi sebelumnya.

Penerapan model ini, diharapkan dapat memberikan kontribusi yang nyata dalam menekan angka stunting, sehingga memungkinkan pemerintah daerah dan pihak terkait untuk mengambil tindakan yang lebih efektif sesuai dengan strategi promotif yang berikan. Langkah-langkah ini menandai perkembangan terbaru dalam penggunaan teknologi dan analisis data untuk mendukung kebijakan kesehatan masyarakat yang lebih efektif dan berkelanjutan.

METODE

Kerangka kerja penelitian ini

dirancang untuk memberikan alur analisis yang sistematis dalam memetakan status gizi balita pada kasus stunting di Kabupaten Asahan. Setiap tahap dalam framework merepresentasikan langkah kunci mulai dari pengumpulan data hingga penemuan pengetahuan (*knowledge discovery*) yang dapat digunakan untuk mendukung strategi intervensi, dapat dilihat pada Gambar 1.



Gambar 1 Framework Penelitian

Penelitian ini menggunakan kerangka kerja analisis data terstruktur untuk memetakan status gizi balita pada kasus stunting di Kabupaten Asahan. Proses diawali dengan data selection, yaitu pengumpulan data primer stunting dari 25 kecamatan melalui Dinas Pengendalian Penduduk dan Keluarga Berencana (DPPKB) yang diperkaya dengan data eksternal dari Dinas Kependudukan dan Pencatatan Sipil terkait variabel sosiodemografis. Selanjutnya dilakukan pre-processing untuk membersihkan data dari nilai hilang, duplikasi, inkonsistensi, dan outlier, sehingga diperoleh dataset yang valid dan konsisten.

Tahap transformasi dilakukan untuk mengkategorikan dan mengkonsolidasikan atribut sesuai variabel operasional, sehingga format data sesuai kebutuhan analisis. Proses K-Modes clustering kemudian digunakan untuk mengelompokkan balita berdasarkan kesamaan karakteristik kategorikal, dengan jumlah klaster optimal ditentukan menggunakan metode *elbow*. Setelah itu,

dilakukan Exploratory Data Analysis (EDA) pada klaster dengan nilai K optimal untuk memahami distribusi variabel, mengidentifikasi pola, dan menganalisis keterkaitan faktor risiko.

Tahap akhir adalah knowledge discovery, yaitu penarikan pengetahuan baru dari hasil klasterisasi dan EDA untuk menghasilkan profil kelompok risiko stunting yang dapat digunakan sebagai dasar perumusan strategi promotif dan pencegahan yang lebih tepat sasaran di tingkat daerah.

HASIL DAN PEMBAHASAN

Data Selection (Pengumpulan data)

Proses *Knowledge Discovery in Databases* (KDD) diawali dari *data selection*/pengumpulan data primer yakni data stunting dari 25 Kecamatan yang ada di Kabupaten Asahan dari Dinas Pengendalian Penduduk dan Keluarga Berencana (DPPKB). Atribut data yang tersedia terdiri dari NIK Balita, Jenis Kelamin, Umur, Nama Ortu, Alamat, Tinggi Badan (TB), Berat Badan (BB), BB/U, TB/U.

Minimnya data demografi yang tersedia sehingga perlu dilakukan enrichment data yang merupakan proses memperkaya data yang sudah ada dengan data atau informasi lain (eksternal) yakni penambahan data dari Dinas Capil Kab.Asahan meliputi: Pendidikan Ayah, Pendidikan Ibu, Pekerjaan Ayah, Pekerjaan Ibu.

Pre-Processing

Tahap *preprocessing* dilakukan untuk memastikan kualitas data sebelum memasuki proses analisis. Data yang telah terhimpun umumnya masih mengandung permasalahan seperti *missing value*, data yang tidak konsisten (*inconsistent*), duplikasi data (*redundant*), maupun data pencilan (*outlier*). Pada penelitian ini, data dengan kondisi tersebut diidentifikasi melalui proses pemeriksaan awal (*data cleaning*). Data yang tidak memenuhi kriteria kualitas, seperti data dengan atribut kosong atau nilai yang tidak logis,

dihapus atau diabaikan agar tidak memengaruhi hasil analisis. Proses ini bertujuan untuk menghasilkan dataset yang bersih, konsisten, dan siap diolah pada tahap berikutnya.

Transformasi

Tahap berikutnya adalah *data transformation*, yaitu proses mengubah data ke dalam format atau struktur yang lebih sesuai dengan kebutuhan analisis. Pada penelitian ini, dataset stunting yang digunakan merupakan data kategorikal. Proses transformasi dilakukan dengan mengkonsolidasikan data serta mengelompokkannya ke dalam skala tertentu guna mengurangi jumlah variasi pada atribut tertentu. Proses dalam pegabungan data serta bentuk data dibagi dalam beberapa kelompok dilakukan pada atribut jenis kelamin, pendidikan pekerjaan. Transformasi ini dilaksanakan berdasarkan variabel operasional yang

telah ditentukan sebelumnya, sehingga data yang dihasilkan siap digunakan pada tahap analisis berikutnya.

Tabel 1 Variabel Operasional

No	Atribut	Value
1	Jenis Kelamin	Perempuan Laki-Laki
2	Kelompok Umur	Bayi: 0-11 bln Balita: 12 – 59 bln
3	Pendidikan	SD/Sederajat SLTP/Sederajat SLTA/Sederajat Diploma I-III Strata-I
4	Pekerjaan	Petani Nelayan Buruh Pegawai/Karyawan Wiraswasta

Tabel 2 Dataset Stunting Kabupaten Asahan

No	Nama Bayi	Jenis Kelamin	Kelompok Umur	Pendidikan Ayah	Pekerjaan Ayah	Pendidikan Ibu	Pekerjaan Ibu	BB/U	TB/U	BB/TB
1	Juwita Tasya	Perempuan	Bayi	Diploma IV/Strata I	SLTA/Sederajat	Karyawan	Mengurus Rumah Tangga	Sangat Kurang	Pendek	Gizi Kurang
2	Melinda Nazwa	Perempuan	Bayi	SLTA/Sederajat	SLTA/Sederajat	Wiraswasta	Mengurus Rumah Tangga	Sangat Kurang	Sangat Pendek	Gizi Buruk
3	Mutia	Perempuan	Balita	SLTA/Sederajat	SLTA/Sederajat	Wiraswasta	Mengurus Rumah Tangga	Normal	Pendek	Normal
...										
255	Haikal Hidayah	Laki-Laki	Balita	SD/Sederajat	SD/Sederajat	Petani	Mengurus Rumah Tangga	Sangat Kurang	Sangat Pendek	Gizi Kurang

Data Mining

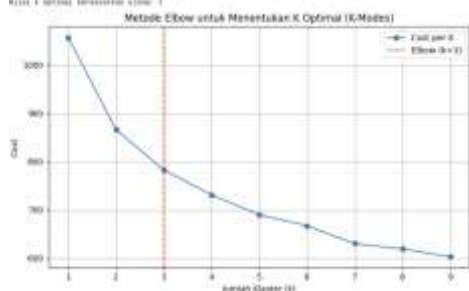
Data mining adalah suatu proses yang bertujuan menemukan pola atau pengetahuan yang tersembunyi dari sekumpulan data, kemudian mengubahnya menjadi informasi yang memiliki makna

(Apriyani et al., 2023). Setelah data melewati tahap *preprocessing* dan transformasi, tahap berikutnya adalah menemukan pola serta membentuk klaster berdasarkan karakteristik yang dimiliki data tersebut

Menentukan jumlah kluster merupakan tahap yang krusial, karena perbedaan jumlah kluster dapat menghasilkan segmentasi serta interpretasi yang bervariasi (Kustanto et al., 2024). Untuk mendapatkan kluster yang optimal dan mampu merepresentasikan struktur alami data, digunakan pendekatan *elbow method*. Metode ini mengidentifikasi jumlah kluster terbaik dengan mengamati titik tekuk (*elbow point*) pada grafik evaluasi sebagai indikator penentuan jumlah kluster (Rahmawati et al., 2024).

```
# 4. Fungsi titik siku optimal
knee = knncluster(k_range, cost, curve='cost', direction='decreasing')
optimal_k = knee.knee
print(f"Nilai k optimal berdasarkan elbow: {optimal_k}")

# 5. Visualisasi Elbow
plt.figure(figsize=(8, 5))
plt.plot(k_range, costs, marker='o', label='Cost per K')
plt.axvline(optimal_k, color='red', linestyle='--', label=f'Elbow (k={optimal_k})')
plt.title('Metode Elbow untuk Menentukan K Optimal (K-Modes)')
plt.xlabel('Jumlah Kluster (k)')
plt.ylabel('Cost')
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()
# Nilai k optimal berdasarkan elbow: 3
```



Gambar Hasil Penentuan Jumlah Kluster dengan pendekatan Elbow Method

K-Modes

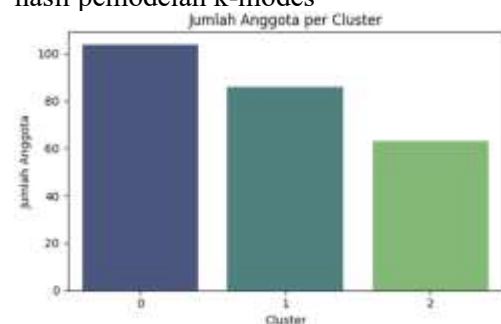
Tahap berikutnya adalah melakukan proses klusterisasi menggunakan algoritma *K-Modes* dengan jumlah $K=Uji = 3$ sesuai dengan *elbow point* optimal berdasarkan pendekatan *Elbow Method*. Algoritma *K-Modes* yang dihasilkan yang merupakan pengembangan dari algoritma *K-Means* untuk menangani data kategorikal. Algoritma *K-Means* sendiri dikenal sebagai salah satu metode *partitional clustering* berbasis pusat yang paling banyak digunakan untuk data numerik. Perluasan konsep ini pada *K-Modes* dilakukan dengan menyesuaikan perhitungan jarak, yakni mengganti metode pengukuran jarak numerik dengan metrik ketidakmiripan (*dissimilarity measure*) yang sesuai untuk data kategorikal. Selain itu, nilai rata-rata pusat

kluster pada *K-Means* digantikan oleh nilai modulus pada *K-Modes*. Proses pembaruan modulus dilakukan dengan pendekatan berbasis frekuensi, sehingga metode ini mampu menghasilkan pengelompokan yang optimal untuk variabel-variabel kategorikal (Ramdhan et al., 2021).

EDA (Exploratory Data Analysis)

Proses Exploratory Data Analysis (EDA) pada penelitian ini dilaksanakan secara terfokus pada kluster dengan nilai K optimal yang diperoleh dari hasil klusterisasi data stunting. Tahap ini memiliki peranan strategis untuk memastikan bahwa data yang digunakan benar-benar merepresentasikan kondisi empiris di lapangan serta layak dianalisis lebih lanjut. Mengingat faktor penyebab stunting bersifat heterogen, mencakup jenis kelamin, kelompok umur, Pendidikan ayah dan ibu, pekerjaan ayah dan ibu BB/U, TB/U, BB/TB. EDA membantu memetakan distribusi setiap variabel, mengidentifikasi potensi outlier, serta menelaah pola kecenderungan umum yang muncul di dalam kluster tersebut (Ramdhan et al., 2025).

Analisis eksplorasi data dimulai dengan mengidentifikasi distribusi responden pada masing-masing kluster hasil pemodelan *k-modes*

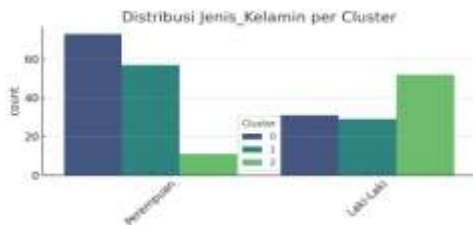


Gambar 3 Distribusi Jumlah Anggota pada Cluster

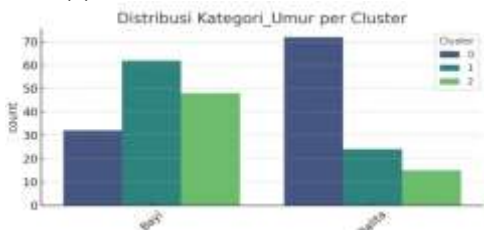
Analisis awal terhadap data hasil pengelompokan menggunakan metode *k-modes* menunjukkan bahwa jumlah anggota dalam setiap kluster tidak terdistribusi secara merata. Cluster 0 memiliki jumlah anggota terbesar (104 individu), diikuti Cluster 1 (86 individu), dan Cluster 2 (63 individu). Perbedaan ini

mengindikasikan adanya variasi signifikan pada karakteristik yang membentuk setiap klaster, di mana Cluster 0 merepresentasikan proporsi populasi terbesar, sedangkan Cluster 2 mewakili kelompok dengan karakteristik yang relatif jarang ditemukan.

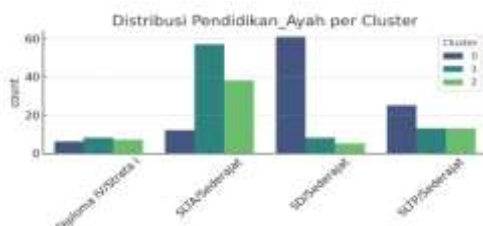
Distribusi variabel kategorikal per klaster memperlihatkan adanya pola khas pada masing-masing kelompok. Berdasarkan hasil perhitungan modus, setiap klaster memiliki kombinasi atribut dominan yang berbeda, dengan persentase kemunculan modus bervariasi antara >50% hingga >80%. Hal ini menegaskan bahwa proses klasterisasi berhasil memisahkan individu ke dalam kelompok yang homogen secara internal namun heterogen antar kelompok.



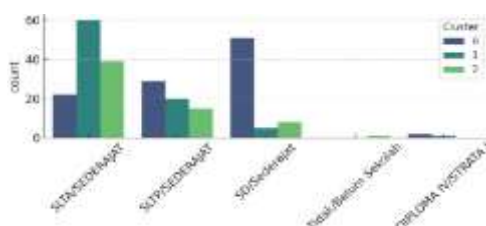
(a) Distribusi Jenis Kelamin



(b) Distribusi Kelompok Umur



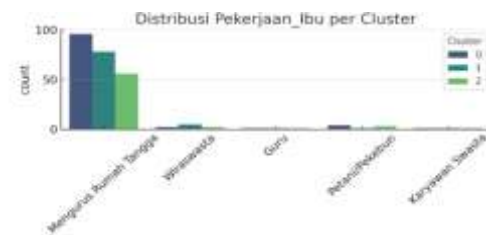
(c) Distribusi Pendidikan Ayah



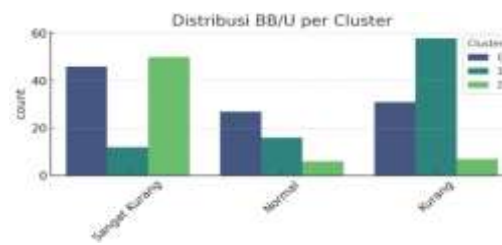
(d) Distribusi Pendidikan Ibu



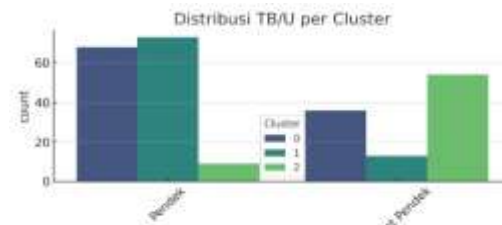
(e) Distribusi Pekerjaan Ayah



(f) Distribusi Pekerjaan Ibu



(g) Distribusi BB/U



(h) Distribusi TB/U



(h) Distribusi BB/TB

Gambar 4 Distribusi Variabel Kategorikal Pada Cluster

Berdasarkan hasil analisis distribusi variabel pada masing-masing klaster, terlihat perbedaan profil yang cukup mencolok. Klaster 0 didominasi oleh anak perempuan (70,2%), mayoritas berada

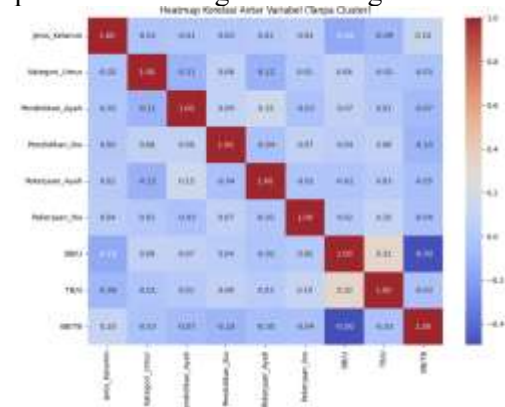
pada kategori balita (69,2%), dengan tingkat pendidikan orang tua tergolong rendah (ayah 58,7% berpendidikan SD/ sederajat; ibu 49,0% SD/ sederajat). Pekerjaan ayah sebagian besar adalah wiraswasta (44,2%), sementara ibu cenderung tidak bekerja secara formal (92,3%). Status gizi menunjukkan proporsi tinggi BB/U sangat kurang (44,2%) dan TB/U pendek (65,4%), namun BB/TB normal pada 56,7% anak. Pola ini mengindikasikan adanya masalah gizi kronis tanpa penurunan signifikan pada status gizi akut.

Klaster 1 mayoritas terdiri dari anak perempuan (66,3%), dengan proporsi tertinggi pada kategori bayi (72,1%). Pendidikan orang tua relatif lebih baik dibanding Klaster 0, dengan mayoritas berpendidikan SLTA/ sederajat (ayah 66,3%, ibu 69,8%). Pekerjaan ayah umumnya wiraswasta (58,1%), sedangkan sebagian besar ibu berperan sebagai ibu rumah tangga (90,7%). Status gizi didominasi oleh BB/U kurang (67,4%) dan TB/U pendek (84,9%), namun BB/TB normal pada 69,8% anak. Temuan ini menandakan adanya masalah gizi kronis yang merata pada kelompok bayi, sementara kondisi gizi akut relatif terkendali.

Klaster 2 didominasi oleh anak laki-laki (82,5%), mayoritas berusia bayi (76,2%), dengan tingkat pendidikan orang tua sebanding dengan Klaster 1 (ayah 60,3%, ibu 61,9% SLTA/ sederajat). Pekerjaan ayah sebagian besar wiraswasta (42,9%), dan mayoritas ibu adalah ibu rumah tangga (88,9%). Status gizi pada klaster ini tergolong paling memprihatinkan, dengan BB/U sangat kurang (79,4%), TB/U sangat pendek (85,7%), dan BB/TB normal hanya 47,6%. Hal ini menunjukkan adanya kombinasi kekurangan gizi kronis dan akut yang lebih parah dibandingkan dua klaster lainnya.

Untuk memperoleh pemahaman yang lebih komprehensif mengenai keterkaitan antara variabel sosiodemografis dan indikator status gizi balita, dilakukan analisis korelasi yang divisualisasikan dalam bentuk heatmap,

dapat dilihat pada Gambar 4. Analisis ini bertujuan untuk mengidentifikasi arah dan kekuatan hubungan antarvariabel, sehingga dapat memberikan gambaran awal mengenai faktor-faktor yang saling berhubungan dan berpotensi memengaruhi kejadian stunting. Pendekatan ini sekaligus menjadi dasar untuk menentukan variabel prioritas yang relevan dalam perumusan strategi intervensi gizi.



Gambar 5 Heatmap Korelasi Antar Variabel

Hasil analisis korelasi menunjukkan adanya hubungan yang konsisten antara indikator status gizi anak. Berat badan menurut umur (BB/U) dan tinggi badan menurut umur (TB/U) memiliki korelasi positif sedang sebesar 0,32, yang menunjukkan bahwa anak dengan berat badan sesuai usianya cenderung memiliki tinggi badan yang proporsional terhadap usia yang sama. Sebaliknya, BB/U memiliki korelasi negatif kuat dengan berat badan menurut tinggi badan

(BB/TB) sebesar -0,50, yang mengindikasikan bahwa anak dengan berat badan rendah untuk usianya sering kali memiliki rasio berat terhadap tinggi yang berada di bawah standar optimal. Temuan ini sejalan dengan konsep gizi yang menyatakan bahwa kekurangan gizi kronis (stunting) kerap disertai oleh kekurangan gizi akut (wasting).

Pengaruh variabel biologis relatif kecil. Jenis kelamin memiliki korelasi negatif lemah dengan BB/U (-0,16), dan kategori umur memiliki korelasi yang sangat rendah terhadap TB/U (-0,01) maupun BB/U (0,08). Hal ini mengimplikasikan bahwa faktor biologis

tersebut memiliki peran yang lebih terbatas dibandingkan determinan lingkungan, asupan pangan, maupun praktik pengasuhan. Studi sebelumnya juga menempatkan pola konsumsi dan kualitas perawatan sebagai faktor dominan dalam menentukan status gizi.

Faktor sosiodemografis menunjukkan pola hubungan yang menarik. Pendidikan ayah berkorelasi positif kecil dengan pekerjaan ayah (0,15) dan dengan pendidikan ibu (0,09), mengindikasikan adanya kecenderungan pasangan memiliki tingkat pendidikan yang relatif sepadan serta hubungan antara pendidikan formal dengan peluang memperoleh pekerjaan yang lebih baik. Namun, pekerjaan ayah dan pekerjaan ibu

menunjukkan korelasi yang sangat rendah terhadap indikator gizi, mengisyaratkan bahwa jenis pekerjaan semata bukan determinan utama status gizi anak. Pendidikan ibu, meskipun memiliki korelasi lemah dengan variabel gizi, berpotensi memainkan peran strategis dalam perbaikan status gizi anak melalui peningkatan literasi kesehatan, pemahaman kebutuhan gizi, dan penerapan pola makan yang tepat.

Selanjutnya, hasil k-modes clustering mengelompokkan data menjadi tiga cluster utama yang menunjukkan perbedaan signifikan pada distribusi karakteristik sosiodemografis dan status gizi anak dapat dilihat pada Tabel 3.

Tabel 3 Ringkasan Karakteristik Dominan dan Starategi Promotif Stunting

Variabel	Cluster 0	Cluster 1	Cluster 2
Jenis Kelamin	Perempuan (70.2 %)	Perempuan (66.3%)	Perempuan (82.5 %)
Kategori Umur	Balita (69.2 %)	Bayi (72.1 %)	Bayi (76.2 %)
Pendidikan Ayah	SD/Sederajat (58.7 %)	SLTA/Sederajat (66.3 %)	SLTA/Sederajat (60.3 %)
Pendidikan Ibu	SD/Sederajat (49.0 %)	SLTA/Sederajat (69.8 %)	SLTA/Sederajat (61.9 %)
Pekerjaan Ayah	Wiraswasta (44.2 %)	Wiraswasta (58.1 %)	Wiraswasta (42.9 %)
Pekerjaan Ibu	Mengurus RT (92.3 %)	Mengurus RT (90.7 %)	Mengurus RT (88.9 %)
BB/U	Sangat Kurang (44.2 %)	Kurang (67.4 %)	Sangat Kurang (79.4 %)
TB/U	Pendek (65.4 %)	Pendek (84.9 %)	Pendek (85.7 %)
BB/TB	Normal (56.7 %)	Normal (69.4 %)	Normal (47.6 %)
Interpretasi	Berat badan dan tinggi badan cenderung selaras; risiko stunting relatif rendah, namun keterbatasan pendidikan orang tua berpotensi menurunkan kualitas pengasuhan	Berat badan rendah sering disertai rasio BB/TB kurang baik; pengetahuan gizi mungkin ada tetapi penerapan praktik pengasuhan belum optimal	Kasus gizi buruk kombinasi stunting dan wasting tertinggi; pendidikan ibu rendah berpotensi memperburuk ketidakseimbangan asupan dan pengasuhan

Secara keseluruhan, kombinasi hasil korelasi dan analisis cluster ini menegaskan bahwa intervensi perbaikan gizi anak memerlukan pendekatan multidimensional. Selain perbaikan faktor

ekonomi, strategi harus mencakup peningkatan literasi gizi keluarga, promosi pola konsumsi sehat, penguatan peran pendidikan ibu, dan dukungan kebijakan yang memungkinkan pemantauan

pertumbuhan anak secara rutin. Pendekatan ini akan lebih efektif jika disesuaikan dengan karakteristik spesifik setiap kelompok yang telah teridentifikasi.

SIMPULAN

Integrasi Exploratory Data Analysis (EDA) dan K-Modes Clustering mampu memberikan pemetaan komprehensif terhadap variasi status gizi balita pada kasus stunting di Kabupaten Asahan. Pendekatan ini berhasil mengidentifikasi tiga klaster utama dengan karakteristik sosiodemografis dan status gizi yang berbeda, sehingga memudahkan penentuan kelompok prioritas intervensi. Analisis korelasi mengungkap adanya hubungan yang konsisten antara indikator gizi kronis (stunting) dan gizi akut (wasting), serta menegaskan bahwa faktor sosiodemografis, khususnya pendidikan ibu, berpotensi menjadi determinan penting dalam perbaikan status gizi anak. Temuan ini memperkaya pemahaman ilmiah mengenai keterkaitan antara faktor internal dan eksternal dalam kejadian stunting, serta menawarkan kerangka analisis berbasis data yang dapat digunakan untuk mendukung perumusan strategi promotif dan pencegahan stunting yang lebih tepat sasaran di tingkat daerah.

UCAPAN TERIMA KASIH

Penulis menyampaikan apresiasi yang setinggi-tingginya kepada Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia (Kemendikbudristek) atas dukungan pendanaan melalui skema Penelitian Dasar Pemula yang disetujui dalam Program Penelitian Nasional Tahun 2025. Ucapan terima kasih yang mendalam juga ditujukan kepada Pimpinan Universitas Royal dan Ketua Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) Universitas Royal atas arahan, dukungan, dan fasilitasi yang diberikan selama proses penelitian. Penulis juga menghaturkan terima kasih kepada Dinas Pengendalian

Penduduk dan Keluarga Berencana (DPPKB) Kabupaten Asahan dan Dinas Kependudukan dan Pencatatan Sipil (Disdukcapil) Kabupaten Asahan atas kerja sama yang baik dalam penyediaan data dan informasi lapangan, yang berkontribusi signifikan terhadap kelancaran dan keberhasilan penelitian ini.

DAFTAR PUSTAKA

- Apriyani, P., Dikananda, A. R., & Ali, I. (2023). Penerapan Algoritma K-Means dalam Klasterisasi Kasus Stunting Balita Desa Tegalwangi. *Hello World Jurnal Ilmu Komputer*, 2(1), 20–33. <https://doi.org/10.56211/helloworld.v2i1.230>
- Fauzan Adzim, Budianita, E., Nazir, A., & Syafria, F. (2023). Klasifikasi Status Stunting Balita Menggunakan Metode C4.5 Berbasis Web. *ZONAsi: Jurnal Sistem Informasi*, 5(3), 215–225. <https://doi.org/10.31849/zn.v5i3.15828>
- Handayani, M., Fitri, M., Sibuea, L., Tinggi, S., Informatika, M., & Royal, D. K. (2024). Optimasi Model Clustering Dalam Pemetaan Stunting Di Kabupaten Asahan. *Journal of Science and Social Research*, 4307(4), 2192–2197. <http://jurnal.goretanpena.com/index.php/JSSR>
- Kresna, I., Pratama, P., Triastuti, N., & Sudarmaji. (2024). Faktor Kejadian Stunting Pada Balita. *CoMPHI Journal: Community Medicine and Public Health of Indonesia Journal*, 4(2). <https://doi.org/10.37148/comphijournal.v4i2.168>
- Kustanto, Y., Arumi, R., Sasongko, D., Ully Artha, E., & Prabowo, N. A. (2024). Implementasi K-Modes Clustering Untuk Pengelompokan Data Bermain Game Pada Mahasiswa Ditinjau Dari Durasi Belajarnya. *Media Online*, 4(5), 2495–2505.

- <https://doi.org/10.30865/klik.v4i5.1619>
- Pertiwi, A. N. A. M., & Hendrati, L. Y. (2023). Literature Review: Analisis Penyebab Kejadian Stunting Pada Balita Di Provinsi Jawa Timur. *Amerta Nutrition*, 7(2SP), 320–327. <https://doi.org/10.20473/amnt.v7i2sp.2023.320-327>
- Pratistha, R. N., & Kristianto, B. (2024). Implementasi Algoritma K-Means dalam Klasterisasi Kasus Stunting pada Balita di Desa Randudongkal. *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, 5(2), 1193–1205. <https://doi.org/10.35870/jimik.v5i2.6>
- Rahmawati, T., Wilandari, Y., & Kartikasari, P. (2024). Analisis Perbandingan Silhouette Coefficient Dan Metode Elbow Pada Pengelompokan Provinsi Di Indonesia Berdasarkan Indikator Ipm Dengan K-Medoids. *Jurnal Gaussian*, 13(1), 13–24. <https://doi.org/10.14710/j.gauss.13.1.13-24>
- Ramdhan, W., Sitompul, O. S., Nababan, E. B., & Nasution, S. (2025). a Multi-View Clustering Approach and Exploratory Data Analysis for Dominant Demographic Knowledge Discovering on the Cause Behind the Pandemic. *ICIC Express Letters*, 19(2), 175–184. <https://doi.org/10.24507/icicel.19.02.175>
- Ramdhan, W., Sitompul, O. S., Nababan, E. B., & Sawaluddin. (2021). A Framework for Dominant Factors Revelation of the Outbreak's Cause. *2021 International Conference on Data Science, Artificial Intelligence, and Business Analytics, DATABIA 2021 - Proceedings*, 52–57. <https://doi.org/10.1109/DATABIA53375.2021.9649732>
- Reza, A. A. R., & Muhammad Syaifur Rohman. (2024). Prediction Stunting Analysis Using Random Forest Algorithm and Random Search Optimization. *Journal of Informatics and Telecommunication Engineering*, 7(2), 534–544. <https://doi.org/10.31289/jite.v7i2.10628>
- Wargiyo, S. Y., Susliansyah, & Sumarno, H. (2025). *Rancangan Aplikasi Algoritma C4 . 5 pada Stunting Balita Menggunakan Bahasa Phyton*. 9(1), 254–265.
- Weston, J., Bickert, B., Stasiuk, C., Alzhouri, F., & Ebrahim, D. (2022). Dynamic Analysis of Demographic Sentiment. *2022 IEEE International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2022*. <https://doi.org/10.1109/IEMTRONIC555184.2022.9795706>