

Klasifikasi *Stunting* Balita Nusa Tenggara Timur Berdasarkan Faktor Sosial Ekonomi Kesehatan dengan *Supervised Learning*

Bagus Candra Setiawan^{1,*}, Alma Ainul Haqi², Arum Nilawati³, M. Fajrin Ramadhan⁴, Muhammad Riefky⁵

^{1, 2), 3), 4), 5)} Studi Demografi dan Pencatatan Sipil, Universitas Sebelas Maret
Jl. Kolonel Sutarto No.150K, Jebres, Kec. Jebres, Kota Surakarta, Jawa Tengah 57126

e-mail: baguscandra98@student.uns.ac.id¹⁾, haqialma629@student.uns.ac.id²⁾,

arumnilawati27@student.uns.ac.id³⁾, m.fajrinramadhan@student.uns.ac.id⁴⁾,

muhammadriefky999@staff.uns.ac.id⁵⁾

*corresponding author

(Naskah masuk : 27 Agustus 2025 Diterima untuk diterbitkan : 17 April 2026)

ABSTRAK

Provinsi Nusa Tenggara Timur (NTT) masih menghadapi tantangan serius dalam penurunan prevalensi *stunting*, di mana 27,3% kabupaten/kota pada tahun 2024 tergolong dalam kategori *stunting* tinggi. Riset ini menggunakan data sekunder kuantitatif dari Badan Pusat Statistik (BPS) Provinsi NTT tahun 2024. Data dianalisis menggunakan algoritma klasifikasi *supervised learning* yaitu *K-Nearest Neighbor (KNN)* dan *Decision Tree*. Variabel yang dipakai mencakup faktor sosial ekonomi dan kesehatan, seperti persentase *stunting*, persentase penduduk miskin, Rata-rata Lama Sekolah (RLS), persentase balita *wasting*, dan persentase sanitasi layak. Tujuan dari penelitian ini yaitu untuk menganalisis prevalensi *stunting* balita di NTT tahun 2024 berdasarkan faktor sosial ekonomi dan kesehatan, serta menentukan visualisasi dari algoritma klasifikasi terbaik berdasarkan nilai akurasi. Temuan menunjukkan algoritma *Decision Tree* memberikan hasil terbaik dengan nilai akurasi, sensitivitas, dan spesifisitas sebesar 100%, mengungguli *KNN* yang hanya mencapai akurasi 80%. Analisis kepentingan fitur dari model *Decision Tree* menunjukkan bahwa tiga faktor utama yang paling berpengaruh adalah rata-rata lama sekolah (29,1%), persentase balita *wasting* (28%), dan persentase penduduk miskin (25,4%). Berdasarkan temuan penelitian menegaskan bahwa percepatan penurunan *stunting* di NTT perlu difokuskan pada tiga strategi prioritas, yaitu peningkatan kualitas pendidikan, intervensi gizi balita (penanganan *wasting*), serta pengentasan kemiskinan.

Kata Kunci: Balita, *Decision Tree*, *KNN*, Nusa Tenggara Timur, *Stunting*.

ABSTRACT

The province of East Nusa Tenggara (NTT) still faces serious challenges in reducing the prevalence of *stunting*, with 27.3% of its districts/cities categorized as having high *stunting* rates in 2024. This research uses quantitative secondary data from the Central Bureau of Statistics (BPS) of East Nusa Tenggara Province in 2024. The data was analyzed using supervised learning classification algorithms, namely *K-Nearest Neighbor (KNN)* and *Decision Tree*. The variables used include socioeconomic and health factors such as the percentage of *stunting*, the percentage of poor population, Average Length of Schooling (RLS), the percentage of *wasting* in children under five, and the percentage of adequate sanitation. The purpose of this study is to analyze the prevalence of *stunting* in children under five in NTT in 2024 based on socioeconomic and health factors, and to determine the visualization of the best classification algorithm based on accuracy value. The findings show that the *Decision Tree* algorithm provides the best results with an accuracy, sensitivity, and specificity value of 100%, outperforming *KNN* which only achieved 80% accuracy. Feature importance analysis from the *Decision Tree* model shows that the three main influential factors are average years of schooling (29.1%), percentage of *wasting* in children under five (28%), and percentage of poor population (25.4%). Based on the research findings, it is confirmed that accelerating the reduction of *stunting* in NTT needs to focus on three priority strategies: improving the quality of education, nutritional interventions for children under five (addressing *wasting*), and poverty alleviation.

Keywords: *Decision Tree*, East Nusa Tenggara, *KNN*, *Stunting*, Toddlers.

I. PENDAHULUAN

Salah satu masalah kesehatan masyarakat yang strategis di Indonesia adalah *stunting*. Menurut Survei Status Gizi Indonesia (SSGI) tahun 2022, prevalensi *stunting* nasional adalah 21,6%, lebih rendah dari target RPJMN 2020–2024 sebesar 14%. Pada tahun 2021, Provinsi NTT melaporkan prevalensi *stunting* nasional tertinggi yaitu 37,8% [1]. Sebagai kondisi, *stunting* disebabkan oleh kekurangan gizi kronis dan sering dipengaruhi faktor misalnya asupan nutrisi yang buruk, pola asuh yang buruk, dan infeksi berulang [2]. Meskipun pemerintah telah meluncurkan berbagai program percepatan penurunan prevalensi *stunting*, faktor sosial ekonomi dan kondisi kesehatan lingkungan masih menjadi kendala signifikan.

Data ambang batas prevalensi *stunting* tingkat nasional sebesar 19,8% [3]. Tingkat kemiskinan yang tinggi dan kurangnya literasi gizi di kalangan penduduk yang bermukim pada wilayah pedesaan dan terpencil di Provinsi NTT menyebabkan *stunting* yang tinggi. Kondisi tersebut menuntut adanya intervensi yang bersifat lintas sektor, terintegrasi, dan komprehensif untuk memastikan efektivitas penanganan permasalahan gizi kronis tersebut [4]. Selain itu, prevalensi *stunting* di Provinsi NTT juga dipengaruhi oleh akses masyarakat yang buruk terhadap layanan kesehatan dan sanitasi, yang secara signifikan menghambat pencegahan dan keparahan *stunting* [5]. Kondisi tersebut menegaskan urgensi strategis untuk menjadikan Provinsi NTT sebagai lokasi fokus penelitian, mengingat tingkat prevalensi yang melampaui ambang kritis nasional, adanya disparitas antarwilayah, dan kompleksitas tantangan struktural yang dihadapi.

Faktor sosial ekonomi dan kesehatan, terutama RLS, persentase penduduk miskin, dan persentase sanitasi layak, memengaruhi fenomena *stunting* di Provinsi NTT pada tahun 2024. Tingkat pendidikan memiliki peranan strategis dalam membentuk kapasitas pemahaman masyarakat, khususnya ibu, terhadap gizi dan pola asuh anak, sedangkan kemiskinan menjadi determinan struktural yang membatasi kemampuan rumah tangga untuk memperoleh pangan bergizi dan mengakses layanan kesehatan yang memadai [6]. Cakupan ASI eksklusif, tingginya prevalensi *wasting*, dan keterbatasan kesempatan kerja, meskipun memiliki pengaruh relatif lebih rendah, tetap relevan dalam menjelaskan variasi prevalensi *stunting* antarwilayah [7]. Pemilihan variabel-variabel tersebut didasarkan pada landasan empiris yang kuat serta relevansi kontekstual di Provinsi Nusa Tenggara Timur sehingga mampu merepresentasikan determinan utama *stunting* sekaligus memberikan pijakan ilmiah bagi perumusan intervensi kebijakan yang terarah, terukur, dan berbasis bukti.

Penelitian ini mengangkat permasalahan mengenai pemetaan dan klasifikasi wilayah dengan prevalensi *stunting* tinggi di Provinsi NTT pada tahun 2024, berdasarkan variabel faktor sosial ekonomi dan Kesehatan. Tingginya angka prevalensi *stunting* menandakan adanya kompleksitas penyebab yang belum sepenuhnya ditangani dengan pendekatan konvensional. Oleh karena itu, pendekatan berbasis teknologi data seperti algoritma *machine learning* dipandang perlu untuk meningkatkan efektivitas deteksi wilayah prioritas. Algoritma KNN mengklasifikasikan data uji berdasarkan dominasi kelas dari K tetangga terdekat, dengan perhitungan jarak yang umumnya menggunakan jarak Euclidean untuk menilai kesamaan antar data [8]. Sementara itu, *decision tree* adalah model pembelajaran terawasi dengan struktur hierarkis, di mana setiap simpul internal mewakili fitur yang dipilih berdasarkan nilai keuntungan tertinggi, cabang-cabang menunjukkan aturan keputusan, dan simpul daun menunjukkan hasil pengambilan keputusan [9]. Dalam hal ini yaitu KNN dan *Decision Tree* dipilih karena kemampuannya dalam melakukan klasifikasi berdasarkan pola dan relasi non-linear antar variabel [10].

Beberapa penelitian sebelumnya telah menunjukkan bahwa faktor-faktor seperti RLS, persentase sanitasi layak, persentase penduduk miskin, dan cakupan ASI eksklusif berkontribusi secara signifikan terhadap *stunting*. Penelitian yang dilakukan oleh Pratama & Irwandi (2021) menemukan bahwa cakupan ASI eksklusif merupakan faktor yang melindungi terhadap *stunting* [11]. Nadiem et al. (2024) meneliti hubungan antara air minum, sanitasi, kebersihan, dan berat badan kurang pada anak-anak di Punjab, Pakistan. Hasilnya menunjukkan bahwa anak-anak 1,8 kali lebih kecil kemungkinannya mengalami gizi kurang jika mereka memiliki akses ke sumber air minum, sanitasi, dan fasilitas kebersihan yang lebih baik [12]. Sementara Nurahadiyatika et al. (2022) menekankan pengaruh signifikan kemiskinan [13]. Riset yang dilakukan oleh Uskunia et al. (2024) menegaskan pendekatan yang digunakan mayoritas masih menggunakan analisis statistik klasik seperti regresi linier [14]. Studi

Putri et al. (2024) membandingkan kinerja algoritma *machine learning*, temuannya *Random Forest* dan KNN efektif dalam prediksi *stunting* [15]. Namun, belum banyak penelitian yang menggunakan algoritma *machine learning*, khususnya KNN dan *Decision Tree*, di tingkat provinsi seperti Provinsi Nusa Tenggara Timur. Dengan demikian, berdasarkan pengetahuan peneliti, hal tersebut menjadi dasar kebaruan dalam penelitian ini.

Berdasarkan uraian pendahuluan, tujuan riset ini untuk mengidentifikasi perubahan dalam prevalensi *stunting* balita di Provinsi NTT pada tahun 2024 berdasarkan faktor sosial ekonomi dan kesehatan. Penelitian juga akan menentukan hasil visualisasi algoritma terbaik dengan skor akurasi tertinggi. Salah satu keuntungan dari penelitian ini adalah bahwa itu akan membantu dalam pengembangan teknik klasifikasi spasial yang berbasis pembelajaran mesin untuk mengevaluasi prevalensi *stunting*, dengan fokus pada menggabungkan variabel sosial ekonomi dan kesehatan. Secara praktis, riset ini dapat digunakan oleh pembuat kebijakan di pemerintah Provinsi NTT untuk membuat kebijakan intervensi berbasis bukti (*evidence-based*), memberi tenaga kesehatan program edukasi gizi, dan meningkatkan pemahaman masyarakat tentang risiko *stunting* di Provinsi NTT. Selain itu, penelitian ini dapat berfungsi sebagai referensi untuk membangun sistem peringatan dini untuk mengendalikan *stunting* di daerah dengan prevalensi tinggi.

II. METODE PENELITIAN

Dalam penelitian ini, pendekatan kuantitatif digunakan untuk mengidentifikasi dan mengkategorikan tingkat *stunting* di Provinsi NTT pada tahun 2024 berdasarkan faktor sosial ekonomi dan kesehatan. Data tahun 2024 yang bersumber dari BPS Provinsi NTT. Tabel I berikut menunjukkan variabel yang dipilih yaitu prevalensi *stunting*, persentase penduduk miskin, RLS, Tingkat Pengangguran Terbuka (TPT), cakupan ASI eksklusif, persentase balita *wasting*, dan persentase sanitasi layak.

Tabel I

Data Sosial Ekonomi dan Kesehatan Provinsi Nusa Tenggara Timur Tahun 2024

Kabupaten /Kota	Prevalensi <i>Stunting</i>	Kategori <i>Stunting</i>	Persentase Penduduk Miskin	RLS	TPT	Cakupan ASI Eksklusif	Persentase Balita <i>Wasting</i>	Persentase Sanitasi Layak
Sumba Barat	23.60	Tinggi	26.52	7.17	3.21	100.00	9.20	60.76
Sumba Timur	15.10	Rendah	27.04	7.77	3.46	97.84	5.50	65.88
Kupang	14.00	Rendah	21.37	7.62	3.36	100.00	9.50	73.34
Timor Tengah Selatan	24.20	Tinggi	24.68	7.21	2.36	96.28	11.60	70.72
Timor Tengah Utara	25.90	Tinggi	20.89	8.17	1.82	93.35	8.70	80.84
Belu	13.90	Rendah	13.86	7.50	5.41	86.17	10.70	85.78
Alor	13.10	Rendah	19.87	8.61	2.27	95.41	11.10	85.83
Lembata	7.90	Rendah	24.22	8.27	2.18	95.79	7.50	88.61
Flores Timur	20.00	Tinggi	11.25	8.06	3.58	100.00	10.20	96.98
Sikka	19.30	Rendah	11.89	6.99	2.33	92.73	15.30	85.35
Ende	10.50	Rendah	22.57	8.29	2.05	95.84	9.40	92.69
Ngada	9.80	Rendah	11.87	8.83	2.68	98.80	6.00	86.56
Manggarai	11.50	Rendah	19.01	7.91	1.17	97.34	3.70	66.26
Rote Ndao	18.40	Rendah	25.78	7.83	2.42	98.24	9.90	85.16
Manggarai Barat	11.80	Rendah	16.74	8.21	3.47	96.93	6.10	82.03

Kabupaten /Kota	Prevalensi <i>Stunting</i>	Kategori <i>Stunting</i>	Persentase Penduduk Miskin	RLS	TPT	Cakupan ASI Eksklusif	Persentase Balita <i>Wasting</i>	Persentase Sanitasi Layak
Sumba Tengah	15.30	Rendah	30.84	7.23	1.89	96.59	4.30	52.51
Sumba Barat Daya	39.20	Tinggi	27.20	6.39	2.64	94.49	9.00	52.60
Nagekeo	9.60	Rendah	12.30	8.23	2.17	100.00	8.20	83.52
Manggarai Timur	8.30	Rendah	24.90	7.96	0.51	100.00	4.00	55.75
Sabu Raijua	20.80	Tinggi	28.13	6.99	3.99	89.29	9.80	83.51
Malaka	15.40	Rendah	13.92	7.14	1.58	91.85	12.30	73.70
Kota Kupang	18.80	Rendah	8.24	11.64	8.60	96.50	11.40	91.69

Sumber: BPS Provinsi NTT

Berdasarkan ambang batas nasional prevalensi *stunting* sebesar 19,8% menurut Survei Status Gizi Indonesia (SSGI) 2024, setiap kabupaten/kota diklasifikasikan ke dalam dua kategori, yaitu *stunting* tinggi apabila prevalensi lebih dari 19,8% dan *stunting* rendah apabila prevalensi kurang dari atau sama dengan 19,8%.

A. Pengumpulan dan Praproses Data

Data penelitian diperoleh dari BPS Provinsi NTT tahun 2024, mencakup 22 kabupaten/kota dengan tujuh variabel yaitu persentase penduduk miskin, RLS, TPT, cakupan ASI eksklusif, persentase balita *wasting*, persentase sanitasi layak, dan prevalensi *stunting* sebagai variabel target. Dari keseluruhan data, 80% atau 17 data dipakai untuk data latih dan 20% atau 5 data sebagai data uji, dengan pembagian acak digunakan metode *train-test split*. Sebelum analisis, data dibersihkan dan distandarisasi menggunakan *StandardScaler* untuk menyeragamkan skala antarvariabel numerik [16].

B. Metode Klasifikasi

K-Nearest Neighbor (KNN)

KNN menentukan kelas mayoritas K tetangga terdekat dengan mengukur jarak antara data uji dan data latih. Euklidis Jarak digunakan sebagai ukuran jarak [17]. Rumus *Euclidean Distance* ditulis sebagai Persamaan (1) berikut.

$$d(\rho, q) = \sqrt{\sum_{i=1}^n (\rho_i - q_i)^2} \quad (1)$$

Keterangan:

$d(\rho, q)$ = jarak euklidis yang ada antara dua titik data, ρ dan q

n = jumlah atribut

ρ_i, q_i = nilai fitur ke- i untuk titik ρ dan q

Decision Tree

Algoritma *Decision Tree* membagi data berdasarkan atribut yang memberikan pemisahan terbaik, diukur dengan *Entropy* dan *Information Gain* [18]. Rumus *Entropy* ditulis sebagai Persamaan (2) berikut.

$$Entropy(S) = - \sum_{i=1}^c \rho_i \log_2(\rho_i) \quad (2)$$

Keterangan:

S = himpunan data

c = jumlah kelas

ρ = proporsi data untuk kelas ke- i

Rumus *Information Gain* dihitung dengan Persamaan (3) berikut.

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} |S_v| |S| Entropy(S_v) \quad (3)$$

Keterangan:

A = atribut yang dievaluasi

v = nilai-nilai unik dari atribut A

S_v = subset data dengan nilai atribut $A = v$
 $|S_v|/|S|$ = proporsi subset terhadap keseluruhan data

C. *Pengujian dan Evaluasi*

20 persen data uji dan 80 persen data latih terdiri dari dua bagian dataset. Untuk mengevaluasi kinerja model, metrik klasifikasi seperti akurasi, presisi, sensitivitas (*recall*), dan spesifisitas digunakan, yang didasarkan pada matriks *confusion*. [19]. Rumus-rumus metrik evaluasi ditulis sebagai Persamaan (8), Persamaan (9), Persamaan (10), dan Persamaan (11) berikut.

1. Akurasi adalah proporsi jumlah peramalan yang benar dari seluruh data yang diuji.

$$Akurasi = \frac{Prediksi\ Benar}{Total\ Prediksi} \tag{8}$$

2. Presisi adalah persentase peramalan positif yang benar dari semua prediksi positif yang dihasilkan oleh model.

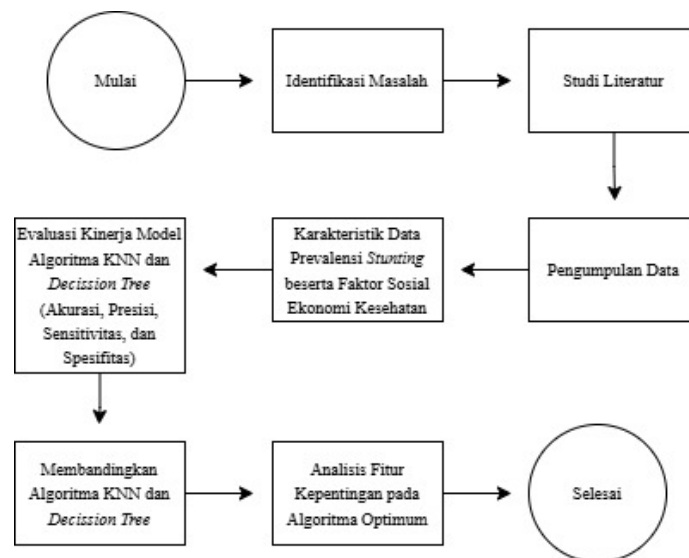
$$Presisi(P) = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP)+False\ Positive\ (FP)} \tag{9}$$

3. Sensitivitas (*recall* positif) adalah proporsi data nyata yang diidentifikasi secara akurat sebagai positif oleh model.

$$Sensitivitas = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP)+False\ Negative\ (FN)} \tag{10}$$

4. Spesifisitas (*recall* negatif) adalah proporsi dari seluruh data yang sebenarnya negatif yang diklasifikasikan dengan benar oleh model.

$$Spesifisitas = \frac{True\ Negative\ (TN)}{True\ Negative\ (TN)+False\ Positive\ (FP)} \tag{11}$$

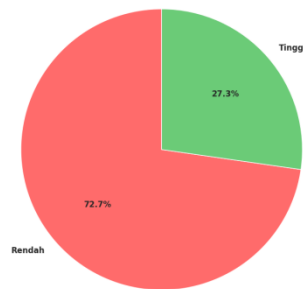


Gambar 1. Flowchart Penelitian

III. HASIL DAN PEMBAHASAN

A. Karakteristik Data

Tahap awal analisis dilakukan dengan mengelompokkan prevalensi *stunting* di Provinsi NTT tahun 2024 berdasarkan ambang batas nasional sebesar 19,8%. Proses pengelompokan ini memberikan gambaran awal mengenai kondisi kerentanan *stunting* antarwilayah, sekaligus menjadi dasar untuk memahami pola distribusi yang ada sebelum dilakukan analisis lebih lanjut menggunakan algoritma klasifikasi. Hasil pengelompokan menunjukkan adanya perbedaan yang cukup nyata, di mana sebagian besar kabupaten/kota berada pada kategori *stunting* rendah, sementara sebagian lainnya masih tergolong *stunting* tinggi sebagaimana disajikan pada Gambar 2 berikut.



Gambar 2. Distribusi Kategori *Stunting*

Berdasarkan Gambar 1, distribusi prevalensi *stunting* di Provinsi NTT tahun 2024 menunjukkan bahwa dari total 22 kabupaten/kota, sebanyak 16 wilayah atau sekitar 72,7 persen termasuk dalam kategori *stunting* rendah, sedangkan 6 wilayah atau 27,3 persen tergolong *stunting* tinggi. Ketimpangan distribusi ini menggambarkan kondisi *class imbalance* yang dapat memengaruhi kinerja model klasifikasi karena algoritma cenderung lebih akurat mengenali kelas dengan jumlah data yang lebih dominan dibandingkan kelas yang jumlahnya lebih sedikit.

Tabel II

Descriptive Statistics

	<i>Mean</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Std. Deviation</i>
X_1	20.14	8.24	30.84	6.64
X_2	7.91	6.39	11.64	1.03
X_3	2.87	0.51	8.6	1.65
X_4	96.07	86.17	100	3.64
X_5	8.79	3.7	15.3	2.93
X_6	77.28	52.51	96.98	13.3

Berdasarkan Tabel II menyajikan hasil analisis deskriptif, rata-rata persentase penduduk miskin (X_1) di NTT mencapai 20,14% dengan variasi antara 8,24% hingga 30,84% yang menunjukkan adanya kesenjangan signifikan antarwilayah. rata-rata lama sekolah (X_2) tercatat sebesar 7,91 tahun, dengan rentang 6,39 hingga 11,64 tahun, yang mengindikasikan perbedaan akses pendidikan di antara kabupaten/kota. tingkat pengangguran terbuka (X_3) relatif rendah, dengan rata-rata 2,87% dan nilai maksimum 8,60%, meskipun di beberapa wilayah angkanya masih cukup tinggi untuk ukuran pedesaan. Cakupan ASI eksklusif (X_4) memiliki nilai rata-rata tinggi yaitu 96,07% dengan variasi 86,17% hingga 100%, meskipun masih terdapat daerah dengan capaian di bawah standar ideal. Prevalensi balita *wasting* (X_5) mencapai rata-rata 8,79% dan maksimum 15,3%, yang menandakan adanya masalah gizi akut di sejumlah wilayah. Sementara itu, persentase sanitasi layak (X_6) rata-rata mencapai 77,28%, tetapi masih ada wilayah dengan cakupan rendah sekitar 52,51%.

B. *Evaluasi Kinerja Model Algoritma*

Evaluasi kinerja algoritma KNN dan *Decision Tree* dilakukan untuk menilai tingkat akurasi, presisi, sensitivitas, serta spesifisitas dalam mengklasifikasikan prevalensi *stunting*. Tahapan ini bertujuan untuk mengetahui sejauh mana kedua metode mampu memberikan hasil klasifikasi yang tepat untuk menentukan algoritma yang paling cocok untuk penelitian ini.

K- Nearest Neighbors (KNN)

Tabel III

Skor Akurasi

k	Skor Akurasi
K=1	40%
K=2	80%
K=3	80%
K=4	80%
K=5	60%

Berdasarkan Tabel III menyajikan pengujian algoritma KNN dilakukan dengan menggunakan variasi nilai K mulai dari 1 hingga 5. Hasil akurasi yang diperoleh memperlihatkan perbedaan antar nilai K, di mana K=2, K=3, dan K=4 menghasilkan performa terbaik dengan tingkat akurasi masing-masing sebesar 80%. Sementara itu, K=1 hanya mencapai akurasi 40% dan K=5 turun menjadi 60%. Perbedaan ini menunjukkan bahwa pemilihan nilai K sangat memengaruhi kemampuan model dalam melakukan klasifikasi.

Confusion Matrix - KNN

Aktual	Rendah	4	0
	Tinggi	1	0
		Rendah	Tinggi
		Prediksi	

Gambar 3. Confusion Matrix K- Nearest Neighbors (KNN)

Evaluasi matriks *confusion* pada algoritma KNN sebagaimana disajikan pada Gambar 3 sebagai berikut:

Akurasi : 80%
 Presisi : 0%
 Sensitivitas : 0%
 Spesifisitas : 100%

Berdasarkan hasil evaluasi matrik *confusion* yang disajikan pada Gambar 3 algoritma KNN memperoleh akurasi sebesar 80%, yang berarti dari lima kabupaten/kota yang dianalisis, empat wilayah dengan prevalensi *stunting* rendah berhasil diprediksi dengan benar. Namun, tidak ada satupun wilayah dengan prevalensi *stunting* tinggi yang dapat dikenali, karena seluruh kasus *stunting* tinggi justru diklasifikasikan sebagai *stunting* rendah. Kondisi ini tercermin pada nilai presisi dan sensitivitas sebesar 0%, yang menunjukkan ketidakmampuan model dalam mendeteksi kelas minoritas. Sebaliknya, nilai spesifisitas mencapai 100%, menandakan bahwa seluruh wilayah dengan prevalensi *stunting* rendah dapat diprediksi secara tepat tanpa menghasilkan kesalahan klasifikasi. Dengan demikian, meskipun akurasi model terlihat cukup tinggi, kelemahan mendasar KNN terletak pada kegagalannya mengidentifikasi wilayah *stunting* tinggi akibat ketidakseimbangan data (*class imbalance*), sehingga performanya kurang optimal untuk mendukung penentuan wilayah prioritas penanganan *stunting*.

Decision Tree

Confusion Matrix - Decision Tree

Aktual	Rendah	4	0
	Tinggi	0	1
		Rendah	Tinggi
		Prediksi	

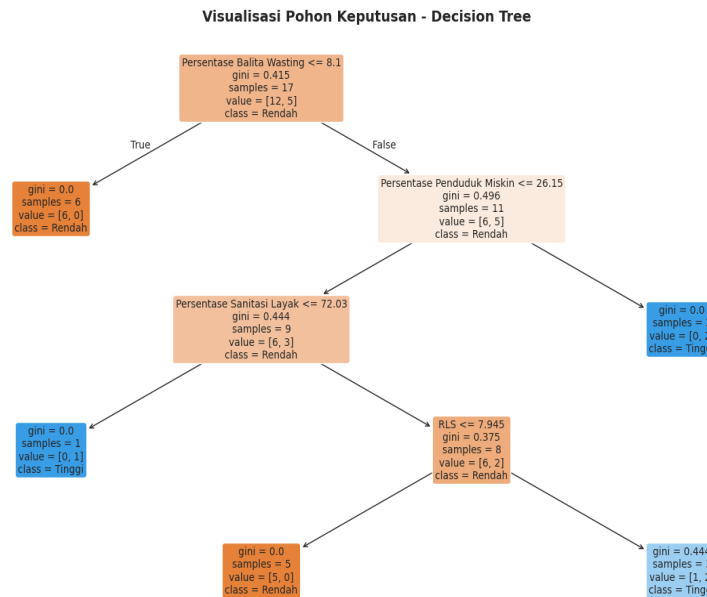
Gambar 4. Confusion Matrix Decision Tree

Evaluasi *confusion matrix* pada algoritma *Decision Tree* sebagaimana disajikan pada Gambar 4 sebagai berikut:

Akurasi : 100%
 Presisi : 100%
 Sensitivitas : 100%

Spesifisitas : 100%

Berdasarkan *confusion matrix* yang disajikan pada Gambar 3 algoritma *Decision Tree* memperoleh akurasi sebesar 100%, yang berarti seluruh kabupaten/kota pada data uji berhasil diprediksi sesuai dengan kondisi sebenarnya. Semua wilayah dengan prevalensi *stunting* rendah terklasifikasi dengan tepat sebagai *stunting* rendah, dan seluruh wilayah dengan prevalensi *stunting* tinggi juga dikenali dengan benar sebagai *stunting* tinggi. Kondisi ini tercermin dari nilai presisi, sensitivitas, dan spesifisitas yang sama-sama mencapai 100%, menandakan bahwa model tidak menghasilkan kesalahan prediksi baik pada kelas mayoritas maupun minoritas. Oleh karena itu, algoritma *Decision Tree* sangat baik dalam mengidentifikasi status *stunting* dan menyelesaikan masalah ketidakseimbangan data, atau ketidakseimbangan kelas, yang merupakan kelemahan utama KNN. Gambar 4 menunjukkan visualisasi algoritma *Decision Tree*.



Gambar 5. Visualisasi *Decision Tree*

Struktur pohon keputusan sebagaimana disajikan pada Gambar 5, diperoleh sebagai berikut.

1. Persentase balita *wasting* $\leq 8,1$; maka *stunting* rendah
2. Persentase balita *wasting* $> 8,1$, persentase penduduk miskin $\leq 26,15$, persentase sanitasi layak $\leq 72,03$; maka *stunting* tinggi
3. Persentase balita *wasting* $> 8,1$, persentase penduduk miskin $\leq 26,15$, persentase sanitasi layak $> 72,03$, RLS $\leq 7,945$; maka *stunting* rendah
4. Persentase balita *wasting* $\leq 8,1$, persentase penduduk miskin $\leq 26,15$, persentase sanitasi layak $> 72,03$, RLS $> 7,945$; maka *stunting* tinggi
5. Persentase balita *wasting* $> 8,1$, persentase penduduk miskin $> 26,15$; maka *stunting* tinggi

Berdasarkan hasil visualisasi *Decision Tree* yang disajikan pada Gambar 4 diketahui bahwa variabel persentase balita *wasting* menjadi titik awal pemisahan yang menentukan klasifikasi status *stunting*. Jalur yang paling relevan ditunjukkan pada jalur kelima, yaitu ketika persentase balita *wasting* $> 8,1\%$ dan persentase penduduk miskin $> 26,15\%$, maka kabupaten/kota cenderung dikategorikan sebagai *stunting* tinggi. Aturan ini konsisten dengan kondisi empiris di Indonesia, di mana prevalensi *wasting* yang melampaui ambang batas nasional menjadi indikator kuat permasalahan gizi akut, sementara faktor kemiskinan memperkuat kerentanan wilayah terhadap *stunting*. Temuan ini juga menunjukkan bahwa kemiskinan tidak selalu menjadi determinan tunggal, sebab pada beberapa kasus tingginya angka *wasting* sudah cukup untuk mengklasifikasikan wilayah ke dalam kategori *stunting* tinggi meskipun tingkat kemiskinan relatif lebih rendah. Dengan demikian, jalur kelima menegaskan bahwa masalah gizi akut, khususnya *wasting*, memiliki kontribusi besar terhadap peningkatan prevalensi *stunting*.

C. Perbandingan Algoritma K-Nearest Neighbors (KNN) dan Decision Tree

Perbandingan kinerja algoritma KNN dan *Decision Tree* dilakukan untuk memperoleh gambaran menyeluruh mengenai efektivitas kedua metode dalam mengklasifikasikan prevalensi *stunting*.

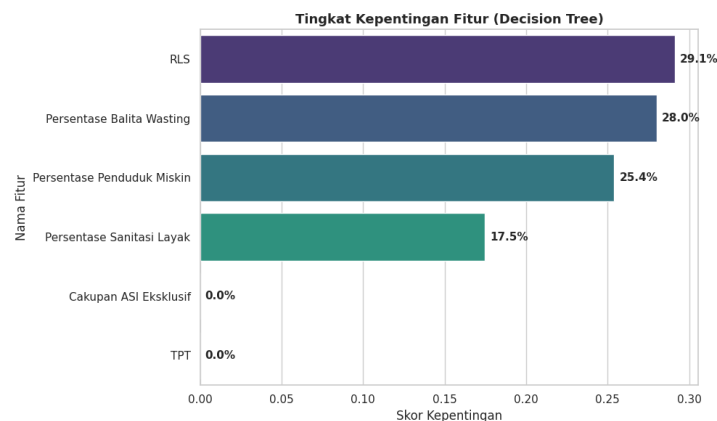
Tahapan ini penting untuk menentukan model mana yang paling optimal digunakan, khususnya ditinjau dari capaian akurasi dan indikator evaluasi lainnya. Tabel IV berikut menyajikan ringkasan hasil pengujian kedua algoritma tersebut.

Tabel IV
Hasil Akurasi KNN dan *Decision Tree*

Metode	Akurasi (%)
KNN	80%
<i>Decision Tree</i>	100%

Berdasarkan hasil pengujian model yang disajikan pada Tabel IV diketahui bahwa algoritma KNN menghasilkan akurasi sebesar 80%, sedangkan algoritma *Decision Tree* mencapai akurasi 100%. Perbedaan ini menunjukkan bahwa *Decision Tree* lebih mampu menyesuaikan pola dari data yang digunakan, meskipun nilai akurasi yang sempurna juga mengindikasikan potensi *overfitting* sehingga perlu dilakukan evaluasi lebih lanjut. Sementara itu, akurasi KNN yang lebih rendah memperlihatkan adanya keterbatasan algoritma ini dalam menangani data dengan jumlah sampel yang relatif kecil serta ketidakseimbangan kelas

D. Analisis Kepentingan Fitur



Gambar 6. Tingkat Kepentingan Fitur *Decision Tree*

Berdasarkan Gambar 6 yang menyajikan hasil analisis tingkat kepentingan fitur berdasarkan algoritma *Decision Tree*. Variabel yang memiliki kontribusi terbesar dalam membedakan kategori *stunting* adalah rata-rata lama sekolah dengan skor kepentingan sebesar 29,1%, diikuti oleh persentase balita *wasting* sebesar 28%, serta persentase penduduk miskin sebesar 25,4%. Selanjutnya, persentase sanitasi layak berkontribusi sebesar 17,5%, meskipun nilainya relatif lebih rendah dibandingkan variabel lain. Sementara itu, cakupan ASI eksklusif dan tingkat pengangguran terbuka tidak memberikan kontribusi terhadap klasifikasi karena nilai kepentingannya 0%. Dalam konteks algoritma *Decision Tree*, nilai kepentingan ini merepresentasikan sejauh mana suatu variabel berperan dalam meningkatkan akurasi pemisahan kelas pada model. Hasil tersebut mengindikasikan bahwa aspek pendidikan, status gizi, dan kondisi ekonomi lebih dominan dalam menjelaskan perbedaan prevalensi *stunting* antarwilayah dibandingkan indikator kesehatan lain seperti cakupan ASI eksklusif atau tingkat pengangguran terbuka. Dengan demikian variabel RLS, persentase balita *wasting*, dan persentase penduduk miskin menjadi determinan utama dalam klasifikasi status *stunting* di Provinsi NTT.

E. Pembahasan

Hasil analisis deskriptif menunjukkan bahwa prevalensi *stunting* di Provinsi NTT tahun 2024 masih menjadi persoalan serius, di mana 27,3% kabupaten/kota tergolong dalam kategori *stunting* tinggi. Kondisi ini memperlihatkan adanya ketimpangan antarwilayah, meskipun sebagian besar daerah telah berada di bawah ambang batas nasional sebesar 19,8%. Tingginya angka *stunting* dipengaruhi oleh faktor sosial, ekonomi, dan kesehatan yang saling berkaitan. Rata-rata persentase penduduk miskin mencapai 20,14% dengan variasi cukup besar antarwilayah, yang menegaskan adanya disparitas

kesejahteraan rumah tangga. Faktor pendidikan yang direpresentasikan oleh rata-rata lama sekolah sebesar 7,91 tahun turut berkontribusi terhadap risiko *stunting*, karena tingkat pendidikan yang rendah berkorelasi dengan keterbatasan pengetahuan gizi dan praktik pola asuh. Sementara itu, prevalensi balita *wasting* dengan rata-rata 8,79% mencerminkan adanya masalah gizi akut yang secara langsung memengaruhi pertumbuhan anak.

Berdasarkan hasil visualisasi pada algoritma terbaik yaitu *Decision Tree* berdasarkan skor akurasi terbaik yaitu 100% mengindikasikan bahwa jalur utama klasifikasi dipengaruhi oleh persentase balita *wasting* lebih dari 8,1% dan persentase penduduk miskin lebih dari 26,15%, yang secara konsisten mengarahkan wilayah pada kategori *stunting* tinggi. Temuan ini selaras dengan hasil analisis kepentingan fitur, di mana rata-rata lama sekolah sebesar 29,1%, persentase balita *wasting* sebesar 28%, dan persentase penduduk miskin sebesar 25,4% muncul sebagai variabel dominan. Sementara itu, persentase sanitasi layak hanya menyumbang 17,5%, sedangkan cakupan ASI eksklusif dan tingkat pengangguran terbuka tidak memberikan kontribusi berarti. Dengan demikian, hasil penelitian ini menegaskan bahwa pendidikan, gizi, dan kondisi ekonomi merupakan faktor kunci yang membedakan prevalensi *stunting* antarwilayah di Provinsi NTT.

IV. KESIMPULAN

Temuan penelitian menunjukkan bahwa 27,3% kota dan kabupaten di NTT mengalami *stunting* tinggi pada 2024, yang dipengaruhi oleh faktor sosial-ekonomi dan kesehatan seperti kemiskinan, rendahnya pendidikan, dan tingginya prevalensi *wasting*. Algoritma *Decision Tree* memberikan kinerja terbaik dengan akurasi sempurna, menegaskan bahwa RLS, prevalensi *wasting*, dan persentase penduduk miskin menjadi variabel utama pembeda wilayah *stunting* tinggi dan rendah. Oleh karena itu, fokus penurunan *stunting* perlu diarahkan pada peningkatan pendidikan, intervensi gizi, dan pengentasan kemiskinan.

Meski berhasil mengidentifikasi faktor utama, penelitian ini masih membuka peluang pengembangan dengan penggunaan data panel multi-tahun untuk menganalisis tren dan kausalitas secara lebih mendalam. Penggunaan metode validasi yang lebih kuat seperti *k-fold cross-validation* dianjurkan untuk menghindari *overfitting* akibat dataset kecil. Selain itu, analisis data mikro tingkat rumah tangga dapat memberikan wawasan lebih detail tentang determinan *stunting*.

DAFTAR PUSTAKA

- [1] S. F. Nashriyah, M. R. Makhful, and Y. P. Devi, "Gambaran Spasial Hubungan Antara Faktor Lingkungan Dan Ekonomi Dengan Stunting Balita Di Provinsi Nusa Tenggara Timur," *J. Spat. Wahana Komun. dan Inf. Geogr.*, vol. 23, no. 2, pp. 1–8, Mar. 2023, doi: 10.21009/spatial.232.01.
- [2] R. Pratiwi, S. R., and R. F., "Dampak Status Gizi Pendek (Stunting) Terhadap Prestasi Belajar," *J. Ilm. Ilmu Keperawatan*, vol. 12, no. 2, pp. 11–23, 2021, doi: 10.36089/nu.v12i2.317.
- [3] BPS NTT, "Jumlah dan Persentase Balita Stunting Menurut Kabupaten/Kota Tahun 2024," Badan Pusat Statistik Provinsi Nusa Tenggara Timur. [Online]. Available: <https://ntt.bps.go.id/id/statistics-table/2/MTQ4OSMy/jumlah-balita-stunting-menurut-kabupaten-kota.html>
- [4] A. L. Kobun and A. M. Susilo, "Stunting: Poverty and Nutrition Status (Case Study at Tetaf Health Center, Timor Tengah Selatan Regency (TTS) East Nusa Tenggara Province (NTT)," *Int. J. Econ. Bus. Manag. Res.*, vol. 6, pp. 1–11, 2022.
- [5] M. A. L. Suratri *et al.*, "Risk Factors for Stunting among Children under Five Years in the Province of East Nusa Tenggara (NTT), Indonesia," *Int. J. Environ. Res. Public Health*, vol. 20, no. 2, pp. 1–13, 2023, doi: 10.3390/ijerph20021640.
- [6] U. Agustine, S. D. R. P. Santoso, G. F. Boa, S. Mugiarti, and Y. E. S. Gunawan, "Determinant of Stunting among Toddler in East Nusa Tenggara, Indonesia," *Babali Nurs. Res.*, vol. 3, no. 3, pp. 232–245, 2022, doi: 10.37363/bnr.2022.33148.
- [7] S. F. Nashriyah, M. R. Makhful, and Y. P. Devi, "Gambaran Spasial Hubungan Antara Faktor Lingkungan dan Ekonomi dengan Stunting Balita di Provinsi Nusa Tenggara Timur," *J. Spat. Wahana Komun. dan Inf. Geogr.*, vol. 23, no. 2, pp. 1–8, 2023.
- [8] Z. Sultana, A. Ferdousi, F. Tasnim, and L. Nahar, "An Improved K-Nearest Neighbor Algorithm for Pattern Classification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 8, pp. 760–767, 2022, doi: 10.14569/IJACSA.2022.0130887.
- [9] B. Luna-Benoso, J. C. Martínez-Perales, U. S. Morales-Rodríguez, R. Flores-Carapia, and V. M. Silva-García, "A New Classification Model Using a Decision Tree Generated from Hyperplanes in Dimensional Space," *Appl. Artif. Intell.*, vol. 38, no. 1, Dec. 2024, doi: 10.1080/08839514.2024.2426377.
- [10] S. Lonang, A. Yudhana, and M. K. Biddinika, "Analisis Komparatif Kinerja Algoritma Machine Learning untuk Deteksi Stunting," *J. Media Inform. Budidarma*, vol. 7, no. 4, p. 2109, 2023.
- [11] M. R. Pratama and S. Irwandi, "Hubungan Pemberian ASI Eksklusif dengan Stunting di Puskesmas Hinai Kiri, Kecamatan Secanggang, Kabupaten Langkat," *J. Kedokt. STM (Sains dan Teknol. Med.)*, vol. 4, no. 1, pp. 17–25, 2021.
- [12] M. Nadeem, M. Anwar, S. Adil, W. Syed, M. B. A. Al-Rawi, and A. Iqbal, "The Association between Water, Sanitation, Hygiene, and Child Underweight in Punjab, Pakistan: An Application of Population Attributable Fraction," *Journal of Multidisciplinary Healthcare.*, vol. 17, pp. 2475-2487, 2024, doi: 10.2147/JMDH.S461986
- [13] F. Nurahdiyatika, D. R. Atmaka, and A. I. Imani, "Peningkatan Ketahanan Pangan dan Pengentasan Status Kemiskinan dalam Konvergensi Penurunan Angka Stunting," *Media Gizi Indones.*, vol. 17, pp. 215–220, 2022.

- [14] M. Uskunia, N. Tameno, and N. T. Kiak, "Pengaruh Sosial Ekonomi Keluarga Dan Strategi Penanggulangan Stunting di Puskesmas Nurobo Kabupaten Malaka," *Monet. J. Ekon. dan Keuang.*, vol. 2, no. 2, pp. 254–264, 2024.
- [15] I. P. Putri, T. Terttiaavini, and N. Arminarahmah, "Analisis Perbandingan Algoritma Machine Learning untuk Prediksi Stunting pada Anak," *Malcom Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, pp. 257–265, 2024.
- [16] A. Prasatya and N. Hendrastuty, "Analisis Sentimen: Perbandingan Performa Algoritma Naive Bayes, Support Vector Machine, Random Forest, dan K-Nearest Neighbor Dalam Pemecatan Shin Tae Yong pada Media X," *Build. Informatics, Technol. Sci.*, vol. 6, no. 4, p. 2612–2623, 2025.
- [17] P. Cunningham and S. J. Delany, "K-Nearest Neighbour Classifiers - A Tutorial," *ACM Comput Surv.*, vol. 54, no. 6, pp. 1–25, 2022.
- [18] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *J. Appl. Sci. Technol. Trends*, vol. 2, no. 1, pp. 20–28, 2021.
- [19] A. K. Iman and E. I. H. Ujjianto, "Analisis Sentimen Pemindahan Ibu Kota Indonesia Menggunakan K-Nearest Neighbor," *J. Pendidik. dan Teknol. Indones.*, vol. 4, no. 12, pp. 759–768, 2025.