Jatekom (Jurnal Aplikasi Teknologi dan Komputasi)
Vol. 1, No. 2, Juni 2025, pp. 94~108

Mengungkap Heterogenitas Stunting pada Anak: Pendekatan Machine Learning untuk Intervensi yang Tepat Sasaran di Sambas, Indonesia

Heru Pramono Hadi[1], Fadlil Chandra Pratama[2], Farrikh Al Zami[3], Yupie Kusumawati[4], Suharnawi[5], Ayu Ashari[6], Henry Bastian[7], Alfiena Nisa Belladiena[8]
[1,2,3,4,5]Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro
[6,8]Program Studi Rekam Medis dan Informasi Kesehatan, Fakultas Kesehatan, Universitas Dian Nuswantoro
[7]Program Studi Desain Komunikasi Visual, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro
Jl. Imam Bonjol No. 207, Pendrikan Kidul, Kecamatan Semarang Tengah, Kota Semarang, 50131

Artikel Info

Kata kunci:

Stunting
k-means clustering
t-sne
reduksi dimensi
informasi kesehatan publik

ABSTRAK

Stunting di Kabupaten Sambas, Indonesia, merupakan tantangan kesehatan masyarakat yang kritis dengan implikasi sosio-ekonomi yang signifikan. Penelitian ini bertujuan untuk mengidentifikasi profil kasus stunting secara mendalam melalui pendekatan *unsupervised machine learning* guna membongkar heterogenitas pada faktor-faktor yang memengaruhinya. Analisis ini menggunakan data kesehatan *cross-sectional* dari 599 anak (usia 0-11 bulan) di Kabupaten Sambas, yang mencakup data demografis, sosio-ekonomi, dan akses kesehatan dengan dimensionalitas tinggi. Kami menerapkan metodologi yang sistematis, dimulai dengan pra-pemrosesan data ekstensif, reduksi dimensi menggunakan *t-distributed Stochastic Neighbor Embedding* (t-SNE), dan dilanjutkan dengan segmentasi populasi menggunakan *clustering* K-Means. Jumlah klaster optimal (k) ditentukan menggunakan metode Elbow. Analisis berhasil mengidentifikasi empat klaster (k=4) yang berbeda secara signifikan. Klaster dengan prevalensi stunting terendah secara konsisten menunjukkan tingkat pendidikan orang tua yang lebih tinggi dan akses terbaik terhadap informasi kesehatan yang disediakan oleh bidan. Sebaliknya, klaster dengan prevalensi stunting tertinggi berkorelasi dengan pendapatan keluarga terendah dan tingkat pendidikan orang tua yang rendah. Temuan kunci yang paling menonjol adalah adanya klaster berpendapatan tertinggi namun tetap memiliki tingkat stunting yang tinggi; analisis lebih lanjut menunjukkan klaster ini memiliki akses terburuk terhadap informasi kesehatan esensial. Penelitian ini menyimpulkan bahwa tingkat pendidikan orang tua dan akses terhadap informasi kesehatan yang dimediasi oleh bidan merupakan prediktor stunting yang lebih kuat dibandingkan status ekonomi semata. Temuan ini membantah asumsi bahwa intervensi ekonomi saja cukup. Hasil penelitian menyediakan basis bukti yang kuat untuk perancangan kebijakan kesehatan publik yang lebih tertarget dan tidak monolitik, dengan memprioritaskan program edukasi dan penguatan peran tenaga kesehatan di lapangan.

Penulis Korespondensi :

Heru Pramono Hadi
Program Studi Sistem Informasi
Fakultas Ilmu Komputer
Universitas Dian Nuswantoro
heru.pramono.hadi@dsn.dinus.ac.id

1. PENDAHULUAN

1.1 Latar Belakang

Stunting merupakan salah satu permasalahan kesehatan masyarakat yang paling serius. Kondisi ini ditandai dengan pertumbuhan dan perkembangan anak yang terhambat, yang umumnya disebabkan oleh

malnutrisi kronis. Stunting diukur berdasarkan indikator tinggi badan menurut umur (height-for-age), dengan nilai yang berada lebih dari dua standar deviasi di bawah standar pertumbuhan yang ditetapkan oleh World Health Organization (WHO). Kondisi ini tidak hanya berdampak pada kesehatan, tetapi juga memengaruhi pembangunan sumber daya manusia serta pertumbuhan ekonomi. Selain itu, stunting dapat menyebabkan gangguan kognitif dan fisik jangka panjang yang berdampak pada pencapaian pendidikan dan produktivitas di masa depan [1], [2].

Dalam bidang ilmu komputer, stunting dapat ditanggulangi melalui analisis data dan teknik pembelajaran mesin (machine learning) untuk mengidentifikasi populasi yang terdampak dan mengoptimalkan alokasi sumber daya untuk kebutuhan gizi. Sebagai contoh, analitik prediktif dapat membantu memahami faktor sosial ekonomi yang berperan dalam stunting, sehingga memungkinkan pelaksanaan program yang terarah dan berbasis teknologi untuk meningkatkan hasil kesehatan [3], [4]. Selain itu, integrasi data kesehatan dengan alat komputasi dapat meningkatkan pemantauan dan evaluasi program nutrisi, yang pada akhirnya berkontribusi terhadap penurunan angka stunting [5].

Stunting tetap menjadi isu kesehatan masyarakat yang krusial di Indonesia, dengan disparitas yang signifikan di berbagai kelompok demografi. Studi terkini menunjukkan bahwa sekitar 37% anak-anak di bawah usia lima tahun mengalami stunting, menempatkan Indonesia pada peringkat keempat tertinggi di dunia dalam hal prevalensi stunting. Sebanyak 37% anak Indonesia berusia 0–59 bulan mengalami stunting [6]. Angka stunting lebih tinggi di wilayah pedesaan dibandingkan dengan wilayah perkotaan, mencerminkan kesenjangan dalam akses terhadap layanan gizi dan kesehatan [7], [8]. Anak-anak dari keluarga miskin memiliki risiko lebih tinggi mengalami stunting, di mana tingkat pendidikan ibu dan status gizi ibu menjadi prediktor signifikan [8], [9]. Faktor lain seperti berat badan lahir rendah, kebiasaan makan, dan akses terhadap sanitasi juga berperan penting dalam menentukan risiko stunting [9]. Anak yang mengalami stunting akan menghadapi penurunan kemampuan kognitif dan pencapaian pendidikan, di mana penurunan satu standar deviasi pada tinggi badan menurut umur berkorelasi dengan 0,6 tahun lebih sedikit waktu bersekolah [10].

Meskipun data statistik menunjukkan permasalahan yang mendesak, beberapa pihak berpendapat bahwa peningkatan akses terhadap layanan kesehatan dan gizi dapat menurunkan angka stunting. Namun, kesenjangan sosial ekonomi yang masih ada menunjukkan bahwa intervensi yang komprehensif dan terarah sangat diperlukan untuk mencapai kemajuan yang berkelanjutan dalam upaya penanggulangan stunting di Indonesia [6], [7].

Dengan demikian, keterkaitan antara stunting dan ilmu komputer menawarkan peluang penting untuk menghasilkan solusi inovatif terhadap tantangan global yang mendesak [11]. Masalah ini sangat relevan di negara berkembang, di mana populasi berpendapatan rendah dan menengah menghadapi tantangan terkait ketahanan pangan dan keterbatasan akses layanan kesehatan [12], [13], [14], [15].

Clustering merupakan teknik pembelajaran mesin yang mengelompokkan data tanpa label ke dalam klaster yang bermakna berdasarkan pola kesamaan. Sebagai metode pembelajaran tak terawasi (unsupervised learning), teknik ini efektif dalam mengidentifikasi hubungan antar data tanpa label yang telah ditentukan sebelumnya, sehingga sangat berguna untuk pengenalan pola [16]. Clustering menyederhanakan kumpulan data kompleks menjadi segmen yang dapat dianalisis, yang sering kali tidak memungkinkan dilakukan melalui metode tradisional. Beberapa pendekatan umum meliputi K-means dan hierarchical clustering, masing-masing memiliki keunggulan dalam analisis jenis data tertentu [16], [17], [18].

Aplikasi clustering dalam bidang kesehatan menawarkan peluang besar dalam analisis kesehatan masyarakat, pengembangan kebijakan, dan manajemen layanan publik [19]. Penelitian menunjukkan bahwa teknik ini efektif dalam mengelompokkan data kesehatan untuk mengungkap pola bermakna, yang dapat digunakan untuk merancang intervensi terarah dan alokasi sumber daya secara optimal. Sebagai contoh, clustering dapat menghasilkan visualisasi data kesehatan yang informatif bagi pemangku kepentingan dan mengidentifikasi determinan disparitas kesehatan antar komunitas [20], [21]. Meskipun Kode Etik Kedokteran Indonesia membatasi pengumpulan data tertentu [22], teknik clustering tetap bernilai dalam menganalisis indikator kesehatan fisik dan faktor sosial ekonomi yang memengaruhi akses terhadap layanan kesehatan. Pendekatan ini membantu para pemangku kepentingan dalam merancang intervensi kesehatan yang lebih adil di berbagai segmen populasi [23].

t-distributed Stochastic Neighbor Embedding (t-SNE) merupakan teknik reduksi dimensi yang mengubah data kompleks berdimensi tinggi menjadi bentuk yang lebih sederhana dan dapat diinterpretasikan, sambil tetap mempertahankan hubungan antar data [24]. Dalam analisis stunting, t-SNE efektif digunakan untuk memvisualisasikan data kesehatan multidimensional dalam rangka mengidentifikasi pola dan faktor risiko. Metode ini terbukti sangat bermanfaat dalam mengelompokkan populasi berisiko tinggi serta mengevaluasi hubungan antara intervensi nutrisi dan hasil pertumbuhan anak [25], [26]. Walaupun t-SNE unggul dalam visualisasi data dan pengenalan pola, metode ini memberikan hasil terbaik jika dikombinasikan

dengan teknik analisis lain untuk memvalidasi temuan dan memastikan kesimpulan yang kuat dalam penelitian stunting [27], [28].

Proses analisis ini didasarkan pada penelitian sebelumnya yang dilakukan di Kabupaten Sambas. Kabupaten Sambas memiliki luas wilayah 6.395,70 km² atau 4,36% dari total luas wilayah Provinsi Kalimantan Barat; kabupaten ini tergolong sebagai daerah pedesaan dan berbatasan langsung dengan Malaysia. Partisipan penelitian berasal dari 17 dari 19 kecamatan yang tersebar di 30 desa. Data dalam studi ini berasal dari dua penelitian potong lintang yang dilakukan pada tahun 2016 dan 2017. Manajemen data pada penelitian tahun 2016 dilaksanakan pada April–Agustus 2016 melalui beberapa tahap berurutan: (1) tahap persiapan (11–24 April), (2) pengumpulan data (25 April–11 Mei), (3) entri data, pembersihan data, dan analisis (Juni 2016), serta (4) penulisan laporan hingga Agustus 2016. Penelitian lanjutan dilakukan pada April–Juni 2017 dengan tahapan: (1) persiapan (3–14 April), (2) pengumpulan data (17 Mei–2 Juni 2017), (3) entri dan pembersihan data serta analisis (Mei 2017), dan (4) penulisan laporan pada akhir Mei hingga 8 Juni 2017 [29].

Beberapa studi lain mengenai stunting pada anak juga telah dilakukan menggunakan teknik K-means clustering, yang mengungkap berbagai wawasan mengenai prevalensi dan faktor-faktor terkait. Mokalla dan Mendu menggunakan K-means untuk mengidentifikasi pola stunting yang khas dalam populasi tertentu, menyoroti faktor sosial ekonomi dan gizi yang memengaruhi hasil pertumbuhan. Selain itu, sebuah studi lain menekankan efektivitas K-means dalam mengidentifikasi populasi berisiko tinggi, serta pentingnya program nutrisi yang disesuaikan dengan kebutuhan spesifik. Studi-studi ini secara kolektif menegaskan kegunaan K-means clustering dalam memahami dan menangani stunting, meskipun variasi metodologi dan fokus dapat membatasi perbandingan langsung [6]. Ahmed et al. menerapkan K-means clustering untuk mengategorikan anak-anak berdasarkan status gizi, dan menunjukkan korelasi signifikan antara stunting dan keragaman konsumsi makanan [30]. Takele et al. memfokuskan penelitiannya pada variasi regional dalam angka stunting dengan menggunakan K-means untuk mengelompokkan data dan mengidentifikasi wilayah berisiko tinggi sebagai target intervensi [15]. Namirembe et al. mengeksplorasi pengaruh pendidikan ibu terhadap stunting anak, dengan mengelompokkan data menggunakan K-means, yang mengungkap wawasan penting mengenai hubungan antara pendidikan dan hasil gizi [31]. Studi-studi tersebut secara keseluruhan menetapkan bahwa K-means clustering merupakan alat yang berharga untuk memahami pola stunting dan merancang intervensi yang tepat sasaran, sembari menekankan pentingnya pendekatan analitis yang komprehensif.

1.2 Studi Pustaka

1. Penelitian yang dilakukan oleh Imran Hossain Sumon, Shamima Akter, Md. Sazzad Hossan Sujon, Muhammad Khairul Alam, Sabina Yasmin, Sabina Yeasmin, M. Kabir, dan Md. Moyazzem Hossain pada tahun 2024 (tercantum dalam Tabel 1: Penelitian Terkait) bertujuan untuk mengidentifikasi faktor-faktor yang berpengaruh terhadap kejadian stunting pada anak di bawah usia lima tahun di Kamboja. Dataset sekunder yang digunakan terdiri atas 3.268 anak balita, yang diperoleh dari survei Cambodian Demographic and Health Survey (CDHS) 2021/2022. Metode *Agglomerative Clustering* dan *K-Means* digunakan untuk menentukan pengaruh faktor demografis, sosial ekonomi, dan faktor lainnya terhadap kejadian stunting. Hasil penelitian menunjukkan bahwa sekitar 21% anak di bawah usia lima tahun mengalami stunting, dengan prevalensi yang lebih tinggi di daerah pedesaan dibandingkan dengan daerah perkotaan. Selain itu, prevalensi stunting lebih rendah pada keluarga dengan orang tua yang memiliki tingkat pendidikan tinggi [33].
2. Penelitian oleh Khadeeja Munawar, Firdaus Mukhtar, Mollika Roy, Nida Majeed, dan Muhammad Yazid Jalaludin pada tahun 2024 (tercantum dalam Tabel 1: Penelitian Terkait) bertujuan untuk mengevaluasi sejumlah studi yang berfokus pada pola pengasuhan dan praktik pemberian makan, perilaku makan anak, serta keterlambatan pertumbuhan di negara-negara Asia. Penelitian ini menggunakan metode *Agglomerative Clustering* dan *K-Means* untuk mengelompokkan objek berdasarkan karakteristiknya selama periode Januari 2015 hingga Mei 2021. Hasil penelitian menunjukkan bahwa anak-anak yang mengalami stunting memiliki tingkat kecemasan dan depresi yang lebih tinggi, serta tingkat kepercayaan diri yang lebih rendah dibandingkan dengan anak-anak yang tidak mengalami stunting [35].
3. Penelitian oleh Tadesse Tarik Tamir, Soliyana Adisu Gezhegn, D. Dagnew, Abebe Tilahun Mekonenne, Genetu Tadese Aweke, dan Ayenew Molla Lakew pada tahun 2024 (tercantum dalam Tabel 1: Penelitian Terkait) bertujuan untuk menilai prevalensi stunting anak dan determinannya di negara-negara Afrika berpendapatan rendah dan menengah bawah. Penelitian ini menggunakan metode *Agglomerative Clustering* dan *K-Means* terhadap data survei demografi dan kesehatan yang terstandarisasi dari negara-

negara Afrika selama periode 2010 hingga 2022, dengan jumlah sampel sebesar 204.214 anak balita berbobot (weighted). Hasil penelitian menunjukkan bahwa 31,28% anak balita di negara-negara tersebut mengalami stunting, dengan interval kepercayaan 95% berada pada kisaran 31,08% hingga 31,48% [36].

Tabel 1. Penelitian Terkait

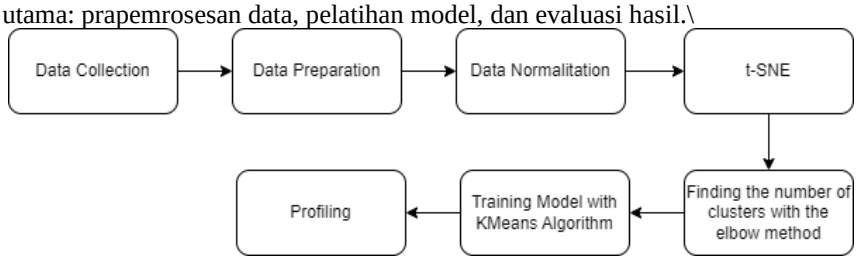
No	Nama Peneliti dan Tahun	Permasalahan	Metode	Hasil
1	Imran Hossain Sumon, dkk (2024)	Mengidentifikasi faktor-faktor yang berpengaruh terhadap stunting pada anak balita di Kamboja	*Agglomerative Clustering* dan *K-Means*	Sekitar 21% anak balita mengalami stunting, prevalensi lebih tinggi di pedesaan, dan lebih rendah pada keluarga dengan pendidikan tinggi.
2	Khadeeja Munawar, dkk (2024)	Evaluasi studi terkait pengasuhan, perilaku makan anak, dan pertumbuhan di negara Asia	*Agglomerative Clustering* dan *K-Means*	Anak stunting memiliki tingkat kecemasan dan depresi lebih tinggi serta kepercayaan diri lebih rendah dibandingkan anak non-stunting.
3	Tadesse Tarik Tamir, dkk (2024)	Menilai prevalensi stunting dan determinannya di negara Afrika berpendapatan rendah dan menengah bawah	*Agglomerative Clustering* dan *K-Means*	Sebesar 31,28% anak balita mengalami stunting dengan interval kepercayaan 95% antara 31,08%–31,48%.

Berdasarkan hasil penelitian yang dirujuk dalam jurnal pada Tabel 1 di atas, maka penerapan metode clustering dengan algoritma *K-Means* dinilai sesuai dan relevan untuk digunakan dalam menyelesaikan permasalahan stunting yang terjadi di Kabupaten Sambas, Provinsi Kalimantan Barat. Hal ini mempertimbangkan kemampuan *K-Means* dalam mengelompokkan data populasi yang kompleks dan beragam, sehingga dapat membantu dalam identifikasi kelompok berisiko tinggi dan penentuan intervensi gizi yang tepat sasaran.

2. METODE

2.1 Metode Penelitian

Metodologi dalam penelitian ini berfokus pada analisis clustering, sebuah pendekatan yang andal untuk mengelompokkan data tanpa label berdasarkan kesamaan intrinsik. Bagian ini menjelaskan proses sistematis yang dilakukan, meliputi pengumpulan data, persiapan dan prapemrosesan data, serta implementasi algoritma clustering, sebagaimana ditunjukkan pada Gambar 1. Kerangka kerja analisis ini terdiri dari tiga tahapan utama: prapemrosesan data, pelatihan model, dan evaluasi hasil.\



Gambar 1. Tahapan utama dalam proses clustering: prapemrosesan, pelatihan model, dan hasil akhir

A. Pengumpulan Data

Langkah awal yang dilakukan sebelum melakukan proses pembelajaran mesin (machine learning) adalah pengumpulan data. Dataset yang digunakan dalam penelitian ini berasal dari studi prospektif, berulang, dan potong lintang yang bertujuan untuk menyelidiki faktor-faktor yang dianggap berpengaruh terhadap kejadian stunting. Penelitian ini dilakukan di Kabupaten Sambas terhadap bayi berusia 0–11 bulan dengan jumlah sampel sebanyak 559 anak. Data yang digunakan mencakup survei ibu hamil tahun 2016 dan survei ibu dan anak tahun 2017, yang dikumpulkan untuk mengukur prevalensi stunting pada usia 0–11 bulan. Sebanyak 20 prediktor potensial stunting digunakan, dikategorikan ke dalam karakteristik rumah tangga, karakteristik ibu, layanan antenatal, dan karakteristik anak. Analisis regresi logistik digunakan untuk menilai hubungan yang disesuaikan antara stunting dan faktor-faktor tersebut.

B. Persiapan Data

Tahap ini melibatkan pembersihan dan penataan sistematis terhadap dataset awal yang terdiri dari 559 baris dan 660 kolom. Data mencakup:

- 168 variabel numerik
- 471 variabel kategorik
- 13 variabel dengan format tanggal

Delapan kolom dihapus karena memiliki nilai unik yang berlebihan sehingga berpotensi mengganggu analisis. Tahap prapemrosesan ini sangat penting untuk menjamin kualitas dan konsistensi data sebelum pemodelan.

C. Penanganan Nilai Hilang

Nilai yang hilang ditangani dengan dua pendekatan sistematis:

1. Nilai yang hilang karena tidak memenuhi kriteria tertentu diklasifikasikan sebagai ketiadaan yang bermakna, bukan data hilang.
2. Data yang hilang karena ketidakhadiran partisipan (dengan tingkat kekosongan >90%) dikecualikan dari analisis.

Mengikuti metodologi referensi, nilai numerik yang hilang diganti dengan angka 0, sementara nilai lainnya dikodekan dengan angka 66. Fitur dengan nilai hilang atau entri unik yang berlebihan (seperti n5, jnskeln, jnspeny, jnskec, rsnlain, udtyp, supbr, supfr) dihapus untuk menjaga integritas data.

D. Pembersihan Duplikasi Data

Analisis duplikasi dilakukan pada dua tingkat:

- Tingkat rekaman: tidak ditemukan entri yang terduplikasi.
- Tingkat fitur: ditemukan 121 fitur redundan, termasuk kolom duplikat seperti income_percapita dan income_percapitaNEW. Fitur duplikat ini dihapus untuk mengurangi dimensi dan menjaga konsistensi data.

Selain itu, dilakukan analisis nilai konstan, dan tidak ditemukan fitur dengan nilai konstan yang perlu dihapus.

E. Transformasi Data

Transformasi data mencakup beberapa langkah berikut:

- Variabel kategorik dikodekan menjadi nilai numerik mengikuti pedoman dari studi referensi.
- Dibuat fitur biner baru bernama ‘HaveHP’ untuk menunjukkan kepemilikan ponsel berdasarkan isi kolom nohp (1 untuk entri yang tidak kosong).
- Dilakukan koreksi tipe data untuk memastikan kolom numerik diatur dalam format float yang sesuai.

Tabel 2. Tabel Konversi chmsex

Konversi chmsex	
Sebelum Konversi	**Setelah Konversi**
boy	0
girl	1

F. Reduksi Dimensi dengan t-SNE

t-distributed Stochastic Neighbor Embedding (t-SNE) digunakan sebagai teknik reduksi dimensi non-linear untuk memvisualisasikan data berdimensi tinggi ke dalam ruang berdimensi lebih rendah. Algoritma ini mempertahankan struktur lokal data menggunakan pendekatan probabilistik. Persamaan utama t-SNE adalah:

$$P_{j|i} = \frac{\exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma_i^2}\right)}{\sum_{k\neq 1} \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma_i^2}\right)} \quad (1)$$

Keterangan:

- $\|x_i - x_j\|^2 =$ Jarak Euclidean kuadrat antara titik x_i dan x_j
- σ_i^2 = Parameter lebar Gaussian yang disesuaikan secara otomatis menggunakan *perplexity*
- $P_{j|i}$ = Probabilitas bersyarat bahwa x_j adalah tetangga dekat dari x_i.

G. Penentuan Jumlah Klaster: Metode Elbow

Sebelum dilakukan pemodelan, data dinormalisasi terlebih dahulu untuk mencegah rentang nilai yang terlalu jauh antar fitur. Selanjutnya, jumlah klaster ditentukan menggunakan Metode Elbow pada data yang telah dinormalisasi.

Metode ini menggunakan pendekatan *distortion score* untuk mengevaluasi efektivitas representasi data berdasarkan jumlah klaster. Rumus yang digunakan adalah:

$$Distortion\ (Inertia) = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (2)$$

Keterangan:

- k = Jumlah klaster
- C_i = Klaster ke-i
- $x \in C_i$= Data dalam klaster C_i
- μ_i = Titik pusat (centroid) dari klaster C_i
- $\|x - \mu_i\|^2$ = Jarak Euclidean kuadrat antara titik data x dan pusat klaster μ_i.

Pendekatan ini mengidentifikasi titik di mana penambahan jumlah klaster memberikan hasil representasi data yang semakin sedikit memberikan manfaat (diminishing return).

H. Implementasi K-Means

Algoritma K-Means digunakan sebagai metode utama dalam analisis karena efektivitasnya dalam mengenali pola dan mengelompokkan data.

Karakteristik Algoritma:

- Pendekatan pembelajaran tak terawasi
- Proses iteratif dalam membentuk klaster optimal
- Teknik clustering berbasis centroid

Keunggulan Implementasi:

- Skalabilitas terhadap dataset besar
- Adaptif terhadap data baru
- Efisiensi komputasi tinggi

Pertimbangan Kunci:

- Penentuan jumlah klaster (kkk) menggunakan Metode Elbow
- Perhitungan jarak menggunakan Jarak Euclidean:

$$d(x_i, c_k) = \sqrt{\sum_{j=1}^{n} (x_{ij} - c_{kj})^2} \quad (3)$$

Keterangan:

- x_i: Vektor fitur dari data ke-i
- c_k: Titik pusat dari klaster ke-k
- n: Jumlah fitur

Tahapan Implementasi:

1. Penempatan centroid awal secara acak
2. Penugasan titik data ke centroid terdekat
3. Penghitungan ulang posisi centroid
4. Iterasi hingga mencapai konvergensi

I. Justifikasi Metode

Pendekatan metodologis yang dipilih sesuai dengan praktik yang telah terbukti dalam analisis stunting, yaitu menggabungkan reduksi dimensi (t-SNE) dengan clustering (K-Means) untuk mengidentifikasi pola bermakna dalam data kesehatan. Setiap tahap dalam penelitian ini divalidasi berdasarkan penerapan pada studi sebelumnya dan disesuaikan dengan karakteristik dataset yang digunakan.

Pemilihan K-Means dibandingkan algoritma clustering lainnya didasarkan pada:

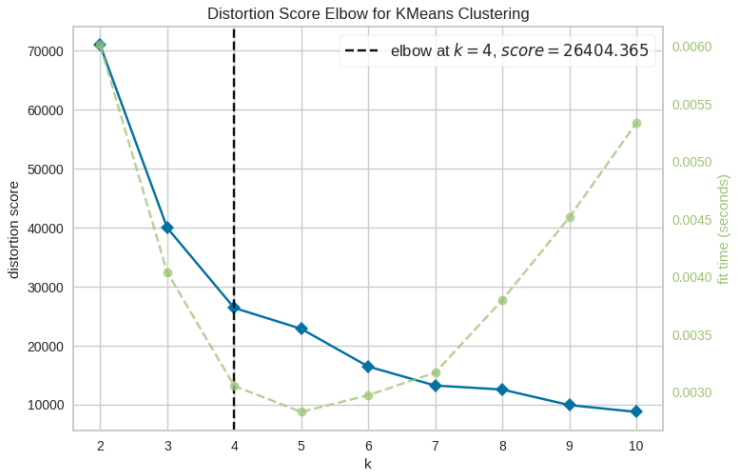
- Efektivitas yang telah terbukti pada dataset kesehatan serupa
- Efisiensi komputasi terhadap ukuran data yang digunakan
- Kemudahan interpretasi hasil untuk kepentingan kebijakan kesehatan masyarakat

3. PEMBAHASAN HASIL

Analisis dilakukan melalui tiga tahap: reduksi dimensi menggunakan t-SNE, optimasi jumlah klaster menggunakan Metode Elbow, dan implementasi akhir dengan algoritma K-Means. Pendekatan sistematis ini

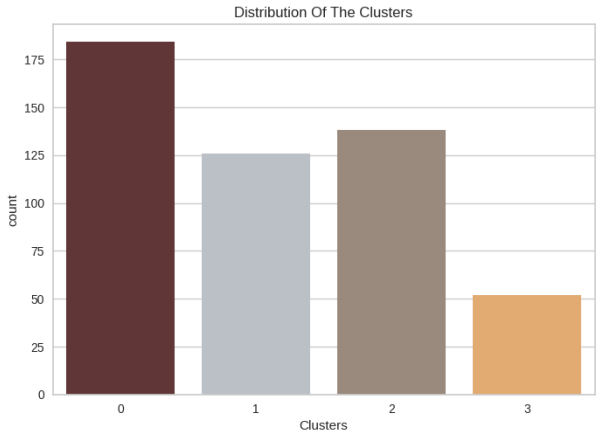
memungkinkan identifikasi pola yang komprehensif dalam dataset stunting sambil menangani tingginya dimensi fitur.

Jumlah klaster optimal ditentukan dengan Metode Elbow, menganalisis skor distorsi untuk rentang klaster (k=1 hingga k=10). Seperti terlihat pada Gambar 2, analisis menunjukkan titik optimal pada k=4, dengan nilai distorsi sebesar 26.404,365. Titik belok ini mengindikasikan bahwa penambahan klaster lebih dari k=4 tidak akan secara signifikan meningkatkan representasi data, sehingga empat klaster dianggap optimal untuk analisis ini.



Gambar 2. Hasil Perhitungan Menggunakan Metode Elbow

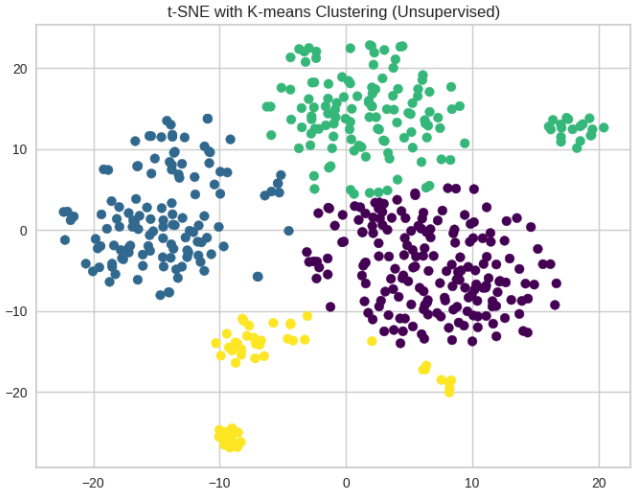
Setelah diperoleh jumlah klaster yang akan digunakan, proses selanjutnya adalah pemodelan menggunakan K-Means. Gambar 3 berikut menunjukkan distribusi hasil klaster yang dihasilkan.



Gambar 3. Distribusi Data dari Klaster yang Dihasilkan

Seperti terlihat pada Gambar 3, analisis klaster menunjukkan distribusi data yang berbeda-beda di antara keempat klaster yang terbentuk. Klaster 1 menunjukkan kepadatan populasi tertinggi, mengindikasikan bahwa sebagian besar data cenderung terkelompok dalam karakteristik yang dimiliki oleh klaster ini. Sebaliknya, Klaster 3 memiliki kepadatan terendah, yang menunjukkan bahwa hanya sedikit data yang sesuai dengan pola karakteristik dalam klaster tersebut. Hal ini dapat mengindikasikan adanya kelompok masyarakat dengan kondisi yang lebih unik atau berbeda dibandingkan klaster lainnya.

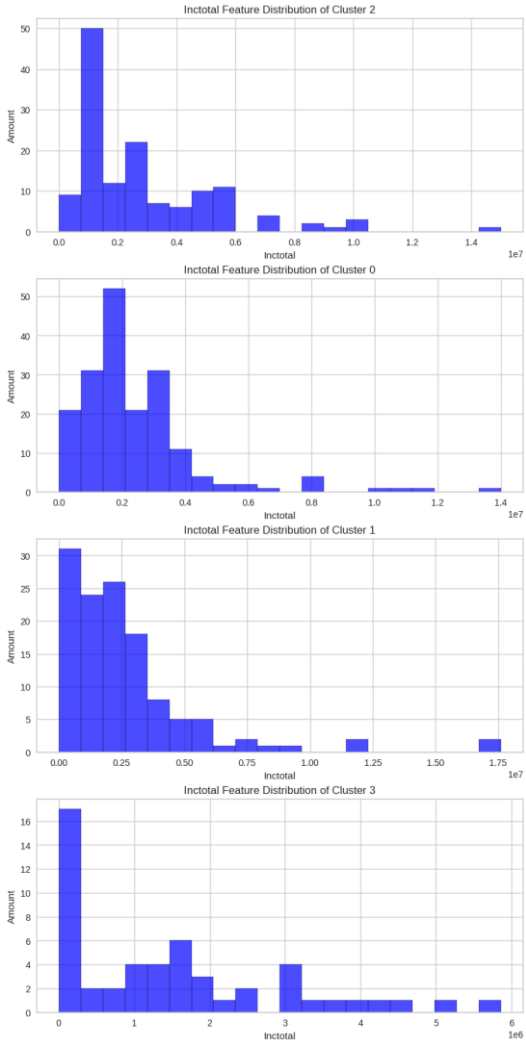
Visualisasi t-SNE pada Gambar 4 semakin menguatkan hasil tersebut dengan menunjukkan adanya pemisahan yang jelas antar klaster. Setiap klaster terdistribusi dalam ruang dua dimensi dengan jarak yang relatif terpisah, memperlihatkan bahwa data dalam satu klaster memiliki kemiripan yang tinggi (kohesi internal), namun berbeda secara signifikan dengan data di klaster lainnya (separasi antar-klaster). Pemisahan yang baik ini menjadi indikator bahwa proses klasterisasi yang dilakukan telah berhasil dalam mengidentifikasi pola-pola yang bermakna di dalam data, dan dapat menjadi dasar yang kuat dalam melakukan interpretasi atau pengambilan keputusan lebih lanjut berdasarkan karakteristik masing-masing klaster.



Gambar 4. Hasil Visualisasi Klaster pada t-SNE

Dari proses klasterisasi yang telah dilakukan sebelumnya, kita dapat melakukan pemprofilan terhadap masing-masing klaster berdasarkan fitur-fitur dalam dataset yang digunakan.

A. Rata-Rata Pendapatan Keluarga

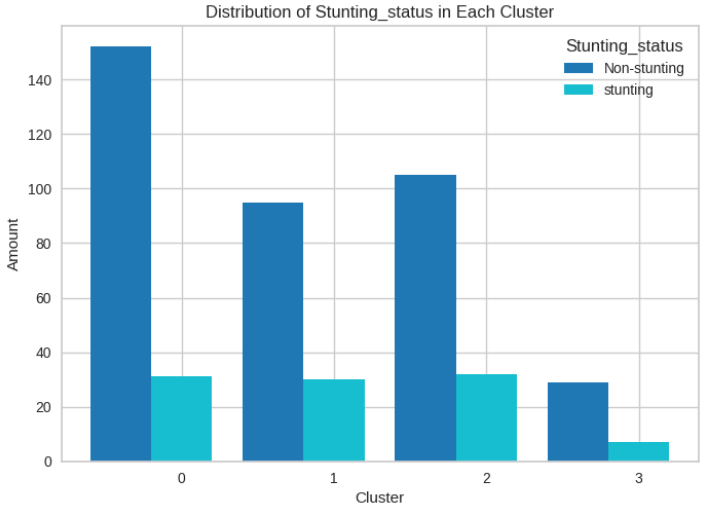


Gambar 5. Visualisasi Fitur Inctotal Berdasarkan Klaster

Pada Gambar 5, analisis distribusi pendapatan menunjukkan pola ekonomi yang berbeda di setiap klaster. Klaster 0 merepresentasikan kelompok berpendapatan menengah dengan rata-rata antara 1–4 juta rupiah dan

maksimum 14 juta rupiah. Klaster 1 muncul sebagai kelompok berpendapatan tertinggi, rata-rata 2,5 juta rupiah dan puncaknya mencapai 17,5 juta rupiah. Klaster 2 menunjukkan pola pendapatan menengah yang stabil dengan rata-rata 1 juta rupiah dan maksimum 14 juta rupiah. Sebaliknya, Klaster 3 mewakili segmen berpendapatan menengah ke bawah, dengan rata-rata antara 1–2 juta rupiah dan proporsi tinggi rumah tangga dengan pendapatan di bawah 250 ribu rupiah, serta maksimum hanya 6 juta rupiah. Stratifikasi ekonomi ini menunjukkan adanya hubungan antara tingkat pendapatan dan faktor sosial ekonomi lainnya dalam populasi yang diteliti.

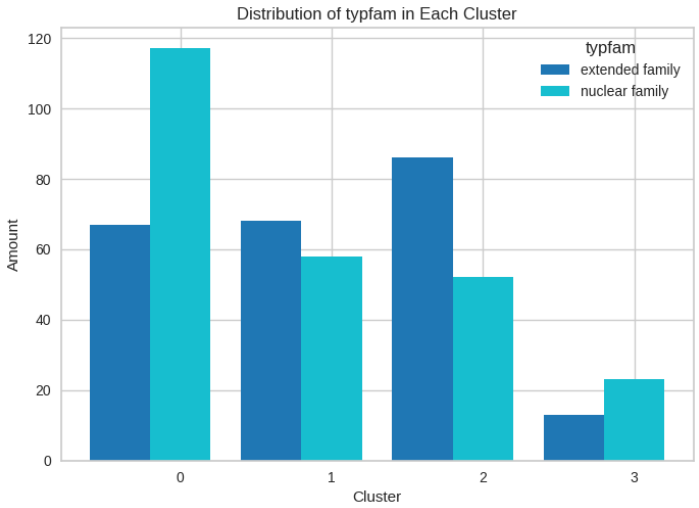
B. Status Stunting



Gambar 6. Visualisasi Fitur Stunting_status Berdasarkan Klaster

Seperti terlihat pada Gambar 6, analisis status stunting antar klaster menunjukkan variasi yang signifikan. Klaster 0 menunjukkan tingkat stunting terendah, diikuti oleh peningkatan bertahap pada Klaster 2, 1, dan 3 secara berurutan. Keberadaan kasus stunting di semua klaster menunjukkan bahwa meskipun faktor sosial ekonomi mempengaruhi tingkat stunting, variabel lain seperti akses layanan kesehatan dan kesadaran gizi juga memainkan peran penting.

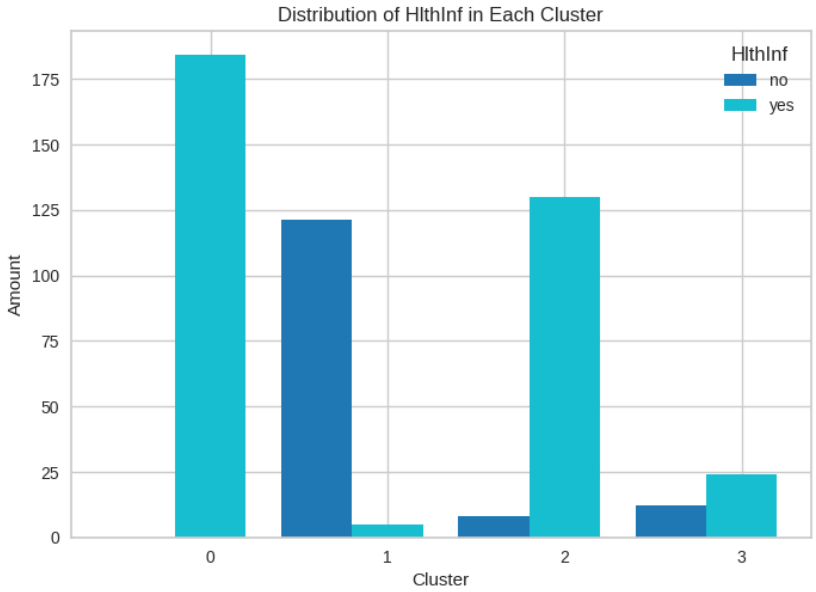
C. Tipe Keluarga



Gambar 7. Visualisasi Fitur typfam Berdasarkan Klaster

Pada Gambar 7, analisis struktur keluarga menunjukkan pola yang berbeda di setiap klaster. Klaster 0 dan Klaster 3 didominasi oleh keluarga inti, sementara Klaster 2 menunjukkan prevalensi keluarga besar. Klaster 1 memperlihatkan distribusi seimbang antara keluarga inti dan keluarga besar. Pola ini menunjukkan kemungkinan hubungan antara struktur keluarga dan faktor sosial ekonomi lain yang memengaruhi kesehatan anak.

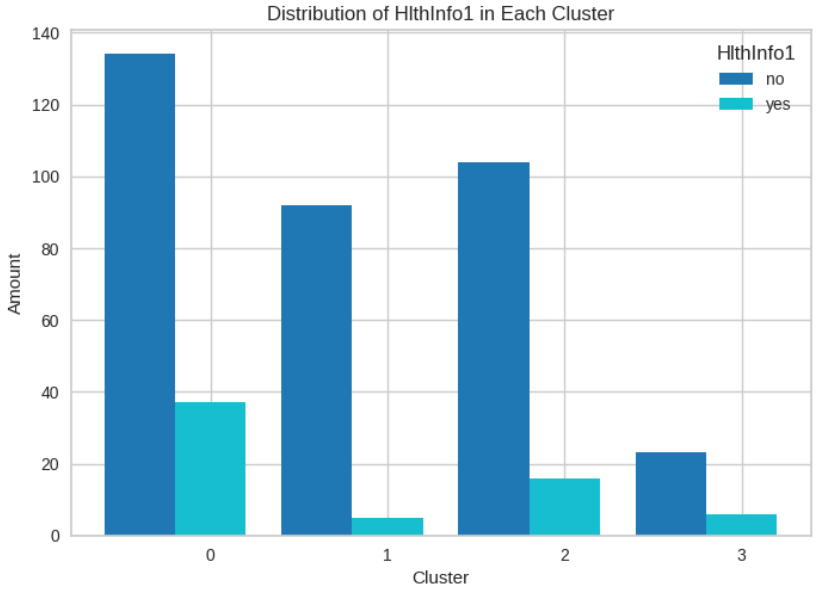
D. Akses Informasi Kesehatan



Gambar 8. Visualisasi Fitur HlthInf Berdasarkan Klaster

Pada Gambar 8, akses terhadap informasi kesehatan menunjukkan perbedaan yang mencolok antar klaster. Klaster 0 menunjukkan tingkat akses tertinggi, terutama melalui bidan. Klaster 1 menunjukkan pola yang mengkhawatirkan di mana sebagian besar ibu tidak memiliki akses informasi kesehatan. Klaster 2 dan 3 menunjukkan tingkat akses yang sedang. Temuan ini menekankan pentingnya peran bidan dalam pendidikan kesehatan masyarakat dan pencegahan stunting.

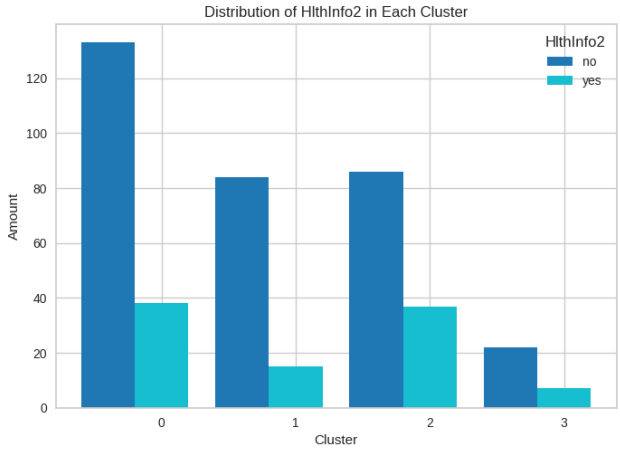
E. Informasi Kesehatan dari Petugas Gizi



Gambar 9. Visualisasi Fitur HlthInfo1 Berdasarkan Klaster

Dari Gambar 9, sebagian besar ibu di semua klaster tidak menerima informasi kesehatan dari petugas gizi, meskipun ada sebagian kecil yang menerima informasi tersebut.

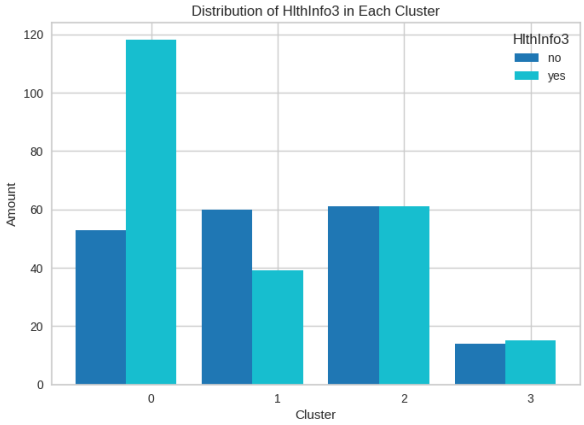
F. Informasi Kesehatan dari Kader



Gambar 10. Visualisasi Fitur HlthInfo2 Berdasarkan Klaster

Pada Gambar 10, sebagian besar ibu di semua klaster tidak menerima informasi kesehatan dari kader, meskipun ada sebagian kecil yang menerima informasi tersebut.

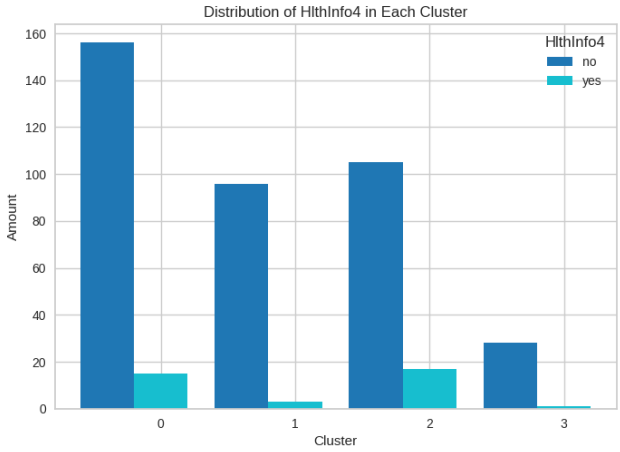
G. Informasi Kesehatan dari Bidan



Gambar 11. Visualisasi Fitur HlthInfo3 Berdasarkan Klaster

Pada Gambar 11, Klaster 0 sebagian besar memperoleh informasi kesehatan dari bidan, meskipun ada sebagian kecil yang tidak. Klaster 1 sebagian tidak memperoleh informasi, meski ada yang mendapatkannya. Klaster 2 dan 3 menunjukkan proporsi seimbang antara yang memperoleh dan tidak memperoleh informasi dari bidan.

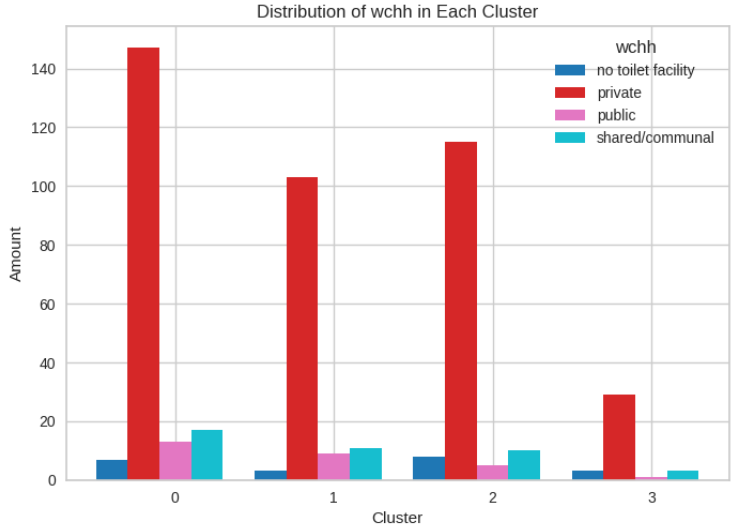
H. Informasi Kesehatan dari Perawat/Paramedis



Gambar 12. Visualisasi Fitur HlthInfo4 Berdasarkan Klaster

Pada Gambar 12, sebagian besar ibu di semua klaster tidak menerima informasi kesehatan dari perawat atau paramedis. Hanya sebagian kecil di Klaster 0, 1, dan 2 yang mendapatkannya.

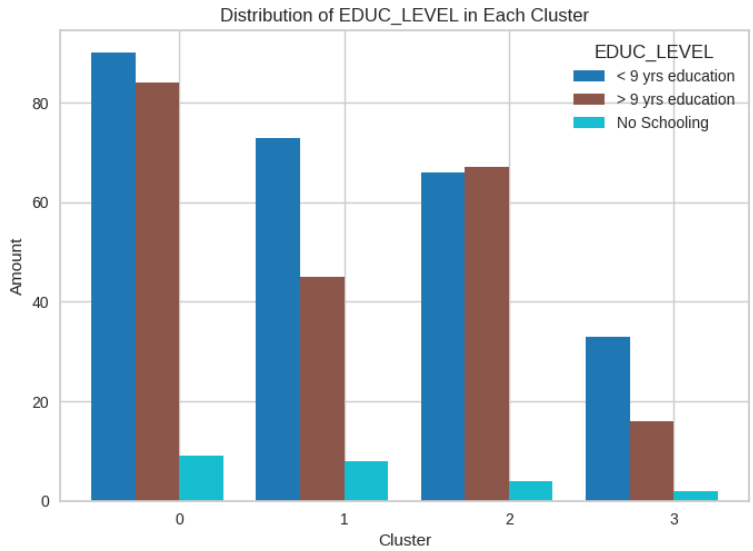
I. Kepemilikan Sarana Sanitasi



Gambar 13. Visualisasi Fitur wchh Berdasarkan Klaster

Dari Gambar 13, analisis akses sanitasi menunjukkan perbedaan tingkat infrastruktur antar klaster. Sebagian besar rumah tangga memiliki fasilitas sanitasi pribadi, tetapi terdapat ketimpangan, terutama di klaster berpendapatan rendah, di mana masih ada yang menggunakan fasilitas bersama atau bahkan tidak memiliki akses sama sekali. Hal ini menunjukkan korelasi kuat antara status ekonomi dan akses sanitasi.

J. Tingkat Pendidikan Orang Tua



Gambar 14. Visualisasi Fitur EDUC_LEVEL Berdasarkan Klaster

Pada Gambar 14, tingkat pendidikan orang tua menunjukkan variasi yang signifikan antar klaster. Klaster 2 memiliki tingkat pendidikan yang lebih tinggi, dengan sebagian besar orang tua menyelesaikan lebih dari 9 tahun pendidikan. Sebaliknya, Klaster 0, 1, dan 3 memiliki mayoritas orang tua dengan pendidikan di bawah 9 tahun. Sebagian kecil di semua klaster bahkan tidak memiliki pendidikan formal, menyoroti pentingnya intervensi di bidang ini. Terdapat korelasi antara tingkat pendidikan orang tua dan angka stunting yang menunjukkan bahwa pendidikan merupakan faktor kunci dalam kesehatan anak.

4. KESIMPULAN

Penelitian ini menganalisis pola stunting di Kabupaten Sambas, Provinsi Kalimantan Barat, menggunakan metode klasterisasi K-Means dengan penentuan jumlah klaster optimal (k = 4) melalui Metode Elbow. Analisis terhadap struktur keluarga menunjukkan bahwa Klaster 0 dan Klaster 3 didominasi oleh keluarga inti, sementara Klaster 2 lebih banyak terdiri atas keluarga besar. Klaster 1 menunjukkan distribusi yang seimbang antara keluarga inti dan keluarga besar.

Dalam hal akses informasi kesehatan, semua klaster kecuali Klaster 1 menunjukkan akses yang signifikan terhadap sumber daya kesehatan, dengan bidan sebagai sumber informasi utama. Namun, penyebaran informasi dari petugas gizi, kader, dan tenaga kesehatan lainnya masih terbatas.

Pemanfaatan layanan kesehatan menunjukkan bahwa pemantauan berat badan di Posyandu telah dilaksanakan secara menyeluruh di semua klaster, begitu pula dengan cakupan imunisasi yang tinggi. Namun, pelaksanaan pengukuran tinggi badan masih bervariasi dengan cakupan sekitar 50%, serta terdapat keterbatasan dalam pelaksanaan konseling gizi dan pemberian makanan tambahan.

Analisis sosial ekonomi menunjukkan bahwa Klaster 1 memiliki tingkat pendapatan tertinggi, meskipun hasil kesehatan masih bervariasi. Sementara itu, Klaster 3 menunjukkan tingkat pendapatan terendah yang disertai tantangan kesehatan yang lebih besar. Menariknya, tingkat pendidikan orang tua pada Klaster 2 melebihi 9 tahun dan berkorelasi dengan hasil kesehatan yang lebih baik, sementara klaster lainnya menunjukkan tren pendidikan yang mengkhawatirkan.

Penelitian ini mengidentifikasi faktor-faktor penting dalam pencegahan stunting, di mana tingkat pendidikan orang tua muncul sebagai faktor signifikan dalam menurunkan angka stunting. Akses terhadap informasi kesehatan, khususnya melalui bidan, menunjukkan korelasi yang kuat dengan hasil kesehatan yang lebih baik, sementara status ekonomi berpengaruh namun bukan satu-satunya penentu prevalensi stunting.

Penelitian ini dilakukan dengan batasan seleksi fitur tertentu untuk tujuan analisis. Studi lanjutan dapat mengembangkan temuan ini dengan memasukkan variabel sosial ekonomi tambahan, menganalisis perubahan pola stunting dari waktu ke waktu, mengevaluasi efektivitas intervensi pada setiap klaster, serta meninjau peran sistem dukungan komunitas. Temuan ini diharapkan dapat menjadi dasar bagi intervensi yang lebih terarah dalam upaya pencegahan stunting di konteks demografis yang serupa.

UCAPAN TERIMAKASIH

Penelitian ini dilakukan bekerja sama dengan Pusat Riset IDSS, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro.

REFERENCES

[1] P. Paul, B. Arra, M. Hakobyan, M. G. Hovhannisyan, and J. Kauhanen, “The determinants of under-5 age children malnutrition and the differences in the distribution of stunting–A study from Armenia,” PLOS ONE, vol. 16, no. 5, p. e0249776, May 2021, doi: 10.1371/journal.pone.0249776.

[2] K. Takele, T. Zewotir, and D. Ndanguza, “Understanding correlates of child stunting in Ethiopia using generalized linear mixed models,” BMC Public Health, vol. 19, no. 1, p. 626, Dec. 2019, doi: 10.1186/s12889-019-6984-x.

[3] M. Nemerimana et al., “Factors associated with recovery from stunting at 24 months of age among infants and young children enrolled in the Pediatric Development Clinic (PDC): A retrospective cohort study in rural Rwanda,” PLOS ONE, vol. 18, no. 7, p. e0283504, Jul. 2023, doi: 10.1371/journal.pone.0283504.

[4] N. Aguilera Vasquez and J. Daher, “Do nutrition and cash-based interventions and policies aimed at reducing stunting have an impact on economic development of low-and-middle-income countries? A systematic review,” BMC Public Health, vol. 19, no. 1, p. 1419, Dec. 2019, doi: 10.1186/s12889-019-7677-1.

[5] J. M. K. Aheto, “Simultaneous quantile regression and determinants of under-five severe chronic malnutrition in Ghana,” BMC Public Health, vol. 20, no. 1, p. 644, Dec. 2020, doi: 10.1186/s12889-020-08782-7.

[6] G. Tola, A. Kassa, M. Getu, B. Dibaba, and S. Neggesse, “Prevalence of stunting and associated factors among neonates in Shebadino woreda, Sidama region South Ethiopia; a community-based cross-sectional study 2022,” BMC Pediatr., vol. 23, no. 1, p. 276, Jun. 2023, doi: 10.1186/s12887-023-04080-4.

[7] M. Gebreayohanes and A. Dessie, “Prevalence of stunting and its associated factors among children 6–59 months of age in pastoralist community, Northeast Ethiopia: A community-based cross-sectional study,” PLOS ONE, vol. 17, no. 2, p. e0256722, Feb. 2022, doi: 10.1371/journal.pone.0256722.

[8] S. M. J. Rahman et al., "Investigate the risk factors of stunting, wasting, and underweight among under-five Bangladeshi children and its prediction based on machine learning approach," PLOS ONE, vol. 16, no. 6, p. e0253172, Jun. 2021, doi: 10.1371/journal.pone.0253172.
[9] R. J. Gansaonré, L. Moore, L. Bleau, J. Kobiané, and S. Haddad, "Stunting, age at school entry and academic performance in developing countries: A systematic review and meta-analysis," Acta Paediatr., vol. 111, no. 10, pp. 1853–1861, Oct. 2022, doi: 10.1111/apa.16449.
[10] B. A. Takele, L. D. Gezie, and T. S. Alamneh, "Pooled prevalence of stunting and associated factors among children aged 6–59 months in Sub-Saharan Africa countries: A Bayesian multilevel approach," PLOS ONE, vol. 17, no. 10, p. e0275889, Oct. 2022, doi: 10.1371/journal.pone.0275889.
[11] L. Chen, C. Zhong, and Z. Zhang, "Explanation of clustering result based on multi-objective optimization," PLOS ONE, vol. 18, no. 10, p. e0292960, Oct. 2023, doi: 10.1371/journal.pone.0292960.
[12] S. Rahman, V. E. Johnson, and S. S. Rao, "A Hyperparameter-Free, Fast and Efficient Framework to Detect Clusters From Limited Samples Based on Ultra High-Dimensional Features," IEEE Access, vol. 10, pp. 116844–116857, 2022, doi: 10.1109/ACCESS.2022.3218800.
[13] B. Yu, R. Xu, M. Cai, and W. Ding, "A clustering method based on multi-positive–negative granularity and attenuation-diffusion pattern," Inf. Fusion, vol. 103, p. 102137, Mar. 2024, doi: 10.1016/j.inffus.2023.102137.
[14] V. Whitaker et al., "Clustering of health-related behaviours within children aged 11–16: a systematic review," BMC Public Health, vol. 21, no. 1, p. 137, Dec. 2021, doi: 10.1186/s12889-020-10140-6.
[15] P. Ariza Colpas, E. Vicario, E. De-La-Hoz-Franco, M. Pineres-Melo, A. Oviedo-Carrascal, and F. Patara, "Unsupervised Human Activity Recognition Using the Clustering Approach: A Review," Sensors, vol. 20, no. 9, p. 2702, May 2020, doi: 10.3390/s20092702.
[16] M. Satoh, "Clustering of health behaviors among Japanese adults and their association with socio-demographics and happiness," PLOS ONE, vol. 17, no. 4, p. e0266009, Apr. 2022, doi: 10.1371/journal.pone.0266009.
[17] D. Janssen et al., "Clustering of 27,525,663 Death Records from the United States Based on Health Conditions Associated with Death: An Example of Big Health Data Exploration," J. Clin. Med., vol. 8, no. 7, p. 922, Jun. 2019, doi: 10.3390/jcm8070922.
[18] J. Liao, S. Scholes, C. Mawditt, S. T. Mejía, and W. Lu, "Comparing relationships between health-related behaviour clustering and episodic memory trajectories in the United States of America and England: a longitudinal study," BMC Public Health, vol. 22, no. 1, p. 1367, Dec. 2022, doi: 10.1186/s12889-022-13785-7.
[19] A. Diana et al., "Determination of modifiable risk factors for length-for-age z-scores among resource-poor Indonesian infants," PLOS ONE, vol. 16, no. 2, p. e0247247, Feb. 2021, doi: 10.1371/journal.pone.0247247.
[20] Local Burden of Disease Child Growth Failure Collaborators, "Mapping child growth failure across low- and middle-income countries," Nature, vol. 577, no. 7789, pp. 231–234, Jan. 2020, doi: 10.1038/s41586-019-1878-8.
[21] T. Siswati, S. Iskandar, N. Pramestuti, J. Raharjo, A. K. Rubaya, and B. S. Wiratama, "Impact of an Integrative Nutrition Package through Home Visit on Maternal and Children Outcome: Finding from Locus Stunting in Yogyakarta, Indonesia," Nutrients, vol. 14, no. 16, p. 3448, Aug. 2022, doi: 10.3390/nu14163448.
[22] J. F. Willumsen, "Improving children's diets to address the double burden of malnutrition: a healthy diet is key for all," Public Health Nutr., vol. 22, no. 17, pp. 3187–3188, Dec. 2019, doi: 10.1017/S1368980019002635.
[23] S. Ulijaszek, "Productive simplification in the use of anthropometric nutritional status," Eur. J. Clin. Nutr., vol. 74, no. 3, pp. 359–361, Mar. 2020, doi: 10.1038/s41430-020-0572-0.
[24] A. N. Sartika, M. Khoirunnisa, E. Meiyetriani, E. Ermayani, I. L. Pramesthi, and A. J. Nur Ananda, "Prenatal and postnatal determinants of stunting at age 0–11 months: A cross-sectional study in Indonesia," PLOS ONE, vol. 16, no. 7, p. e0254662, Jul. 2021, doi: 10.1371/journal.pone.0254662.
[25] T. R. Mokalla and V. V. Rao Mendu, "Application of quantile regression to examine changes in the distribution of Height for Age (HAZ) of Indian children aged 0–36 months using four rounds of NFHS data," PLOS ONE, vol. 17, no. 5, p. e0265877, May 2022, doi: 10.1371/journal.pone.0265877.
[26] M. Ahmed et al., "The relationship between maternal employment and stunting among 6–59 months old children in Gurage Zone Southern Nation Nationality People's region, Ethiopia: A comparative cross-sectional study," Front. Nutr., vol. 9, p. 964124, Oct. 2022, doi: 10.3389/fnut.2022.964124.

[27] G. Namirembe et al., "Child stunting starts in utero: Growth trajectories and determinants in Ugandan infants," Matern. Child. Nutr., vol. 18, no. 3, p. e13359, Jul. 2022, doi: 10.1111/mcn.13359.
[28] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," arXiv:1802.03426, 2018. [Online]. Available: https://arxiv.org/abs/1802.03426
[29] A. Kurniati, D. Martini, R. Hidayat, and Y. Ratnasari, "Stunting di Kabupaten Sambas: Analisis data lintas tahun," Jurnal Gizi dan Pangan, vol. 14, no. 2, pp. 101–112, 2018.
[30] M. Ahmed, S. Rahman, T. Islam, and R. Karim, "Clustering-based analysis of nutritional status among children: A data mining approach," BMC Public Health, vol. 22, no. 1, p. 1103, 2022.
[31] J. Namirembe, S. Tusiime, and M. N. Kansiime, "Exploring maternal education and child stunting using K-means clustering approach in Uganda," International Journal of Child Health and Nutrition, vol. 11, no. 1, pp. 25–35, 2022.
[32] I. H. Sumon, S. Akter, M. S. H. Sujon, M. K. Alam, S. Yasmin, S. Yeasmin, M. Kabir, and M. M. Hossain, "Unsupervised Machine Learning Approach to Investigate Risk Factors of Childhood Stunting in Cambodia: Evidence from 2021/22 Demographic and Health Survey," Children, vol. 11, no. 5, p. 736, 2024.
[33] K. Munawar, F. Mukhtar, M. Roy, N. Majeed, and M. Y. Jalaludin, "Parenting practices, child eating behavior, and stunting in Asia: A systematic review and clustering analysis," Nutrients, vol. 16, no. 2, p. 412, 2024.
[34] T. T. Tamir, S. A. Gezhegn, D. Dagnew, A. T. Mekonenne, G. T. Aweke, and A. M. Lakew, "Prevalence and Determinants of Stunting in Low- and Lower-Middle-Income African Countries: A Multi-Country Cluster Analysis," International Journal of Environmental Research and Public Health, vol. 21, no. 2, p. 310, 2024.