Journal Computer and Technology
Vol. 3, No. 1, Juli 2025, hlm. 57-64
ISSN : 3048-1880 (Online)
DOI : https://doi.org/10.69916/comtechno.v3i1.343

ANALISIS KLASTER DATA KESEHATAN BAYI DAN ANAK MENGGUNAKAN K-MEANS STUDI KASUS PROVINSI NUSA TENGGARA TIMUR

Supardianto[1], Lalu Mutawalli[*2]

[1]Teknik Informatika, Fakultas Teknologi Informasi dan Komunikasi, Universitas Teknologi Mataram, Indonesia
[2]Sistem Informasi, STMIK Lombok, Indonesia
Email: [1]supardianto88mkom@gmail.com, [2]laluallistilo@gmail.com

SEJARAH ARTIKEL
Diterima: 12.06.2025
Direvisi: 21.07.2025
Diterbitkan: 23.07.2025

ABSTRAK

Masalah gizi buruk dan stunting pada balita di Nusa Tenggara Timur (NTT) mencapai prevalensi yang sangat tinggi, dengan angka prevalensi stunting sebesar 42,6%. Penelitian ini bertujuan untuk memetakan risiko gizi balita di NTT menggunakan algoritma K-Means clustering, sebuah pendekatan berbasis machine learning untuk mengelompokkan wilayah berdasarkan tingkat risiko gizi. Data yang digunakan mencakup indikator status gizi, prevalensi stunting, dan akses layanan kesehatan. Metode Elbow digunakan untuk menentukan jumlah klaster optimal, menghasilkan delapan klaster. Analisis Silhouette Score sebesar 0,4045 menunjukkan struktur klaster cukup jelas meskipun terdapat potensi overlap. Hasil penelitian mengidentifikasi wilayah dengan risiko gizi tinggi, seperti Sumba Barat, Lembata, Ngada, dan Rote Ndao, yang memerlukan intervensi segera. Wilayah dengan risiko sedang, seperti Sumba Timur dan Manggarai Barat, membutuhkan program peningkatan status gizi. Sementara itu, wilayah dengan risiko rendah, seperti Kota Kupang dan Manggarai, tetap perlu pemantauan untuk mencegah risiko gizi lebih. Temuan ini memberikan dasar berbasis data yang kuat bagi pengambil kebijakan dalam merancang intervensi gizi yang lebih efektif dan efisien di NTT.

Kata Kunci: klasterisasi, k-means, risiko, gizi, bayi.

ABSTRACT

The prevalence of malnutrition and stunting among children under five in East Nusa Tenggara (NTT) remains alarmingly high, with a stunting rate of 42.6%. This study aims to map the nutritional risk in NTT using the K-Means clustering algorithm, a machine-learning approach to classify regions based on nutritional risk levels. The data includes indicators such as nutritional status, stunting prevalence, and healthcare access. The Elbow Method determined the optimal cluster number, resulting in eight clusters. A Silhouette Score of 0.4045 indicates reasonably clear clustering, despite potential overlaps. The findings identify high-risk areas, such as Sumba Barat, Lembata, Ngada, and Rote Ndao, requiring immediate interventions. Medium-risk regions, including Sumba Timur and Manggarai Barat, need to be targeted nutritional improvement programs. Meanwhile, low-risk areas such as Kota Kupang and Manggarai require monitoring to prevent overnutrition risks. These findings provide a robust data-driven foundation for policymakers to design more effective and efficient nutrition interventions in NTT.

Keywords: *k means, clustering, risk, health, baby.*

1. PENDAHULUAN

Masalah gizi pada balita di Nusa Tenggara Timur (NTT) sangat tinggi, prevalensi stunting di NTT mencapai 42,6% [1], presentase tersebut mengindikasikan bahwa NTT memiliki risiko jauh lebih tinggi disbanding rata-rata tingkat nasional sebesar 30,8% [2]. Gizi buruk memiliki risiko jangka pendek, dampanya terganggunya perkembangan otak dan pertumbuhan fisik, sedangkan jangka Panjang menyebabkan overweight, obesitas, dan meningkatkan risiko penyakit kardiovaskular [3]. Kondisi ini menciptakan tantangan yang signifikan dalam upaya meningkatkan status gizi balita. Masalah risiko gizi buruk dan stunting pada balita di NTT menunjukan urgensi untuk dilakukan pemetaan risiko gizi secara mendalam.

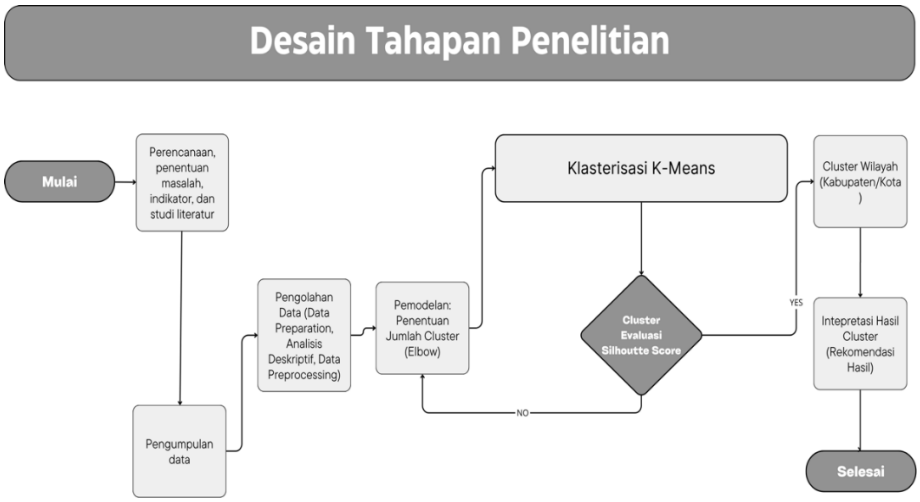
Pemetaan risiko pemetaan risiko gizi tidak hanya berfungsi sebagai alat diagnosis, tetapi juga sebagai fondasi untuk menciptakan strategi pencegahan dan penanganan gizi buruk yang lebih efektif di NTT. Saat ini pengelompokan wilayah di NTT berdasarkan tingat risiko gizi belum dilakukan, meskipun prevalensi stunting di NTT sangat tinggi. Oleh karena itu, penting untuk melakukan pemetaan risiko gizi secara mendalam guna mengidentifikasi wilayah yang paling terdampak, sehingga sumber daya dan program intervensi dapat daifokuskan pada area denga risiko tinggi. Hal ini akan membantu menciptakan strategi penaganan yang berbasis bukti dan mendukung pencapaian target nasional dalam menurunkan prevalensi stunting. Tanpa pemetaan yang tepat, upaya penagnan gizi di NTT beisiko menjadi tidak efisen dan kurang tepat sasaran. Pertanyaan penelitian agar dapat memahami fenomena lebih bagik adalah bagaimana pola distribusi risiko gizi buruk dan stunting pada balita diwilayah NTT. Selain itu, bagaimana pengelompokan wilayah di NTT. Tingkat risiko gizi dapat membantu memprioritaskan intervensi kesehatan, wilayah mana saja di NTT yang memiliki tingkat risiko gizi tertinggi berdasarkan hasil pengelompokan sehingga akan ditemukan solusi yang lebih efektif dalam menangani masalah gizi buruk dan stunting di NTT.

Berdasarkan permasalahan pengelompokan wilayah risiko gizi di NTT, pendekatan machine learning dapat menjadi upaya yang baik dalam merepresentasikan penyelesaian masalah yang ada. Terlebih pada pendekatan machine learing terdapat salah satu algoritma yaitu K-Means dapat menjadi solusi yang konfrehensif untuk mengidentifikasi dan mengelompokan wilayah berdasarkan tingkat risiko gizi. Berdasarkan permasalahan pengelompokan wilayah risiko gizi di Nusa Tenggara Timur (NTT), pendekatan machine learning dapat menjadi solusi yang efektif dalam menangani isu ini. Di antara berbagai algoritma yang tersedia, K-Means clustering muncul sebagai metode yang komprehensif untuk mengidentifikasi dan mengelompokkan wilayah berdasarkan tingkat risiko gizi. Dengan menggunakan pendekatan ini, kita dapat menganalisis data dan mengelompokkan daerah-daerah yang memiliki karakteristik serupa, sehingga memudahkan dalam merancang intervensi yang lebih tepat sasaran[4].

Metode K-Means clustering telah banyak digunakan di berbagai sektor, termasuk dalam bidang kesehatan, untuk mengelompokkan data berdasarkan karakteristik tertentu [5]. Dalam bidang kesehatan gizi untuk balita dan anak, metode K-Means sangat penting karena dapat membantu dalam mengidentifikasi area atau kelompok yang memiliki risiko tinggi terhadap stunting dan masalah gizi buruk [6]. Metode K-Means efektif dalam mengelompokkan data terkait kesehatan, termasuk prevalensi penyakit, status gizi, dan faktor risiko lainnya [7]. Penerapan algoritma K-Means clustering belum pernah dilakukan untuk pemetaan gizi di wilayah NTT, sehingga penerapanya pada penelitian ini dapat mengidentifikasi dan mengelompokan wilayah beresiko tinggi stunting berdasarkan karakteristik sfesifik, sehingga memungkinkan intervensi yang lebih tepat sasaran. Penelitian ini menyediakan instrument analisis yang dapat membantu *stakeholder* merancang strategi penanggulangan stunting yang efektif, sehingga dengan pendekatan berbasis data dapat menjadi model strategis dalam penanganan wilayah penaganan gizi balita di Provinsi Nusa Tenggara Barat.

2. METODE PENELITIAN

Dalam upaya mengatasi masalah gizi buruk dan stunting yang semakin meningkat di Nusa Tenggara Timur (NTT), penelitian ini bertujuan untuk melakukan pemetaan risiko gizi dengan pendekatan yang berbasis data. Melalui langkah-langkah sistematis yang dijelaskan pada gambar 1 tahapan penelitian, menggambarkan alur bagaimana penelitian ini dilakukan secara sistematis.



Gambar 1. Tahapan penelitian

Berikut adalah penjelasan mendalam mengenai tahapan penelitian, meliputi tujuh tahapan utama yaitu, perencanaan, pengumpulan data, pengolahan data, penentuan jumlah cluster, klaster k-means, evaluasi klaster, justifikasi hasil klaster, dan intepretasi hasil.

a. Perencanaan
Tahap awal penelitian dimulai dengan identifikasi masalah utama, yaitu tingginya prevalensi stunting dan gizi buruk di Nusa Tenggara Timur (NTT). Penentuan indikator dilakukan berdasarkan data prevalensi stunting, status gizi, akses layanan kesehatan, dan faktor lingkungan. Studi literatur dilakukan untuk memahami teori dan metode yang relevan, termasuk algoritma K-Means clustering yang telah terbukti efektif dalam pengelompokan data berbasis karakteristik tertentu [5] *Pattern Recognition and Machine Learning* [8] menjadi dasar dalam memahami metode clustering.

b. Pengumpulan data
Sumber data berasal dari Dinas Kesehatan Provinsi Nusa Tenggara Timur yang dipublikaiskan oleh Badan Pusat Statistik Kabupaten Sumba Barat daya yang dipublikasikan pada laman: *https://sumbabaratdayakab.bps.go.id/id/statistics-table/2/MzMyIzI=/jumlah-balita-dan-status-gizi-menurut-kabupaten-kota.html*. Dalam studi ini dianalisis delapan variabel, terdiri dari satu kategorik dan tujuh numerik. Variabel kategorik mencakup wilayah administratif (Kabupaten/Kota), sedangkan variabel numerik mencakup tahun pengumpulan (2021–2023) dan status gizi balita: gizi buruk, kurang, normal, risiko gizi lebih, gizi lebih, serta obesitas berdasarkan jumlah individu dalam tiap kategori klasifikasi gizi.

c. Pengolahan Data
Data yang dikumpulkan dibersihkan dari nilai yang hilang (missing values) atau data yang tidak relevan, dan distandarkan agar memiliki skala yang sama untuk analisis clustering [9]. Distribusi data dianalisis untuk menggambarkan pola awal, seperti rata-rata prevalensi stunting di setiap wilayah [10]. Normalisasi data dilakukan untuk memastikan semua variabel memiliki bobot yang sama, dan outlier yang dapat memengaruhi hasil clustering dihapus [11]. Tahapan preprocessing data balita melibatkan dua teknik utama. Imputasi nilai hilang menggunakan pendekatan median dengan algoritma *SimpleImputer*, mengikuti konsep sentral tendensi dalam distribusi data numerik. Proses normalisasi dilakukan memakai *StandardScaler* berbasis teori z-score, bertujuan menyeragamkan skala fitur numerik agar memiliki rata-rata nol dan standar deviasi satu, mendukung analisis berbasis jarak dan algoritma pembelajaran mesin.

d. Penentuan jumlah cluster
Metode Elbow menentukan jumlah cluster optimal dengan menghitung inertia untuk berbagai jumlah cluster. Titik di mana penurunan inertia melambat disebut elbow point, menunjukkan jumlah cluster ideal [12]. Metode ini efektif dalam menentukan jumlah cluster yang tepat.

e. Pemodelan Klaster K-Means
Algoritma K-Means digunakan untuk mengelompokkan wilayah berdasarkan tingkat risiko gizi. Algoritma ini bekerja dengan meminimalkan jarak antara data dan centroid cluster. Setiap cluster merepresentasikan tingkat risiko tertentu [12].

f. Evaluasi cluster
Evaluasi clustering menggunakan Silhouette Score mengukur seberapa baik data dikelompokkan dalam cluster. Nilai mendekati 1 menunjukkan clustering yang baik, sedangkan nilai mendekati -1 menunjukkan kesalahan pengelompokan [13]. Metode ini berguna untuk menilai kualitas clustering.

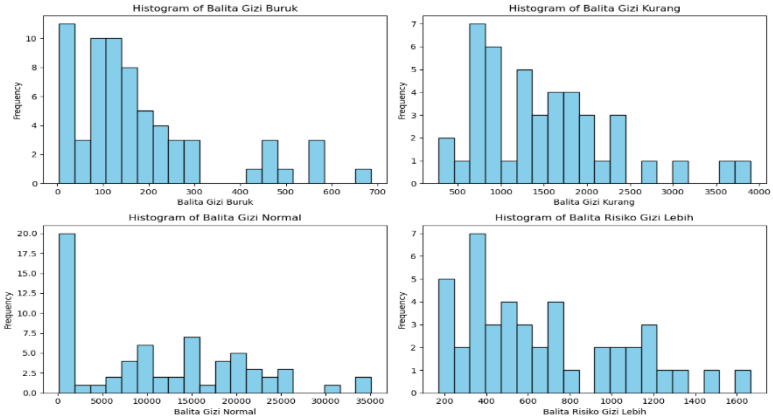
g. Hasil cluster wilayah
Hasil clustering menunjukkan pengelompokan wilayah di NTT berdasarkan tingkat risiko gizi. Setiap cluster dianalisis untuk memahami karakteristik wilayah, seperti prevalensi stunting, akses layanan kesehatan, dan faktor lingkungan. Analisis ini memberikan wawasan yang mendalam tentang distribusi risiko gizi di NTT

h. Intepretasi Hasil
Hasil clustering diinterpretasikan untuk mengidentifikasi wilayah dengan risiko tertinggi. Berdasarkan hasil ini, rekomendasi disusun untuk program intervensi, seperti pemberian makanan tambahan, edukasi gizi, dan peningkatan akses layanan kesehatan. Pendekatan berbasis data ini diharapkan dapat meningkatkan efisiensi dan efektivitas intervensi

3. HASIL DAN PEMBAHASAN

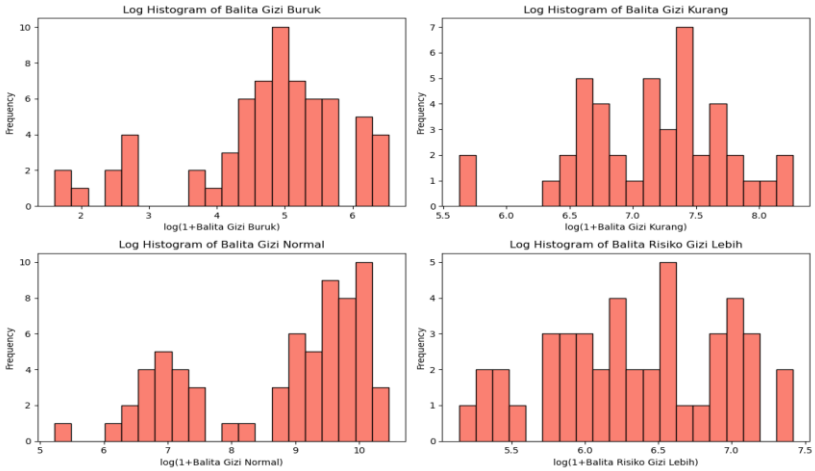
3.1. Impelemtasi

Analisis distribusi data merupakan Langkah awal dalam penelitian ini karena memberikan pemahaman mendalam tentang data. Pada Gambar 2 menunjukkan histogram distribusi status gizi balita, dengan pola distribusi yang tidak merata. Kolom untuk "Balita Gizi Buruk" dan "Gizi Normal" menunjukkan skewed distribution, yang dapat memengaruhi analisis statistik. Selain itu, terdapat outlier pada "Balita Gizi Normal" yang dapat mengganggu analisis kuantitatif. Skala yang berbeda antar kolom juga dapat memengaruhi algoritma seperti K-Means, karena jarak Euclidean menjadi terdistorsi.



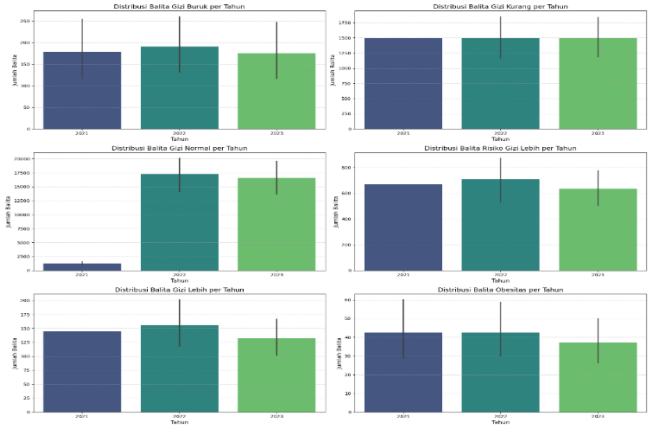
Gambar 2. Distribusi Status Gizi Balita: Histogram Gizi Buruk, Gizi Kurang, Gizi Normal, dan Risiko Gizi Lebih

Menerapkan transformasi logaritmik pada data untuk mengatasi masalah distribusi yang tidak merata. Dengan menggunakan $log(1 + x)$, nilai-nilai kecil dan ekstrem dapat dikelola, sehingga menghasilkan histogram yang lebih seimbang dan informatif. Hasilnya digambarkan pada Gambar 3 menunjukkan distribusi yang lebih layak untuk analisis, dengan pola yang lebih jelas pada setiap kategori status gizi balita. Histogram ini memudahkan identifikasi pola dan outlier, serta meningkatkan akurasi dalam analisis statistik dan clustering.



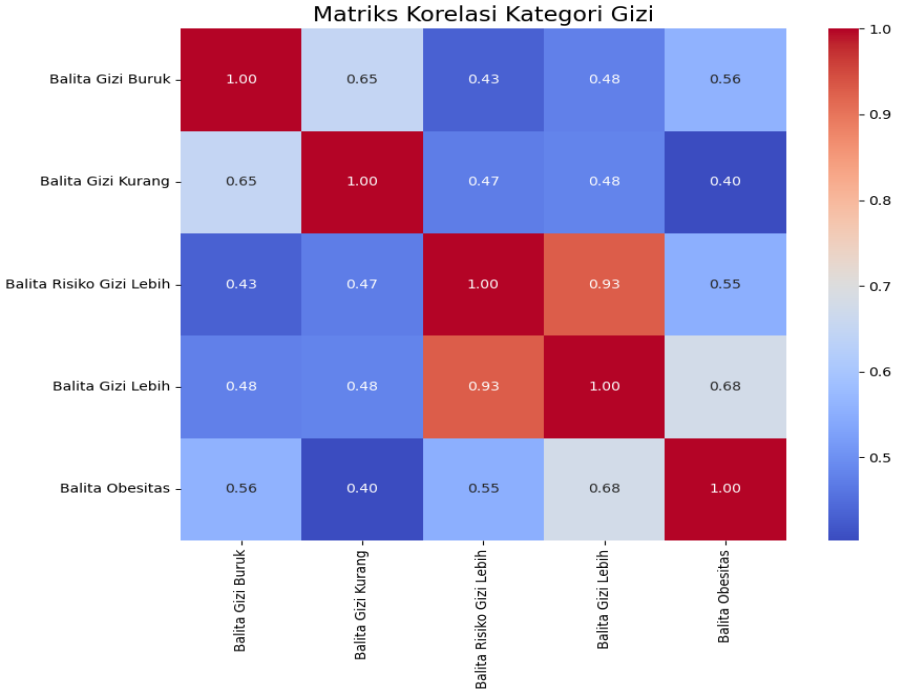
Gambar 3. Log Histogram Distribusi Status Gizi Balita di Nusa Tenggara Timur

Gambar 4 menunjukkan tren distribusi status gizi balita di NTT (2021–2023). Balita gizi buruk dan kurang relatif stabil, sementara balita gizi normal meningkat signifikan pada 2022–2023. Tren ini mencerminkan perbaikan gizi, meskipun risiko gizi lebih dan obesitas tetap perlu perhatian.



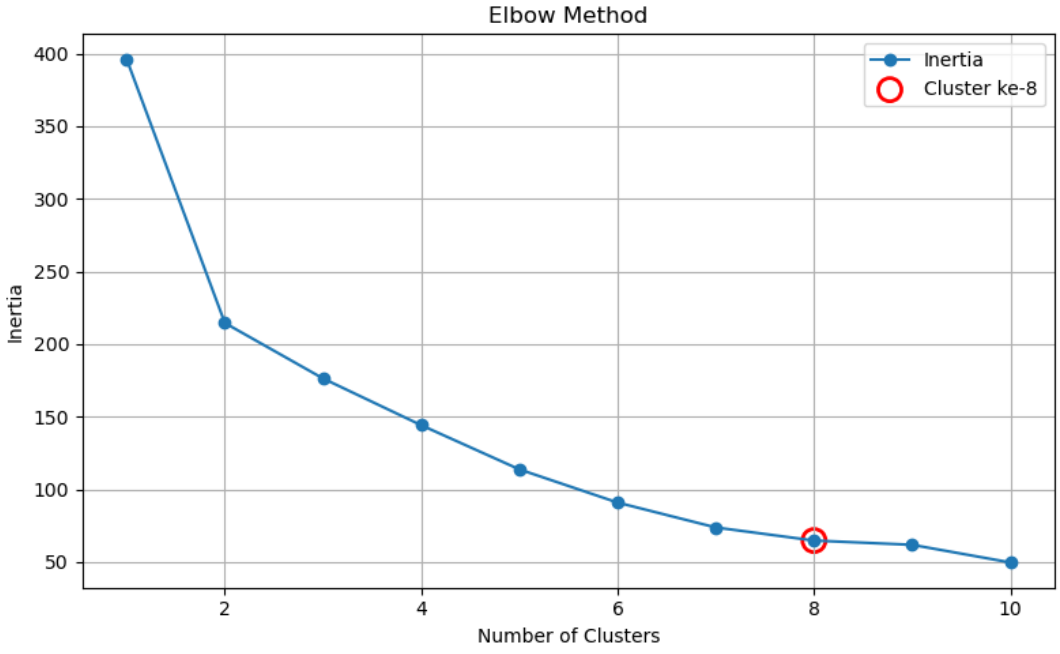
Gambar 4. Tren Distribusi Status Gizi Balita di NTT (2021-2023

Matriks korelasi pada gambar 5 menunjukkan hubungan positif kuat antara "Balita Risiko Gizi Lebih" dan "Balita Gizi Lebih" (0,93), serta hubungan sedang antara "Balita Gizi Buruk" dan "Balita Gizi Kurang" (0,65). Ini mencerminkan keterkaitan antar kategori gizi di NTT.



Gambar 5. Matriks Korelasi Status Gizi Balita di NTT

Gambar 6 menunjukkan Metode Elbow untuk menentukan jumlah klaster optimal dalam analisis klaster. Sumbu-x menunjukkan jumlah klaster, sedangkan sumbu-y menunjukkan nilai **inertia** (jumlah kuadrat jarak antar titik ke pusat klaster). Klaster optimal berada pada titik "siku" grafik, yaitu ketika penurunan inertia mulai melambat. Berdasarkan gambar, klaster ideal adalah **8**, dengan lingkaran merah menandai titik tersebut.



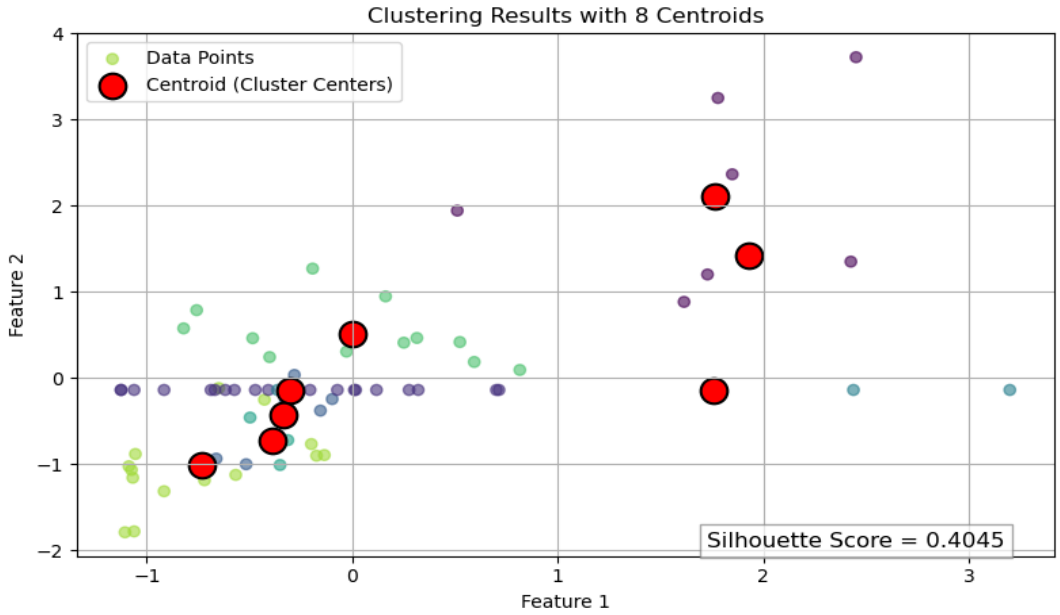
Gambar 6. Penentuan Jumlah Klaster Optimal dengan Metode Elbow

Tabel 1. Menunjukkan nilai *inertia* pada penerapan algoritma *K-Means* dengan jumlah klaster k=1 hingga k=10 terhadap data status gizi balita. Nilai *inertia* tertinggi tercatat pada k=1 (396.00), menandakan distribusi data tanpa segmentasi. Penurunan bertahap terjadi pada k=2 (214.51), k=3 (176.35), k=4 (144.19), dan k=5 (113.77), yang mengindikasikan peningkatan pemadatan klaster. Perubahan *inertia* pada k=6 (90.96), k=7 (73.83), dan k=8 (64.83) menunjukkan efisiensi komputasional yang berlanjut. Nilai pada k=9 (61.88) dan k=10 (49.74) mengalami penurunan melandai, mencerminkan pengaruh kompleksitas klaster terhadap struktur data. Variasi nilai tersebut berimplikasi pada identifikasi titik infleksi yang penting dalam pendekatan Elbow, guna mengevaluasi representasi segmentasi yang optimal. Pemilihan k=8 didukung oleh penurunan inertia yang signifikan hingga titik infleksi, sebagaimana dijelaskan dalam pendekatan Elbow. Nilai inertia mencapai 64.83 pada k=8, menandai transisi dari efisiensi segmentasi menuju tren pelandaian. Kriteria ini mencerminkan kompromi antara kedalaman klasifikasi dan kestabilan struktur data, sehingga k=8 memberikan representasi segmentasi yang optimal secara komputasional dan struktural.

Tabel 1. Contoh Tabel dengan

Jumlah Cluster	**Nilai Intertia**
1	396.000000
2	214.509345
3	176.351453
4	144.190485
5	113.771322
6	90.956027
7	73.825687
8	64.827158
9	61.888917
10	49.742482

Hasil klasterisasi pada gambar 7 menujukkan, dengan **8 centroid** menggunakan algoritma clustering. Titik-titik data diwarnai berdasarkan klaster masing-masing, sementara lingkaran merah menunjukkan posisi centroid klaster. Distribusi data menunjukkan bahwa klasterisasi cukup baik, meskipun terdapat beberapa titik yang mungkin berada di tepi antar klaster. Nilai **Silhouette Score = 0.4045** mengindikasikan bahwa struktur klaster cukup jelas, walau ada potensi overlap antar klaster.



Gambar 6. **Hasil Klaster dengan 8 Centroid dan Sillhotte Score**

3.2. Hasil Pengujian

Hasil analisis pada Tabel 2 menunjukkan distribusi kabupaten/kota di NTT berdasarkan nilai ambang batas cluster. Temuan utama adalah adanya variasi signifikan dalam status gizi antar wilayah. Cluster dengan nilai ambang tinggi (misalnya, Cluster 1 dan 5) cenderung mencakup wilayah dengan akses lebih baik terhadap sumber

daya, sedangkan cluster dengan nilai rendah (Cluster 6 dan 7) menunjukkan wilayah yang mungkin menghadapi tantangan gizi lebih besar. Pola ini mencerminkan ketimpangan regional dalam status gizi, yang relevan untuk intervensi kebijakan berbasis wilayah.

Tabel 2. Distribusi Kabupaten/Kota di NTT Berdasarkan Ambang Batas Cluster Gizi

Cluster	Nilai ambang batas	Nama Kabupaten Kota
1	928.0 – 1300.0	Kupang, Timor Tengah Selatan, Sumba Barat Daya, Kota Kupang, Kupang, Timor Tengah Selatan, Sumba Barat Da ya
2	552.0 – 552.0	Sumba Barat, Sumba Timur, Timor Tengah Utara, Belu, A lor, Lembata, Flores Timur, Sikka, Ende, Ngada, Manggar ai, Rote Ndao, Manggarai Barat, Sumba Tengah, Sumba B arat Daya, Nagekeo, Manggarai Timur, Sabu Raijua, Mala ka
3	687.0 – 1075.0	Sumba Timur, Ende, Manggarai Barat, Manggarai Timur, Sumba Timur, Ende, Manggarai Barat
4	552.0 – 552.0	Kupang, Timor Tengah Selatan, Kota Kupang
5	1200.0 – 1666.0	Manggarai, Manggarai, Manggarai Timur
6	331.0 – 734.0	Timor Tengah Utara, Belu, Alor, Flores Timur, Sikka, Mal aka, Timor Tengah Utara, Belu, Alo, Flores Timur, Sikka, Malaka
7	170.0 – 553.0	Sumba Barat, Lembata, Ngada, Rote Ndao, Sumba Tengah , Nagekeo, Sabu Raijua, Sumba Bara, Lembata, Ngada, Ro te Ndao, Sumba Tengah, Nagekeo, Sabu Raijua
8	1152.0 – 1152.0	Kota Kupang

Berdasarkan hasil ambang batas yang ditunjukkan dalam tabel 1, kabupaten/kota di NTT dapat dikategorikan dalam tiga tingkat risiko gizi. **Risiko gizi tinggi** ditemui di wilayah seperti (Sumba Barat, Lembata, Ngada, Rote Ndao, Sumba Tengah, Nagekeo, dan Sabu Raijua), yang menunjukkan tantangan signifikan terkait prevalensi gizi buruk atau kurang. **Risiko gizi sedang** mencakup (Sumba Timur, Ende, Manggarai Barat, Manggarai Timur), yang menunjukkan kebutuhan intervensi moderat untuk meningkatkan status gizi. Sementara itu, **risiko gizi kurang** ditemukan di Kota Kupang, Manggarai, dan Timor Tengah Selatan, yang relatif lebih baik namun tetap perlu pemantauan untuk mencegah risiko gizi lebih atau obesitas. Klasifikasi ini penting untuk memprioritaskan intervensi gizi sesuai dengan tingkat risiko yang ada di masing-masing wilayah.

4. KESIMPULAN

Penelitian ini berhasil melakukan pemetaan risiko gizi pada balita di Nusa Tenggara Timur (NTT) dengan memanfaatkan algoritma K-Means clustering, yang menghasilkan delapan klaster berdasarkan ambang batas gizi. Wilayah yang teridentifikasi dengan risiko gizi tinggi meliputi Sumba Barat, Lembata, Ngada, Rote Ndao, Sumba Tengah, Nagekeo, dan Sabu Raijua, yang memerlukan perhatian dan intervensi segera. Di sisi lain, risiko gizi sedang ditemukan di Sumba Timur, Ende, Manggarai Barat, dan Manggarai Timur, yang juga memerlukan upaya untuk meningkatkan status gizi. Sementara itu, Kota Kupang, Manggarai, dan Timor Tengah Selatan menunjukkan risiko gizi rendah, tetapi tetap harus dipantau. Pemetaan ini memberikan dasar yang kuat bagi pemerintah dan pemangku kebijakan untuk merancang intervensi yang lebih efektif dan berbasis data dalam menangani masalah gizi di NTT.

DAFTAR PUSTAKA

[1] I. J. Fafo, S. M. Davidson, and K. D. Tauho, “Potret Anak Stunting Usia 2-5 Tahun,” *Jurnal Kesehatan Primer*, vol. 8, no. 2, pp. 110–119, Nov. 2023, doi: https://doi.org/10.31965/jkp.v8i2.1166.

[2] M. F. Akbar, “Prevalensi Stunting Nasional Turun Jadi 19,8 Persen, Menkes RI Targetkan 18,8 Persen di Tahun 2025,” Kabupaten Kutai Karta Negara.

[3] Nursyamsi, A. Nurlinda, and M. Ikhtiar, “Karakteristik Balita Stunting di Wilayah Kerja UPTD Puskesmas Pakkae Kabupaten Barru,”*Journal of Muslim Community Health (JMCH) 2023*, vol. 4, no. 3, pp. 169–175, Oct. 2023, doi: 10.52103/jmch.v4i3.1141.

[4] L. Mutawalli, M. Taufan, A. Tantoni, and I. F. Suhriani, “Segmentasi Risiko Kesehatan Bayi dan Balita Menggunakan Algoritma K-Means,” *BULLETIN OF COMPUTER SCIENCE RESEARCH*, vol. 5, no. 4, pp. 382–391, Jun. 2025, doi: 10.47065/bulletincsr.v5i4.504.

[5] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. California: Morgan Kaufmann, 2006.

[6] A. I. Silitonga, Z. A. Nabila, C. R. Z. Lubis, S. Nurdina, and Haryadi, "Klasterisasigizi Buruk Dan Stuntingdi 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.

[7] A. R. Romadhoni, A. Romadhoni, D. Taqiyyudin, and A. Abid, "Penerapan Algoritma K-Means dalam Analisis Tingkat Kesehatan Pada Populasi Bayi dan Balita di Kota Semarang," *Journal of Data Science Theory and Application*, vol. 3, no. 1, pp. 32–41, May 2024.

[8] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. London: Springer, 2023.

[9] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer-Verlag, 2006.

[10] 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

[11] K. Aggarwal, C. Zhang, J. C. Campbell, A. Hindle, and E. Stroulia, "The power of system call traces: predicting the software energy consumption impact of changes.," in *CASCON*, 2014, pp. 219–233.

[12] 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.

[13] R. Hidayati, A. Zubair, A. H. Pratama, and L. Indana, "Analisis Silhouette Coefficientpada 6 Perhitungan Jarak K-Means Clustering Silhouette Coefficient Analysis in 6 Measuring Distancesof K-Means Clustering," *Techno.COM*, vol. 20, no. 2, pp. 186–197, 2021.