p-ISSN : 2406-7733
e-ISSN : 2597-9922

PENERAPAN METODE *FUZZY K-NEAREST NEIGHBOR* DALAM KLASIFIKASI RISIKO STUNTING PADA ANAK BERDASARKAN FAKTOR ANTROPOMETRI

Nur Choerun Nisa[1], Deny Nugroho Triwibowo[2], Khoirun Nisa[3]
Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Harapan Bangsa
Jl. Raden Patah no.100. Ledug, Kembaran, Purwokerto, Banyumas
E-mail: nurchoerunnisa124@gmail.com[1], denynugroho@uhb.ac.id[2], khoirunnisa@uhb.ac.id[3]

Abstrak - Stunting merupakan kondisi gagal tumbuh pada anak balita akibat kekurangan gizi kronis yang berdampak pada perkembangan fisik dan kognitif anak. Metode klasifikasi konvensional menghasilkan keputusan tegas (*crisp*) yang kurang mampu merepresentasikan ketidakpastian pada data pertumbuhan anak yang tumpang tindih antar kategori status gizi. Penelitian ini bertujuan menerapkan metode *Fuzzy K-Nearest Neighbor* (FKNN) yang mengintegrasikan derajat keanggotaan *fuzzy* ke dalam mekanisme *K-Nearest Neighbor* untuk klasifikasi status gizi balita, serta membandingkan performanya dengan KNN konvensional dan *Random Forest*. *Dataset* yang digunakan adalah dataset Stunting *Toddler* (Balita) *Detection* dari Kaggle, diambil sampel 2.450 data melalui *stratified sampling* proporsional, menyisakan 2.375 data unik dengan tiga fitur (Usia, Jenis Kelamin, Tinggi Badan) dan empat kelas Status Gizi. Tahapan penelitian meliputi praproses data, pembangunan model, optimasi nilai k menggunakan *stratified* 10-*Fold Cross Validation*, serta evaluasi menggunakan *confusion matrix* dengan metrik *precision, recall, F1-Score*, dan akurasi *macro-average*. Hasil pengujian pada 475 data uji menunjukkan FKNN (k=3) mencapai akurasi 91,16%, mengungguli KNN konvensional (89,89%) dan menyamai *Random Forest*. FKNN mencapai *recall* tertinggi pada kelas Stunting (83,33%), menunjukkan kemampuan lebih baik dalam mendeteksi kasus stunting sebenarnya. Nilai keanggotaan *fuzzy* terbukti dapat digunakan sebagai indikator keandalan prediksi pada data *borderline*.

Kata Kunci: stunting, *Fuzzy K-Nearest Neighbor*, klasifikasi, derajat keanggotaan, *Random Forest*

I. PENDAHULUAN

Stunting merupakan permasalahan gizi kronis yang masih menjadi tantangan serius dalam kesehatan masyarakat global, kondisi ini didefinisikan sebagai keadaan anak yang terlalu pendek untuk usianya, yaitu tinggi badan lebih dari dua deviasi standar di bawah median standar pertumbuhan anak World Health Organization (WHO). Stunting dipengaruhi oleh kesehatan dan gizi ibu selama kehamilan, faktor asupan nutrisi, infeksi, dan sanitasi setelah lahir dan faktor sosial ekonomi (Onis et al., 2025). Data *Joint Child Malnutrition Estimates* 2025 mencatat sekitar 150,2 juta anak di bawah lima tahun mengalami stunting secara global pada 2024 (World Health Organization, 2025). Dampak stunting tidak hanya gangguan pertumbuhan fisik namun meluas pada perkembangan kognitif, daya tahan tubuh, dan produktivitas individu dewasa (Imeldawati, 2025).

Penentuan status stunting melibatkan indikator antropometri yang umumnya diplotkan ke kurva pertumbuhan WHO secara manual (Lusiana & Pratiwi, 2023). Meskipun teorinya deterministik, penerapannya sering kali mengalami keterbatasan dalam pengolahan data yang bersifat kompleks (Ratnasari et al., 2024). Metode klasifikasi konvensional seperti *K-Nearest Neighbor* (KNN) menghadapi tantangan akibat kondisi *overlapping* dan ketidakseimbangan kelas, menghasilkan keputusan klasifikasi yang tegas dan seringkali sewenang-wenang pada data *borderline* (Ernawan et al., 2024). Observasi awal terhadap 2.450 sampel data menggunakan KNN memperlihatkan kelemahan berupa anak dengan tinggi badan mendekati ambang −2 SD hingga −3 SD berpotensi salah diklasifikasikan karena *majority voting* KNN konvensional sulit menangkap ambiguitas data. Penurunan akurasi KNN pada berbagai nilai k turut dilaporkan penelitian terdahulu, berbeda dengan metode bermekanisme tambahan untuk penanganan ketidakpastian data seperti *Modified* KNN yang mencapai akurasi 99,36% (D. I. Pratama et al., 2025), sementara kesalahan klasifikasi kemiripan data di area batas kelas turut ditemukan pada studi lain (Pebrianti et al., 2024).

Pendekatan logika *fuzzy* menawarkan alternatif melalui derajat keanggotaan pada setiap kelas (misalnya 0,52 *stunted* dan 0,48 normal), data di sekitar ambang batas −2 SD dikelompokkan secara bertahap alih-alih ditetapkan biner ke salah satu kelas (Munthe et al., 2024). Sebagian besar penelitian klasifikasi stunting masih bertumpu pada KNN dan perbandingannya dengan algoritma seperti *Naïve Bayes* atau *Support Vector Machine*, namun belum melibatkan pendekatan *fuzzy* (Dwinanto et al., 2024). Padahal karakteristik data stunting dengan batas antarkelas yang sempit berpotensi menyebabkan misklasifikasi. Hal ini menunjukkan kesenjangan dalam penerapan *Fuzzy K-Nearest Neighbor* (FKNN) pada klasifikasi stunting.

p-ISSN : 2406-7733
e-ISSN : 2597-9922

Berdasarkan karakteristik data tersebut, penelitian ini menerapkan FKNN untuk klasifikasi multi-kelas status gizi berdasarkan kategori *Z-score* WHO menggunakan *dataset* stunting dari *platform* Kaggle, kinerjanya dibandingkan dengan KNN dan *Random Forest*. Status gizi *Severely Stunted* atau *Stunted* merepresentasikan anak yang telah mengalami stunting, sedangkan status gizi Normal atau Tinggi merepresentasikan anak yang tidak berisiko stunting. Penelitian ini bertujuan menerapkan metode FKNN dalam mengklasifikasikan status gizi sebagai indikator risiko stunting berdasarkan faktor antropometri, hasil klasifikasi yang diperoleh diharapkan lebih merepresentasikan karakteristik data stunting sehingga memberikan kontribusi aplikatif dalam pemanfaatan *data mining* di bidang kesehatan.

II. TINJAUAN PUSTAKA

Yunus dan Pratiwi (2023) mengembangkan sistem penentuan status gizi balita di Puskesmas Cakranegara menggunakan KNN dengan *Euclidean Distance*, mencapai akurasi 88,06% dari 134 data uji. Pendekatan berbasis antropometri ini bersifat objektif, namun sangat bergantung pada akurasi data pengukuran, penggunaan indeks tunggal seperti BB/U berpotensi menimbulkan salah interpretasi klinis, mendorong perlunya metode seperti FKNN untuk mengatasi tumpang tindih data antarkelas (Yunus, M., & Pratiwi, 2023).

Rahmawati dkk. (2024) membandingkan performa SVM dan KNN dalam memprediksi status stunting berdasarkan data antropometri usia, jenis kelamin, dan panjang badan pada klasifikasi empat label (Normal, *Severely Stunted*, *Stunted*, Tinggi). SVM sedikit lebih unggul dengan akurasi 97,11% dibandingkan KNN 97,05%, namun *confusion matrix* kedua model menunjukkan kesalahan klasifikasi pada data *borderline*, SVM salah mengklasifikasikan 7 data *Severely Stunted* sebagai *Stunted* dan 10 data Normal sebagai Tinggi, sementara KNN salah mengklasifikasikan 9 data *Stunted* sebagai Normal, mengindikasikan perlunya metode yang lebih mampu membedakan data yang saling tumpang tindih (Rahmawati et al., 2024).

Ramadhani dkk. (2023) menerapkan FKNN untuk meningkatkan akurasi identifikasi Demam Berdarah *Dengue*, menggabungkan perhitungan jarak *Euclidean* dengan derajat keanggotaan *fuzzy* setelah tahap *Min-Max Normalization* dan *Stratified* 10-*Fold Cross Validation*. Akurasi optimal sebesar 85,01% diperoleh pada K=13, namun performa model sangat bergantung pada pemilihan nilai K, dengan akurasi terendah 75,15% pada K=3, penelitian ini juga tidak membahas interpretasi lebih lanjut atas derajat keanggotaan yang dihasilkan, sekalipun karakteristik data medis yang penuh ketidakpastian tersebut relevan dengan data antropometri stunting (Ramadhani et al., 2023).

Stunting

Stunting merupakan kondisi gagal tumbuh akibat kekurangan gizi kronis, ditandai tinggi badan anak di bawah −2 SD (pendek) hingga −3 SD (sangat pendek) menurut standar pertumbuhan WHO. Status gizi ditentukan melalui indikator antropometri Tinggi Badan menurut Usia (TB/U), dinyatakan dalam *Z-Score* yang mengukur selisih hasil pengukuran terhadap median populasi rujukan dalam satuan standar deviasi (Wahyudin et al., 2023). Berdasarkan Permenkes RI No. 2 Tahun 2020, klasifikasi status gizi terbagi menjadi empat kategori: *Severely Stunted* (*Z-Score* < −3 SD), *Stunted* (−3 SD s.d. < −2 SD), Normal (−2 SD s.d. +3 SD), dan Tinggi (> +3 SD) (Kemenkes RI, 2020).

K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) merupakan algoritma klasifikasi yang bekerja dengan menghitung jarak antara data uji dan seluruh data latih menggunakan rumus *Euclidean Distance* (Gunawan et al., 2025). Sejumlah k data latih dengan jarak terdekat kemudian dipilih sebagai tetangga, kelas dari data uji ditentukan berdasarkan kelas yang paling banyak muncul di antara k tetangga tersebut melalui *majority voting*. KNN dikenal sederhana dalam implementasinya dan mampu beradaptasi terhadap data baru tanpa perlu membangun model secara eksplisit, menjadikannya salah satu algoritma klasifikasi yang banyak digunakan pada berbagai bidang termasuk analisis data Kesehatan (Rina, 2023).

Fuzzy K-Nearest Neighbor (FKNN)

FKNN mengintegrasikan logika *fuzzy* pada tahap penentuan kelas, alih-alih *voting* mayoritas, FKNN menghitung derajat keanggotaan (*membership degree*) data uji terhadap setiap kelas berdasarkan kontribusi tertimbang tetangga terdekat melalui tahap fuzzifikasi dan defuzzifikasi, nilai keanggotaan berada pada rentang 0 hingga 1 dan kelas dengan nilai tertinggi ditetapkan sebagai hasil prediksi (Kumbure, 2022). Tahap awal FKNN meliputi normalisasi data menggunakan *Min-Max Normalization* agar seluruh variabel memiliki rentang nilai yang seimbang, dilanjutkan penghitungan derajat keanggotaan menggunakan rumus *fuzzy* membership (Azkiya' Mahmud, 2025). Mekanisme ini membuat FKNN lebih tahan terhadap data yang mengandung *noise*, tumpang tindih antarkelas, dan ketidakseimbangan jumlah data (Roman et al., 2022).

p-ISSN : 2406-7733
e-ISSN : 2597-9922

Random Forest

Random Forest merupakan algoritma *supervised learning* berbasis *ensemble* yang menggabungkan hasil prediksi dari sejumlah *decision tree* untuk menghasilkan satu keputusan akhir. Setiap *decision tree* dibangun menggunakan data pelatihan hasil *bootstrap sampling*, yaitu pengambilan sampel secara acak dengan pengembalian, sehingga menghasilkan keragaman model antarpohon (Chen et al., 2024). Pada tahap klasifikasi, setiap data uji dilewatkan ke seluruh pohon untuk memperoleh prediksi dari masing-masing pohon, keputusan akhir ditentukan melalui *majority voting* berdasarkan kelas dengan jumlah suara terbanyak (Edeh et al., 2022).

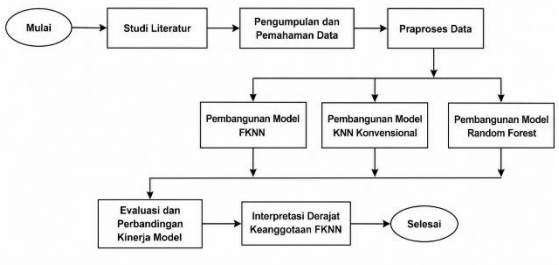
Evaluasi Model

Evaluasi model klasifikasi dilakukan menggunakan *confusion matrix*, yang memberikan analisis distribusi prediksi benar dan salah pada setiap kelas berupa *True Positive, True Negative, False Positive, dan False Negative*. Berdasarkan komponen tersebut dihitung empat metrik evaluasi: akurasi, yang mengukur proporsi data yang diklasifikasikan dengan benar dari keseluruhan data uji; *precision*, yang mengukur proporsi prediksi positif yang benar-benar positif; *recall*, yang mengukur kemampuan model menemukan seluruh data yang seharusnya masuk ke kelas tertentu; serta *F1-Score* sebagai rata-rata harmonik *precision* dan *recall*, digunakan sebagai metrik penyeimbang pada kondisi distribusi kelas yang tidak seimbang (Helmiyah & Pramestiawan, 2025). Pada klasifikasi multi-kelas, keempat metrik tersebut dihitung per kelas menggunakan pendekatan *one-vs-rest*, kemudian digabungkan melalui *macro-average* agar kelas dengan jumlah data lebih sedikit tetap memberikan kontribusi setara terhadap nilai akhir.

III. METODE PENELITIAN

Jenis dan Rancangan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif berbasis komputasi (*computational quantitative research*), di mana analisis dilakukan melalui algoritma untuk menguji hipotesis yang ditetapkan (Adil, et al., 2023). Berdasarkan sifat dan tujuannya, penelitian ini tergolong penelitian eksperimental komputasi (*computational experimental research*), dilakukan melalui serangkaian eksperimen terhadap algoritma dan model klasifikasi *Fuzzy K-Nearest Neighbor* (FKNN), sebuah algoritma berbasis *instance* yang mengintegrasikan teori himpunan *fuzzy* ke dalam kerangka *K-Nearest Neighbor* standar (Simanjuntak et al., 2021). Rancangan penelitian mengikuti alur *data-driven research* yang terdiri atas enam tahapan: (1) studi literatur, (2) pengumpulan dan pemahaman data, (3) praproses data, (4) pembangunan model klasifikasi FKNN, KNN konvensional, dan *Random Forest*, (5) evaluasi dan perbandingan kinerja model, serta (6) interpretasi derajat keanggotaan FKNN, sebagaimana ditunjukkan pada diagram alir penelitian Gambar 1.



Gambar 1. Alur Penelitian

Seluruh tahapan tersebut dijelaskan secara rinci sebagai berikut.

1. Studi literatur

Kajian literatur yang dilakukan mencakup topik-topik stunting dan faktor antropometri, algoritma FKNN, algoritma KNN, dan algoritma Random Forest serta kajian penelitian terdahulu guna mengidentifikasi celah penelitian.

2. Pengumpulan dan pemahaman data

Data yang digunakan merupakan data sekunder dari *platform* Kaggle, yaitu dataset *"Stunting Toddler (Balita) Detection (121K rows)"* yang diunggah oleh Rendi Putra Pradana (2024), terdiri atas 121.000 baris data dengan variabel Umur, Jenis Kelamin, Tinggi Badan, dan Status Gizi sebagai label target berdasarkan Z-score WHO (Kemenkes RI, 2020). Validitas dataset didukung penggunaannya pada beberapa penelitian terdahulu, di antaranya perbandingan *Deep Neural Network* dan *Random Forest* (Lestari et al., 2024), klasifikasi menggunakan *Support Vector Machine* (Candra et al., 2024), dan prediksi menggunakan *Extreme Gradient Boosting* dengan *Bayesian Optimization* (Pratama & Baita, 2025).

Mengingat keterbatasan perangkat keras penelitian (prosesor Intel Core i3, RAM 4GB), sampel sebanyak 2.450 data ditentukan menggunakan rumus Slovin dengan *margin of error* 2% (Sugiyono, 2019), diambil melalui teknik *stratified random sampling* proporsional agar representasi tiap kelas Status Gizi tetap sesuai proporsi populasi.

3. Praproses data

Eksplorasi data (*Exploratory Data Analysis*/EDA) dilakukan sebelum tahap praproses untuk memahami struktur, karakteristik, dan distribusi data, meliputi pemeriksaan dimensi dan tipe data, analisis statistik deskriptif, visualisasi distribusi menggunakan histogram dan *boxplot*, analisis distribusi kelas target, serta analisis korelasi antar fitur.

Analisis korelasi antara fitur Umur dan Tinggi Badan menggunakan koefisien *Pearson*

p-ISSN : 2406-7733
e-ISSN : 2597-9922

menghasilkan nilai r = 0,8456 (p < 0,001), didukung koefisien *Spearman* r = 0,8439 (p < 0,001), menunjukkan hubungan positif yang kuat antara kedua fitur sejalan dengan pola pertumbuhan antropometri anak (Sugiyono, 2019). Analisis korelasi antara fitur Jenis Kelamin dan Status Gizi menggunakan uji *Chi-Square* menghasilkan nilai χ^2 = 7,7837 (df = 3, p = 0,0507) dengan koefisien *Cramér's* V = 0,0395, mengindikasikan asosiasi yang sangat lemah antara kedua variabel pada *dataset* ini

Tahap praproses data meliputi pemeriksaan *missing value* dan data duplikat, *label encoding* pada fitur Jenis Kelamin dan Status Gizi, deteksi *outlier* menggunakan metode *Interquartile Range* (IQR), pembagian data latih dan data uji dengan proporsi 80:20 secara stratifikasi, serta normalisasi *Min-Max* pada fitur numerik yang dihitung hanya dari data latih untuk mencegah *data leakage*. Ketidakseimbangan kelas ditangani melalui *stratified sampling* dan *stratified split*, tanpa penerapan teknik *oversampling* seperti SMOTE.

4. Pembangunan model klasifikasi

Pembangunan model klasifikasi dilakukan terhadap tiga algoritma pada data hasil praproses yang sama, yaitu FKNN sebagai metode utama, KNN konvensional sebagai algoritma dasar pembandingnya, dan *Random Forest* sebagai metode pembanding.

Nilai k optimal untuk FKNN dan KNN dipilih dari kandidat k ∈ {3, 5, 7, 9, 11, 13, 15} menggunakan *Stratified 10-Fold Cross Validation*, sedangkan kombinasi *hyperparameter Random Forest* (jumlah estimator, kedalaman maksimum, jumlah sampel minimum untuk pemisahan *node*, dan jumlah fitur per pemisahan) dioptimasi melalui prosedur validasi silang yang sama untuk menjaga konsistensi pengujian antarmodel. Perhitungan jarak antar data pada FKNN dan KNN menggunakan *Euclidean Distance*, dengan derajat keanggotaan *fuzzy* pada FKNN dihitung mengikuti formula yang dikembangkan oleh Keller et al. (1985) (Keller & Gray, 1985).

5. Evaluasi dan perbandingan kinerja model

Evaluasi dan perbandingan kinerja ketiga model dilakukan menggunakan data uji yang sama, diukur melalui metrik *accuracy*, *precision*, *recall*, dan *F1-Score* yang dihitung per kelas dengan pendekatan *one-vs-rest* kemudian digabungkan melalui *macro-average* (Helmiyah & Pramestiawan, 2025).

6. Interpretasi derajat keanggotaan FKNN

Tahap akhir berupa interpretasi derajat keanggotaan FKNN diarahkan untuk mengidentifikasi kasus data uji pada wilayah tumpang tindih antarkelas, khususnya data dengan nilai Tinggi Badan mendekati ambang batas *Z-score*, guna menilai kemampuan FKNN merepresentasikan tingkat ketidakpastian klasifikasi.

Instrumen Penelitian

Penelitian dilaksanakan menggunakan laptop dengan prosesor Intel *Core* i3-1115G4 dan RAM 4GB, memanfaatkan Google *Colaboratory* sebagai lingkungan pengembangan dengan bahasa pemrograman Python serta *library* NumPy, Pandas, Scikit-learn, Matplotlib, dan Seaborn untuk praproses data, pembangunan model, validasi silang, evaluasi, dan visualisasi hasil penelitian.

IV. HASIL DAN PEMBAHASAN

Hasil Penelitian

1. Hasil Pengumpulan dan Pemahaman Data

Penelitian ini menggunakan *dataset* sekunder *"Stunting Toddler (Balita) Detection (121K rows)"* yang diperoleh dari platform Kaggle. Melalui teknik *stratified random sampling* proporsional, diambil sampel sebanyak 2.450 data guna menyesuaikan dengan kapasitas komputasi perangkat keras yang digunakan. Karakteristik *dataset* final terdiri dari empat kolom utama, yaitu Umur (bulan), Jenis Kelamin, Tinggi Badan (cm), dan Status Gizi. Struktur detail dari *dataset* final disajikan pada Tabel 1.

Tabel 1. Struktur *Dataset* Final

No	Nama Kolom	Tipe Data	Contoh Nilai
1	Umur (bulan)	*Int*64	7
2	Jenis Kelamin	*Object*	laki-laki
3	Tinggi Badan (cm)	*Float*64	75.2
4	Status Gizi	*Object*	normal

Dataset ini menunjukkan kondisi ketidakseimbangan kelas (*class imbalance*), di mana kategori Normal mendominasi sebesar 56% (1.372 data), sedangkan kelas Stunting merupakan minoritas dengan proporsi 11,43% (280 data). Distribusi frekuensi kelas selengkapnya ditunjukkan pada Tabel 2.

Tabel 2. Distribusi Kelas *Dataset*

No	Status Gizi	Jumlah Data	Presentase (%)
1	Normal	1.372	56,00%
2	*Severely Stunted*	402	16,41%
3	Tinggi	396	16,16%
4	*Stunted*	280	11,43%
	Total	2.450	100,00%

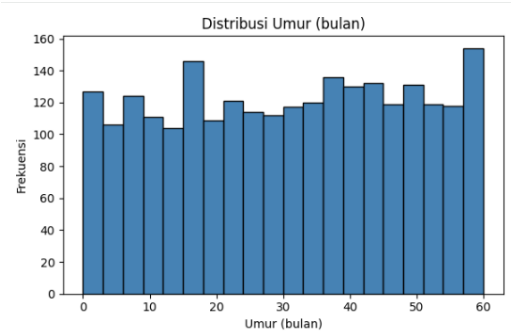
Eksplorasi data awal (EDA) dilakukan untuk mengidentifikasi statistik deskriptif dari fitur numerik. Hasil statistik deskriptif disajikan pada Tabel 3.

p-ISSN : 2406-7733
e-ISSN : 2597-9922

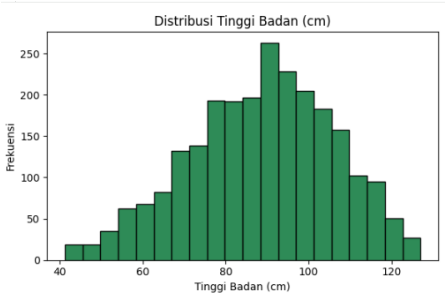
Tabel 3. Statistik Deskriptif Fitur

No	Fitur	Minimum	Maksimum	Rata-rata	Std. Deviasi
1	Umur (bulan)	0,00	60,00	30,28	17,50
2	Tinggi Badan (cm)	41,19	127,00	88,63	17,35

Berdasarkan Tabel 3, fitur Umur memiliki sebaran yang merata dengan rata-rata 30,28 bulan. Visualisasi sebaran ini dapat diidentifikasi melalui Gambar 2 dan Gambar 3, di mana fitur Umur menunjukkan distribusi yang seragam, sedangkan fitur Tinggi Badan menunjukkan bentuk unimodal dengan kecenderungan sedikit condong ke kiri (*left-skewed*).

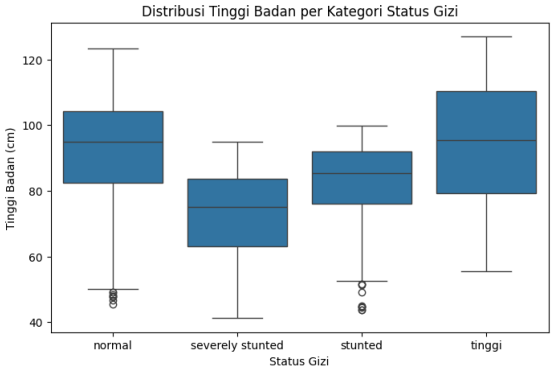


Gambar 2. Distribusi Fitur Numerik Umur



Gambar 3. Distribusi Fitur Numerik Tinggi Badan

Analisis lebih lanjut menggunakan *boxplot* terhadap Tinggi Badan untuk setiap kategori Status Gizi (Gambar 4) menunjukkan adanya tumpang tindih (*overlap*) nilai yang signifikan, terutama pada rentang interkuartil antara kategori Stunting (76–92 cm) and Normal (83–104 cm). Hal ini memvalidasi batas antarkelas status gizi yang tidak sepenuhnya tegas (*borderline*) secara klinis.



Gambar 4. *Boxplot* Tinggi Badan per Kategori Status Gizi

Untuk menguji kekuatan hubungan antarvariabel, dilakukan uji korelasi *Pearson* dan *Spearman* untuk fitur numerik, serta uji *Chi-Square* dan koefisien *Cramér's V* untuk fitur kategorikal. Hasil pengujian dirangkum dalam Tabel 4.

Tabel 4. Hasil Uji Korelasi dan Asosiasi Antar Fitur

No	Pengujian	Fitur	Nilai	Signifikansi
1	*Pearson*	Umur - Tinggi Badan	r = 0,8497	P < 0,0001
2	*Spearman*	Umur - Tinggi Badan	r = 0,8486	P < 0,0001
3	*Chi-Square*	Jenis Kelamin - Status Gizi	χ^2 = 3,3493 (df = 3)	P = 0,3408
4	*Cramér's V*	Jenis Kelamin - Status Gizi	V = 0,0370	-

Hasil pada Tabel 4 mengonfirmasi adanya korelasi positif yang sangat kuat dan signifikan antara Umur dan Tinggi Badan (r = 0,8497), sejalan dengan pola pertumbuhan antropometri anak. Sebaliknya, Jenis Kelamin tidak menunjukkan hubungan signifikan secara statistik dengan Status Gizi (P = 0,3408) dan memiliki asosiasi yang sangat lemah (V = 0,0370).

2. Hasil Praproses Data

Pemeriksaan *Missing Value*: Hasil pemanggilan fungsi *isnull().sum()* menunjukkan nilai 0 untuk seluruh atribut, mengindikasikan data lengkap dan tidak memerlukan imputasi.

Pemeriksaan Data Duplikat: Terdeteksi 75 baris data duplikat yang disebabkan oleh kesamaan nilai antropometri pada anak yang berbeda. Baris tersebut dihapus menggunakan fungsi *drop_duplicates()*, menyisakan 2.375 data unik.

Label Encoding: Konversi nilai kategorikal fitur ke numerik dipetakan secara alfabetis otomatis.

p-ISSN : 2406-7733
e-ISSN : 2597-9922

Jenis Kelamin Laki-laki = 0, Perempuan = 1. Status Gizi Normal = 0, *Severely Stunted* = 1, *Stunted* = 2, Tinggi = 3.

Deteksi *Outlier*: Deteksi pencilan menggunakan metode *Interquartile Range* (IQR) menunjukkan tidak adanya *outlier* ekstrem pada fitur Umur maupun Tinggi Badan.

Pembagian Data Latih dan Uji: Menggunakan rasio 80:20 secara tersertifikasi (*stratified split*) untuk mempertahankan proporsi kelas pada data latih (1.900 data) dan data uji (475 data). Distribusi kelas yang seimbang pasca pembagian data dirinci pada Tabel 5.

Tabel 5. Distribusi Kelas pada Data Latih dan Data Uji

No	Status Gizi	Data Latih (%)	Data Uji (%)
1	Normal	56,05%	56,00%
2	*Severely Stunting*	16,68%	16,63%
3	*Stunting*	11,32%	11,37%
4	Tinggi	15,95%	16,00%

Normalisasi *Min-Max*: Fitur Umur dan Tinggi Badan dinormalisasi ke rentang [0,1] berdasarkan parameter data latih untuk menghindari terjadinya kebocoran data (*data leakage*).

3. Hasil Pembangunan Model Klasifikasi

a. *Fuzzy K-Nearest Neighbor* (FKNN)

Optimasi nilai k menggunakan *Stratified 10-Fold Cross Validation* dilakukan terhadap kandidat $k \in \{3, 5, 7, 9, 11, 13, 15\}$. Hasil pengujian rata-rata akurasi validasi silang disajikan pada Tabel 6.

Tabel 6. Hasil Optimasi Nilai k Menggunakan *Stratified 10-Fold* CV (Model FKNN)

k	Akurasi Rata-rata	Std. Deviasi
3	0,9374	0,0177
5	0,9337	0,0134
7	0,9347	0,0140
9	0,9342	0,0158
11	0,9337	0,0147
13	0,9342	0,0153
15	0,9353	0,0163

Berdasarkan Tabel 8, k = 3 terpilih sebagai nilai optimal dengan akurasi rata-rata tertinggi sebesar 93,74%.

Mekanisme perhitungan FKNN dapat ditelusuri secara matematis. Sebagai ilustrasi, pada data uji ke-20 (label asli Stunting) dengan k = 3, tetangga terdekat dihitung menggunakan jarak *Euclidean*. Rincian tetangga terdekat disajikan pada Tabel 7.

Tabel 7. Rincian Tetangga Terdekat Data Uji Ke-20

Peringkat	Indeks Data Latih	Kelas Tetangga	Jarak Euclidean	Keanggotaan *Training* (U)
1	618	Stunting	0,0105	[0,1633; 0,3267; 0,5100; 0]
2	573	Stunting	0,0177	[0; 0,1633; 0,8367; 0]
3	114	Normal	0,0181	[0,8367; 0; 0,1633; 0]

Melalui integrasi rumus agregasi keanggotaan *fuzzy* Keller et al., diperoleh nilai keanggotaan akhir data uji ke-20 untuk setiap kelas: Normal (0,2639), *Severely Stunting* (0,2270), Stunting (0,5092), dan Tinggi (0,0000). Keputusan akhir mengklasifikasikan data sebagai Stunting karena memiliki nilai keanggotaan tertinggi.

b. *K-Nearest Neighbor* (KNN) Konvensional

Model KNN konvensional juga divalidasi silang untuk menjaga konsistensi pengujian. Hasil rata-rata akurasi disajikan pada Tabel 8.

Tabel 8. Hasil Akurasi Nilai k pada Model KNN Konvensional

k	Akurasi Rata-rata	Std. Deviasi
3	0,9279	0,0153
5	0,9242	0,0145
7	0,9111	0,0132
9	0,9095	0,0163
11	0,9084	0,0211
13	0,8995	0,0153
15	0,8884	0,0184

Berbeda dengan FKNN yang relatif stabil, akurasi KNN konvensional menunjukkan tren penurunan yang signifikan seiring bertambahnya nilai k. Nilai k=3 dipilih sebagai model final KNN konvensional.

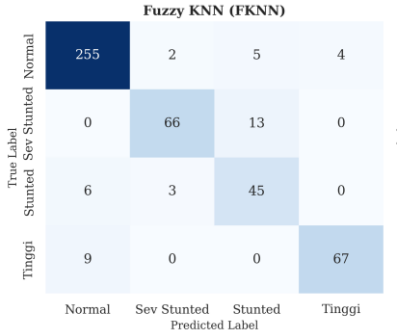
c. *Random Forest*

Melalui *grid search* 10-fold CV pada 96 kombinasi *hyperparameter*, diperoleh parameter optimal: *n_estimators=200*, *max_depth=None*, *min_samples_split=2*, dan *max_features='sqrt'* dengan akurasi 92,63%. Berdasarkan *feature importance* model, fitur Tinggi Badan memiliki kontribusi terbesar (0,613964), diikuti Umur (0,379324), dan Jenis Kelamin dengan kontribusi minimal (0,006712).

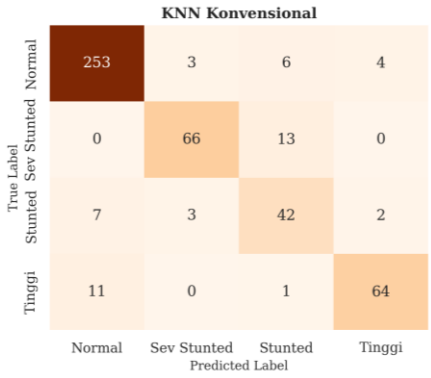
p-ISSN : 2406-7733
e-ISSN : 2597-9922

4. Hasil Evaluasi dan Perbandingan Kinerja Model

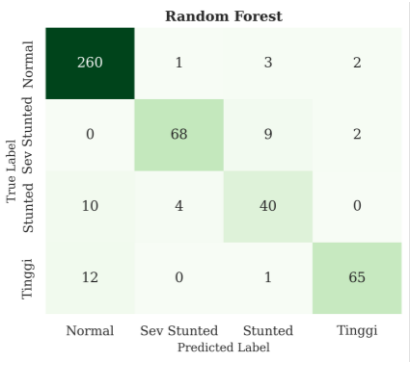
Pengujian dilakukan pada 475 data uji yang belum pernah dilihat sebelumnya. Perbandingan matriks kesalahan (*confusion matrix*) ketiga model disajikan pada Gambar 5, Gambar 6, dan Gambar 7.



Gambar 5. *Confusion Matrix* Model FKNN



Gambar 6. *Confusion Matrix* Model KNN



Gambar 7. *Confusion Matrix* Model *Random Forest*

Berdasarkan *confusion matrix* tersebut, kinerja per kelas untuk metrik *precision, recall,* dan *F1-score* dihitung secara objektif menggunakan pendekatan *one-vs-rest*. Hasil perbandingan disajikan pada Tabel 9.

Tabel 9. Perbandingan Kinerja Per Kelas Ketiga Model

Kelas	Model	*Precision*	*Recall*	*F1-Score*
Normal	FKNN	0,9444	0,9586	0,9515
	KNN	0,9336	0,9511	0,9423
	Random Forest	0,9220	0,9774	0,9489
Severely Stunted	FKNN	0,9296	0,8354	0,8800
	KNN	0,9167	0,8354	0,8742
	Random Forest	0,9315	0,8608	0,8947
Stunting	FKNN	0,7143	0,8333	0,7692
	KNN	0,7000	0,7778	0,7368
	Random Forest	0,7692	0,7407	0,7547
Tinggi	FKNN	0,9437	0,8816	0,9116
	KNN	0,9167	0,8684	0,8919
	Random Forest	0,9559	0,8553	0,9028

Kelas Stunting secara konsisten memiliki nilai kinerja terendah akibat adanya *overlap* data. Meskipun demikian, model FKNN terbukti mencapai *recall* tertinggi pada kelas Stunting (0,8333) dibanding KNN (0,7778) dan *Random Forest* (0,7407). Ringkasan perbandingan metrik kinerja secara keseluruhan (*macro-average*) disajikan pada Tabel 10.

Tabel 10. Perbandingan Kinerja *Macro-Average* Ketiga Model

Model	Accuracy	Precision (Macro)	Recall (Macro)	F1-Score (Macro)
FKNN	91,16%	0,8830	0,8773	0,8781
KNN	89,89%	0,8667	0,8582	0,8613
Random Forest	91,16%	0,8947	0,8586	0,8753

5. Interpretasi Derajat Keanggotaan *Fuzzy K-Nearest Neighbor*

Berbeda dengan metode klasifikasi biner tegas (*crisp*), FKNN menghasilkan derajat keanggotaan *fuzzy* untuk setiap kelas pada setiap baris data uji. Selisih antara derajat keanggotaan tertinggi (*Top*-1) dan tertinggi kedua (*Top*-2) diekstrak sebagai indikator stabilitas keputusan model. Kasus diklasifikasikan sebagai *borderline* jika selisihnya kurang dari 0,1. Dari 475 data uji, teridentifikasi sebanyak 14 kasus (2,95%) berada pada wilayah *borderline* ini.

Untuk melihat kesesuaian keraguan model dengan karakteristik klinis data sebenarnya, nilai Tinggi Badan kasus-kasus tersebut dipetakan terhadap rentang interkuartil (IQR) dari standar antropometri dataset. Rentang tersebut disajikan dalam tabel 11.

p-ISSN : 2406-7733
e-ISSN : 2597-9922

Tabel 11. Rentang Tinggi Badan (IQR) Tiap Kelas

Status Gizi	Batas Bawah (cm)	Median (cm)	Batas Atas (cm)
Normal	82,40	95,10	104,25
Severely Stunting	63,28	75,00	83,80
Stunting	76,00	85,40	92,20
Tinggi	79,45	96,00	110,15

Hasil pemetaan kasus *borderline* disajikan pada Tabel 12.

Tabel 12. Pemetaan Kasus *Borderline* Terhadap Kelas Terdekat

Index	Tinggi Badan (cm)	Label Sebenarnya	Prediksi FKNN	Kelas Terdekat
2127	83,50	Tinggi	Normal	Normal, *Severely* Stunting, Stunting, Tinggi
1109	87,60	Normal	Normal	Normal, Stunting, Tinggi
2046	60,71	Tinggi	Tinggi	*Severely* Stunting
492	84,00	Normal	Normal	Normal, *Severely* Stunting, Stunting, Tinggi
1967	67,60	Stunting	Stunting	*Severely* Stunting
1594	44,71	*Severely* Stunting	Stunting	-
1816	70,80	Stunting	Stunting	*Severely* Stunting
1410	51,90	*Severely* Stunting	Stunting	-
1858	55,30	Stunting	Stunting	-
1136	107,90	Normal	Tinggi	Tinggi

Analisis pemetaan menunjukkan bahwa data dengan Index 2127 and 492 secara geometris berada di dalam rentang *overlap* keempat kelas. Sementara untuk data seperti Index 1594, 1410, dan 1858, ketidakpastian model lebih dipengaruhi oleh interaksi multidimensi antara Tinggi Badan dan Umur.

Selanjutnya, dianalisis perbedaan margin keputusan antara data yang diklasifikasikan dengan benar dan data yang salah diklasifikasikan (Tabel 13).

Tabel 13. Perbandingan Selisih Keanggotaan Prediksi Benar vs Keliru

Status Prediksi	Rata-rata Selisih	Std. Deviasi	Jumlah Data
Benar	0,8713	0,2402	433
Keliru	0,4739	0,3134	42

Rata-rata selisih keanggotaan pada prediksi benar tercatat sangat tinggi (0,8713), mengindikasikan tetangga sekitar yang sangat homogen. Pada prediksi keliru, selisih menyusut drastis menjadi 0,4739, membuktikan model berada dalam kondisi ketidakpastian tinggi di wilayah *borderline*. Meskipun demikian, berkat mekanisme pembobotan invers jarak dalam *defuzzifikasi* FKNN, 9 dari 14 kasus borderline (64,3%) tetap berhasil diklasifikasikan dengan benar

Pembahasan

1. Analisis Fitur Tumbuh Kembang terhadap Status Gizi

Fitur Umur dan Tinggi Badan terbukti secara ilmiah sebagai indikator pertumbuhan utama yang paling krusial dalam klasifikasi stunting. Hal ini dibuktikan dengan korelasi *Pearson* yang sangat tinggi (r = 0,8497) serta kontribusi kepentingan fitur (*feature importance*) pada *Random Forest* yang mencapai 99,3% secara akumulatif. Temuan ini mendukung penyederhanaan komputasi dengan memprioritaskan kedua fitur utama ini. Sebaliknya, Jenis Kelamin tidak memiliki hubungan yang kuat atau bermakna secara statistik terhadap klasifikasi status gizi balita (P = 0,3408 dan V = 0,0370).

2. Analisis Perbandingan Performa Model Klasifikasi

Pengujian menunjukkan bahwa model FKNN berhasil mencapai akurasi keseluruhan sebesar 91,16%, menyamai kinerja *Random Forest* yang telah dioptimasi, dan mengungguli KNN konvensional (89,89%). Keunggulan utama FKNN yang sangat bernilai pada skrining klinis adalah tingkat sensitivitasnya (*recall*) yang sangat tinggi pada kelas minoritas, yaitu 83,33% pada kelas Stunting. Nilai *recall* ini mengungguli KNN konvensional (77,78%) dan secara signifikan melampaui *Random Forest* yang hanya mencapai 74,07%. Hal ini disebabkan oleh bias algoritma pohon keputusan (*Random Forest*) yang cenderung mengarahkan data borderline ke kelas mayoritas (Normal) untuk mengoptimalkan metrik akurasi global, yang menyebabkan 9 dari 54 kasus stunting luput dari deteksi (*false negative*). Pada FKNN, interaksi *fuzzy* dan kontribusi jarak yang proporsional mampu menjaga sensitivitas tinggi ini, meminimalkan risiko luputnya kasus stunting riil di lapangan.

3. Peran Derajat Keanggotaan Fuzzy dalam Mengatasi Ketidakpastian

Integrasi konsep logika *fuzzy* pada FKNN memberikan keunggulan berupa ukuran keyakinan prediktif (*predictive confidence measure*) yang ditunjukkan oleh nilai derajat keanggotaan. Melalui parameter ini, tenaga kesehatan dapat

p-ISSN : 2406-7733
e-ISSN : 2597-9922

mengidentifikasi anak yang berada dalam wilayah abu-abu (*warning zone*) meskipun hasil klasifikasi formalnya adalah Normal (seperti pada Index 2127 dengan selisih keanggotaan hanya 0,0108). Penemuan ini mendukung perancangan triase gizi preventif yang lebih responsif di puskesmas maupun posyandu, tanpa harus menunggu tinggi badan fisik anak merosot melewati ambang batas kaku -2 SD WHO.

V. KESIMPULAN DAN SARAN

Kesimpulan

1. Metode *Fuzzy K-Nearest Neighbor* (FKNN) berhasil diimplementasikan untuk klasifikasi multi-kelas status gizi balita sebagai indikator risiko stunting. Pada nilai k = 3, model FKNN memperoleh akurasi keseluruhan sebesar 91,16%, yang menyamai akurasi model *Random Forest* hasil optimasi *hyperparameter* (91,16%) dan mengungguli KNN konvensional (89,89%).
2. FKNN terbukti sangat efektif dalam mendeteksi kelas minoritas Stunting dengan nilai sensitivitas (*recall*) tertinggi mencapai 83,33%, melampaui KNN konvensional (77,78%) dan *Random Forest* (74,07%). Karakteristik ini meminimalkan tingkat negatif palsu (*false negative*) yang sangat vital dalam konteks skrining kesehatan masyarakat.
3. Nilai derajat keanggotaan *fuzzy* yang dihasilkan oleh FKNN memberikan ukuran keyakinan prediksi (*predictive confidence*) yang objektif. Sebanyak 2,95% kasus diidentifikasi berada di wilayah *borderline* dengan margin perbedaan keanggotaan di bawah 0,1. Pemanfaatan nilai ini sangat bernilai untuk mendeteksi anak-anak dalam kategori 'zona peringatan' sebelum melewati ambang batas fisik *Z-score* stunting.

Saran

1. Penelitian selanjutnya dapat memperluas cakupan dataset dengan mengintegrasikan faktor risiko non-antropometri seperti status sosial ekonomi, riwayat imunisasi, dan pola asupan gizi untuk membangun model klasifikasi stunting yang lebih komprehensif.
2. Pengembangan sistem deteksi dini stunting secara *real-time* berbasis mobile atau web menggunakan API model FKNN ini dapat diimplementasikan untuk memfasilitasi kader posyandu dalam melakukan pencatatan dan skrining klinis yang cepat di lapangan.

DAFTAR PUSTAKA

Adil, A., Liana, Y., Mayasari, R., Lamonge, A. S., Ristiyana, R., Saputri, F. R., ... & Wijoyo, E. B. (2023). *Metode penelitian kuantitatif dan kualitatif: Teori dan praktik*. Get Press indonesia.

Azkiya' Mahmud. (2025). *Analisis Dan Klasifikasi Status Produksi Bebek Pedaging Menggunakan Algoritma Fuzzy K-Nearest Neighbors Thesis Oleh : Azkiya ' Mahmud*.

Candra, A., Erkamim, M., Muharrom, M., & Prayitno, E. (2024). Klasifikasi Stunting Pada Balita Berdasarkan Status Gizi Menggunakan Pendekatan Support Vector Machine (SVM). *Jurnal Ilmiah FIFO*, *16*(2), 171–181. https://doi.org/10.47065/bulletincsr.v4i6.373

Chen, Y., Cheung, K. C., Sun, R. Z., & Yam, S. C. P. (2024). Risk and Statistics in Actuarial Science A user guide of CART and random forests with applications in FinTech and InsurTech. *Japanese Journal of Statistics and Data Science*, *7*(2), 999–1038. https://doi.org/10.1007/s42081-024-00258-x

Dwinanto, R. W., A, A. S. S., & Ardianto, R. (2024). *Klasifikasi Berisiko Stunting Pada Balita : Perbandingan K- Nearest Neighbor , Naïve Bayes , Support Vector Machine*. *8*(2), 264–273.

Edeh, M. O., Dalal, S., Obagbuwa, I. C., Prasad, B. V. V. S., Ninoria, S. Z., Wajid, M. A., & Ademola Olusola Adesina. (2022). Bootstrapping random forest and CHAID for prediction of white spot disease among shrimp farmers. *Scientific Reports*, *12*, 1–12. https://doi.org/10.1038/s41598-022-25109-1

Ernawan, F. E., Andrian, R. B. K., Zarlis, M., Oktavia, T., & Buaton, R. (2024). (2024). *Optimalisasi klasifikasi stunting dengan metode clustering dan rule base time series (RBT)*. CV. Azka Pustaka.

Gunawan, H., Chusyairi, A., & Saputra, M. I. (2025). Penerapan K-Nearest Neighbor dengan Metode Euclidean Distance untuk Klasifikasi Tingkat Ketebalan Cat di PT XYZ. *JUTEKOM Jurnal Teknologi Dan Ilmu Komputer*, *01*(2), 59–72. https://doi.org/10.35134/Jutekom.v9i2.1

Helmiyah, S., & Pramestiawan, R. (2025). *Analisis Komparatif Algoritma Machine Learning dengan Metrik Akurasi , Presisi , Recall , dan F1-Score pada Dataset Kacang Kering*. *6*(3), 152–159.

Imeldawati, R. (2025). *Dampak Terjadinya Stunting terhadap Perkembangan Kognitif Anak : Literature Review*. *1*.

Keller, J. M., & Gray, M. R. (1985). A Fuzzy K-Nearest Neighbor Algorithm. *IEEE Transactions on Systems, Man, and Cybernetics*, *smc-1*(4), 580–585.

Kemenkes RI. (2020). *Peraturan Menteri Kesehatan Republik Indonesia Nomor 2 Tahun 2020 tentang Standar Antropometri Anak*. Jakarta, Kemenkes RI.

Kumbure, M. M. (2022). *Novel Fuzzy K -Nearest Neighbor Methods For Effective Classification And Regression*.

Lestari, W. S., Saragih, Y. M., & Caroline. (2024).

p-ISSN : 2406-7733
e-ISSN : 2597-9922

Comparison of Deep Neural Networks and Random Forest Algorithms for Multiclass Stunting Prediction in Toddlers. *TEKNIKA*, *13*(November), 412–417. https://doi.org/10.34148/teknika.v13i3.1063

Munthe, E. S., Sitorus, I. A. T., Gerald, F., Manoppo, C., & Sari, D. P. (2024). *Sistem Pakar untuk Identifikasi Risiko Proyek Teknologi Informasi Berbasis Metode Fuzzy Logic*. *13*(November), 1710–1716.

Onis, M. De, Dewey, K. G., Borghi, E., Onyango, A. W., Blössner, M., Daelmans, B., Piwoz, E., & Branca, F. (2025). *Orginal Article The World Health Organization ' s global target for reducing childhood stunting by 2025 : rationale and proposed actions*. *9*(2013), 6–26. https://doi.org/10.1111/mcn.12075

Pebrianti, S. W., Astuti, R., & Basysyar, F. M. (2024). Penerapan Algoritma K-Nearest Neighbor Dalam Klasifikasi Status Stunting Di Desa Bojongemas. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(2).

Pratama, D. I., Insani, F., Yanto, F., & Afrianty, I. (2025). Perbandingan Metode K-Nearest Neighbor dan Modifiedk-Nearest Neighbor Untuk Klasifikasi Keluarga Beresiko Stunting. *ZONASI Jurnal Sistem Informasi*, *7*(1), 37–48.

Pratama, R. Y., & Baita, A. (2025). Prediksi Stunting pada Anak Balita Menggunakan Algoritma Extreme Gradient Boosting dan Bayesian Optimization. *Journal of Technology and Informatics (JoTI)*, *7*(2), 175–189. https://e-journals.dinamika.ac.id/joti/article/view/1174/511

Rahmawati, G., Sanmas, S. A., Nudyawati, E., & Syaharani, N. D. (2024). *Studi Perbandingan Performa : Prediksi Status Stunting Pada Anak Berdasarkan Data Antropometri Menggunakan Algoritma Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN)*. *2024*(Senada), 782–790.

Ramadhani, F., Satria, A., & Sari, I. P. (2023). Implementasi Metode Fuzzy K-Nearest Neighbor dalam Klasifikasi Penyakit Demam Berdarah. *Hello World Jurnal Ilmu Komputer*, *2*(2), 58–62.

Ratnasari, Wahidin, A. J., & Andika, T. H. (2024). *Deteksi Dini Stunting Pada Anak Berdasarkan Indikator Antropometri dengan Menggunakan Algoritma Machine Learning*. 378–387. https://doi.org/10.33364/algoritma/v.21-2.2122

Rina, E. (2023). *Algoritma K-Nearest Neighbor (KNN): Penjelasan dan Implementasi untuk Klasifikasi Kanker*. Medium. https://esairina.medium.com/algoritma-k-nearest-neighbor-knn-penjelasan-dan-implementasi-untuk-klasifikasi-kanker-ff9b7fbe0a4

Roman, M., Nawab, H. U., & Ahmad, S. (2022). *K-Nearest Neighbor And Fuzzy K-Nearest Neighbor Algorithm Performance Analysis For Heart Disease Classification*. *19*(1).

Simanjuntak, J., Santoso, E., & Marji. (2021). Klasifikasi Penyakit Infeksi Saluran Pernapasan Akut (ISPA) dengan menerapkan Metode Fuzzy K-Nearest Neighbor. *5*(11), 5023–5029.

Sugiyono. (2019). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. CV. Alfabeta.

Wahyudin, W. C., Hana, F. M., & Prihandono, A. (2023). Prediksi stunting pada balita di rumah sakit kota semarang menggunakan naive bayes. *Jurnal Ilmu Komputer Dan Matematika*, *4*(1), 32–36.

World Health Organization. (2025). *Global nutrition targets 2030: stunting brief*. World Health Organization. https://www.who.int/publications/i/item/B09383

Yunus, M., & Pratiwi, N. K. A. (2023). Prediksi Status Gizi Balita dengan Algoritma K-Nearest Neighbor (KNN) di Puskemas Cakranegara. *JTIM : Jurnal Teknologi Informasi Dan Multimedia*, *4*(4), 221–231. https://journal.sekawan-org.id/index.php/jtim/article/view/328