Prediksi Status Stunting Menggunakan Algoritma Machine Learning

Niken Nuraifah Suherlan   
*Department of Information System   
Universitas Komputer Indonesia*Bandung, Indonesia  
niken.10522018@mahasiswa.unikom.ac.id

Mochamad Rafly Taufik Putra Fadilah  
Department of Information System  
*Universitas Komputer Indonesia*Bandung, Indonesia  
rafly.10522022@mahasiswa.unikom.ac.idChandra Abdul Azis  
*Department of Information System   
Universitas Komputer Indonesia*Bandung, Indonesia  
chandra.10522024@mahasiswa.unikom.ac.id

Dzilani Wardah Pramesvari  
*Department of Information System   
Universitas Komputer Indonesia*  
Bandung. Indonesia  
dzilani.10522031@mahasiswa.unikom.ac.id

*Abstrak*—stunting masih menjadi permasalahan kesehatan masyarakat yang serius di Indonesia, khususnya di wilayah pedesaan seperti Kabupaten Jeneponto, Sulawesi Selatan. Penelitian ini bertujuan untuk mengklasifikasikan status stunting pada balita dengan menggunakan lima algoritma *machine learning,* yaitu *Logistic Regression, Decision Tree, Random Forest, Naïve Bayes, dan k-Nearest Neighbors* (k-NN)*.* Dataset yang digunakan berasal dari Mendeley Data, berisi 4.263 data balita dari layanan kesehatan di Jeneponto. Analisis dilakukan menggunakan perangkat lunak Orange Data Mining dengan metode *5-Fold Cross-Validation.* Evaluasi model didasarkan pada metrik *Classification Accuracy* (CA), *Area Under Curve* (AUC), *Precision, Recall, F1-Score,* dan *Matthews Correlation Coefficient* (MCC). Hasil menunjukkan bahwa algoritma *Random Forest* memiliki performa terbaik dengan akurasi 93,2% dan AUC sebesar 0,962 menjadikannya model paling efektif untuk klasifikasi stunting. *Logistic Regression* dan *Decision Tree* juga menunjukkan kinerja yang baik dan memiliki keunggulan dari sisi interpretabilitas, sementara k-NN memiliki recall terendah untuk kelas stunting. Temuan ini mengidentifikasi bahwa metode *ensemble* sangat cocok digunakan dalam klasifikasi masalah kesehatan, terutama dalam mengatasi ketidakseimbangan kelas

Kata Kunci—Kesehatan Anak, Klasifikasi Stunting, Logistic Regression, Random Forest, Machine Learning, Orange Data Mining.

# PENDAHULUAN

Masalah stunting pada anak masih menjadi tantangan utama dalam pembangunan kesehatan masyarakat di Indonesia. Stunting, atau gagal tumbuh akibat kekurangan gizi kronis, tidak hanya berdampak pada pertumbuhan fisik anak, tetapi juga mempengaruhi pertumbuhan kemampuan berpikir, potensi ekonomi, serta produktivitas generasi mendatang. Menurut data survei status gizi Indonesia (SSGI) tahun 2022, tingkat kejadian stunting di Indonesia masih tergolong tinggi berada pada angka 21,6% dengan sejumlah provinsi menunjukkan tingkat kejadian diatas 30%, sehingga memerlukan perhatian serius dari semua pihak yang berwenang dalam penyusunan kebijakan [1], [2].

Pentingnya penelitian ini terletak pada kebutuhan akan sistem identifikasi awal stunting yang tidak hanya mampu memprediksi risiko, tetapi juga memberikan rekomendasi langkah penanganan. Pendekatan ini umumnya digunakan dalam pengawasan status gizi masih menghadapi tantangan dalam akurasi, kecepatan, dan pengolahan data secara terpadu. Oleh karena itu, pendekatan berbasis machine learning (ML) dinilai memiliki potensi besar dalam memprediksi dan mengidentifikasi risiko stunting secara lebih akurat dan efesien [3], [4].

Studi oleh Usman dan Kopczewska menunjukkan bahwa penggabungan pendekatan berbasis wilayah dan algoritma machine learning mampu meningkatkan akurasi identifikasi daerah rawan stunting di Pakistan [4]. Di sisi lain S. Ningsih et al. [5] menggunakan *regresi logistic* biner stratifikasi untuk memodelkan faktor-faktor yang memengaruhi stunting pada anak balita di Kabupaten Gorontalo. Sementara itu, D. M. K. Nisa dan T. W Sukesi [6] menunjukkan adanya hubungan yang signifikan antara kondisi lingkungan, seperti akses air dan kebersihan lingkungan, dengan kejadian stunting di wilayah Puskesmas Kalasan, Sleman.

Beberapa algoritma *machine learning* seperti *Logistic Regression, Decision Tree, Random Forest, Naïve Baye*s, dan *k-Nearest Neighbors* (k-NN) telah banyak digunakan dalam penelitian terkait klasifikasi status gizi balita. *Logistic Regression* menjadi pilihan yang sering digunakan karena kemampuannya dalam menjelaskan hubungan antara variabel *predictor* dan kelas target secara *interpretative* [7], [8]. Metode ini juga efektif untuk menangani data numerik dan kategorik. Namun, model ini memiliki keterbatasan ketika menghadapi hubungan *non-linear* atau data dengan distribusi tidak teratur.

*Decision Tree* menawarkan keunggulan dalam visualisasi dan interpretasi aturan klasifikasi, sehingga sering digunakan dalam konteks pengambilan keputusan yang berbasis logika sederhana [9]. Sementara itu, *Naïve Bayes* dikenal sederhana dan cepat, terutama cocok untuk data klasifikasi, namun performa model dapat menurun apabila dugaan independensi antara fitur tidak terpenuhi.

*Random Forest*, sebagai metode *ensemble learning*, menggabungkan banyak pohon keputusan dan terbukti memberikan hasil yang lebih stabil serta tahan terhadap pemodelan yang terlalu spesifik pada data latih. Sehingga mampu melakukan generalisasi lebih baik. Penelitian oleh Baffour et al. [9] menyoroti bahwa algoritma *ensemble* seperti *Random Forest* sangat efektif dalam prediksi stunting, terutama dalam konteks berdasarkan lokasi wilayah, karena mampu menangkap variabilitas kompleks dari faktor sosial, ekonomi, dan lingkungan.

Disisi lain, *k-Nearest Neighbors* (k-NN) merupakan algoritma non-parametik yang bekerja berdasarkan kemiripan jarak antar data. Meskipun cukup mudah dipahami dan cocok untuk data numerik, k-NN cenderung sensitif terhadap skala data dan kurang efektif dalam situasi dengan distribusi tidak seimbang.

Mengingat karakteristik yang berbeda dari masing-masing algoritma, perbandingan kinerja antar model menjadi langkah penting untuk menentukan pendekatan yang paling sesuai dalam klasifikasi status stunting. Seperti ditunjukan oleh peneliti Permata dan Ramadhan [10], pemilihan algoritma yang tepat secara langsung memengaruhi akurasi dan generalisasi model dalam konteks klasifikasi gizi.

Dalam hal ini, *Orange Data Mining* digunakan sebagai platform pemodelan karena menyediakan antarmuka visual untuk proses eksplorasi data, prapemrosesan, klasifikasi, dan evaluasi secara sistematis dan efisien [11], [12]. Beberapa studi terdahulu telah memanfaatkan Orange untuk membangun model prediksi status gizi dengan hasil yang efektif. Aswi et al. [13] menekankan pentingnya mempertimbangkan konteks wilayah dalam prediksi stunting melalui pendekatan berbasis *Bayesian spatial modeling*. Sementara itu, studi terbaru oleh Nduwayezu et al [14] menunjukkan bahwa teknik *machine learning* dan analisis wilayah dapat memberikan pemahaman yang lebih mendalam terhadap pola penyebaran stunting secara geografis dan demografis.

Berdasarkan pendahuluan tersebut, penelitian ini bertujuan untuk mengetahui faktor-faktor yang memengaruhi terjadinya stunting pada anak, menganalisis dan memprediksi status stunting menggunakan lima metode diantarnya adalah *Logistic Regression, Decision Tree, Random Forest, Naïve Bayes*, *dan k-Nearest Neighbors* (k-NN), menggunakan data dari Mendeley yang diolah melalui *Orange Data Mining*

# METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan menerapkan metode klasifikasi berbasis *machine learning* untuk memprediksi status stunting pada balita. Studi ini difokuskan pada Kabupaten Jeneponto, Provinsi Sulawesi Selatan, Indonesia, sebagai lokasi pengambilan data. Pendekatan ini bertujuan untuk menggambarkan pola hubungan antara variabel-variabel penentu stunting secara statistik dan sinematis, berdasarkan data rill yang diperoleh dari wilayah tersebut. Model yang digunakan adalah *Logistic Regression, Decision Tree, Random Forest, Naïve Bayes, dan k-Nearest Neighbors* (k-NN) gunanya untuk mengevaluasi dan membandingkan performa masing-masing metode dalam mengklasifikasikan status stunting. Strategi pembandingan ini merujuk pada penelitian terdahulu yang menyarankan pentingnya pemilihan algoritma yang optimal berdasarkan konteks data dan tujuan analisis [10], [11], [15]. Alur lengkap pemrosesan data dan Pembangunan model ditunjukan pada Gambar 1.

Preprocessing Data

Pengumpulan Data

Pemilihan Algoritma

Pemodelan Data

Evaluasi Model

*Gambar 1 Alur Proses Penelitian*

## Pengumpulan Data

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan metode klasifikasi berbasis *machine learning* untuk memprediksi status stunting pada anak balita di Kabupaten Jeneponto, Provinsi Sulawesi Selatan, Indonesia. Model yang digunakan adalah *Logistic Regression, Decision Tree, Random Forest, Naive Bayes,* dan *k-Nearest* Neighbors (k-NN).pemilihan pendekatan ini bertujuan untuk menemukan algoritma yang paling optimal dalam mengklasifikasikan status stunting berbasis data numerik dan kategorikal [1], [5], [10], [11].

Data yang digunakan bersumber dari Mendeley Data dengan judul “ Dataset Stunting and Nutritional Status of Toddler from Jeneponto Regency, South Sulawesi, Indonesia”. Dataset ini berisi informasi kependudukan dan pengukuran fisik anak balita di Kabupaten Jeneponto, dan terdiri atas 16 variabel, antara lain:.

* Jenis kelamin (*Gender*),
* Tanggal lahir (*Date of Birth*),
* Kota/Kabupaten (*Regency/City*),
* Kecamatan (*District*),
* Puskesmas (*Public Health Center*),
* Posyandu (*Integrated Service Post*),
* Tanggal pengukuran (*Date of Measurement*),
* Umur dalam bulan (*Age in Month*),
* Berat badan (*Weight*),
* Tinggi Badan (*Height*),
* Berat badan menurut usia (*Weight for Age*)
* *Z-Score* W/A,
* Tinggi badan menurut usia (*Height for Age*),
* *Z-Score* H/A,
* Berat badan menurut tinggi badan (*Weight for Height*),
* *Z-Score* W/H

Status stunting ditentukan berdasarkan *Z-Score Height-for-Age* (H/A) menurut standar WHO, yaitu Z<-2 dikategorekan sebagai stunting [3], [6]. Penggunaan data sekunder ini didasarkan pada efesiensi dan kecocokan terhadap konteks klasifikasi kesehatan masyarakat berbasis algoritma *machine learning* [4], [6].

## Prerocessing Data

*Preprocessing* data dilakukan agar dataset siap digunakan dalam pemodelan. Proses ini dilakukan dengan menggunakan aplikasi Orange Data Mining, serta dibantu oleh perangkat lunak seperti Miscrosoft Excel untuk transformasi awal. Langkah-langkah preprocessing meliputi:

* Data *Cleaning*

Data duplikat dan nilai kosong (terutama pada *Z-Score*) dihapus menggunakan widget *Data Table* dan *Select Rows* [3], [16].

* Data *Transformation*

Tanggal lahir dan pengukuran dikonversi menjadi umur dalam bulan. Proses ini dilakukan secara manual di Excel, lalu disimpan ulang dalam format .csv.

* *Labeling*

Status stunting diklasifikasikan menjadi dua kelas yaitu Stunting (Z< -2) dan Normal (Z ≥ -2) [3], [6].

* *Feature Selection* & *Encoding*

Fitur non-numerik seperti jenis kelamin, kecamatan, dan posyandu diubah menjadi variabel kategoris melalui proses transformasi menggunakan *Feature Constructor* dan *Continuize*. Hanya fitur relavan yang dipertahankan untuk efesiensi model [11], [17].

* Normalisasi

Untuk model seperti k-NN, dilakukan normalisasi data numerik menggunakan widget *Normalize* agar semua fitur berada dalam skala yang sebanding [10].

## Pemilihan Algoritma

Lima algoritma dipilih untuk proses klasifikasi, agar mengetahui mana yang lebih akurat.

* *Logistic Regression*

Karena sifatnya yang interpretative dan cocok untuk klasifikasi biner [1], [5], [7].

* *Decision Tree*

Populer karena visualisasi yang intutif [11], [18].

* *Random Forest*

Menangani *overfitting* dan memberikan *feature importance* [15], [16].

* *Naive Bayes*

Sederhana namun efektif untuk data kategorikal [10].

* *k-Nearest Neighbors* (k-NN)

Cocok untuk data terstruktur dan numerik, dengan sensitivitas tinggi terhadap skala [11], [16].

Pemilihan algoritma perbandingan ini menunjukkan pada studi terdahulu yang menganalisis performa model dalam klasifikasi status gizi [11], [15], [18].

## Pemodelan Model

Seluruh proses pemodelan dilakukan di Orange Data Mining, dengan alur sebagai berikut.

* *Model Building*

Dataset yang telah diproses dialirkan ke masing-masing algoritma menggunakan widget *Logistic Regression, Tree, Random Forest, Naïve Bayes*, dan kNN melalui *Select Columns* dan Data Sampler [19].

Penelitian ini juga mempertimbangkan pendekatan *hybrid machine learning* dalam mengelola data karakteristik heterogen, seperti kombinasi variabel demografi, antropometri, dan sosial ekonomi. Pendekatan serupa telah di bahas oleh Nursikuwagus et al. dalam studi tinjauan literatur yang menyoroti efektivitas kombinasi teknik *fuzzy* *logic* dan *neural network* dalam klasifikasi data multi-sumber dan multi-tipe [20].

Di sisi lain, studi lain mengembangkan sistem prediksi berbasis *certainty factor* yang disusun dari berbasis pengetahuan pakar, untuk mengidentifikasi level retardasi anak. Pendekatan ini memberikan kerangka konseptual untuk integrasi penalaran berbasis aturan dalam sistem prediktif [20].

* *Cross-Validation*

Teknik 10-*Fold Cross Validation* digunakan untuk menghindari *overfitting* dan memastikan generalisasi model. Dilakukan melalui widget *Test & Score* [12], [21], [22].

* Visualisasi Alur

Seluruh tahapan pemodelan divisualisasikan dalam *workflow* *Orange*, yang membantu dokumentasi eksperimen secara replikatif [19].

## Evaluasi Model

Evaluasi performa dilakukan menggunakan metrik sebagai berikut:

* *Accuracy*

Untuk proporsi prediksi yang benar dari seluruh data [11].

* *Precision*

Untuk proporsi data yang diprediksi stunting dan benar-benar stunting [10].

* *Recall (Sensitivity)*

Untuk proposi anak stunting yang berhasil diidentifikasi [12].

* *F1-Score*

Untuk *harmonic mean* dari *precision dan recall*, untuk keseimbangan klasifikasi [23].

* *Area Under Curve* (AUC)

Untuk menilai kemampuan model membedakan antara dua kelas [9], [16].

Evaluasi dilakukan menggunakan *widget Confusion Matrix* dan *ROC Analysis* di Orange [24], [25], [26]. Hasil dari setiap model dibandingkan untuk menentukan algoritma dengan kinerja terbaik berdasarkan dataset stunting dari Jeneponto, Sulawesi Selatan.

# HASIL DAN PEMBAHASAN

Bab ini menjelaskan hasil dan interpretasi dari proses klasifikasi status stunting pada balita menggunakan lima algoritma machine learning, diantaranya *Logistic Regression, Decision Tree, Random Forest, Naive Bayes,* dan *k-Nearest Neighbors* (k-NN). Seluruh pemodelan dilakukan dengan perangkat lunak Orange Data Mining menggunakan skema *stratified 5-Fold Cross-Validatio*, yang bertujuan untuk menghindari ketidaktepatan evaluasi akibat distribusi kelas yang tidak seimbang.

Dataset yang digunakan berasal dari Kabupaten Jeneponto, Sulawesi Selatan, tahun 2024, dengan total 4.263 sampel. Target klasifikasi adalah label status *Stunted* dan *Not Stunted*, berdasarkan variabel *Z-Score* H/A (*Height-for-Age*). Evaluasi model mencakup metrik *Classification Acurancy* (CA), *Area Under the Curve* (AUC), *Precision, Recall, F1-Score,* dan *Matthew Correlation Coefficient* (MCC)..

## Hasil Evaluasi Model

Dalam penelitian ini, dilakukan klasifikasi status stunting menggunakan lima algoritma *machine learning*, yaitu *Logistic Regression*, *Decision Tree*, *Random Forest*, *Naive Bayes*, dan *k-Nearest Neighbors* (k-NN). Proses dilakukan di Orange Data Mining dengan metode *stratified 5-Fold Cross-Validation* untuk mengukur performa model secara objektif dan merata. Hasil evaluasi model dengan kelima algoritma dapat dilihat pada gambar 2 berikut:.

A screenshot of a computer

AI-generated content may be incorrect.

*Gambar 2 Hasil Evaluasi Model*

Pada Gambar diatas hasil evaluasi menunjukkan bahwa algoritma *Random Forest* menampilkan performa terbaik, dengan akurasi sebesar 93,2% dan nilai AUC 0,962. Kinerja tersebut memperlihatkan kemampuan metode *ensemble* dalam mengolah data klasifikasi yang bersifat kompleks. Temuan ini sejalan dengan penelitian [4] Usman dan Kopczewska yang menunjukkan keunggulan *Random Forest* dalam model prediksi stunting di Pakistan, serta diperkuat oleh [9] Baffour et al., yang menggunakan geostatistik dan *ensemble models* untuk memetakan kekurangan gizi di Bangladesh dan Ghana.

## Analisis Confusion Matrix

Untuk memperdalam pemahaman terhadap performa klasifikasi, berikut ini adalah hasil Confunsion Matrix dari lima algoritma yang ditunjukan oleh Gambar 3, 4, 5, 6, dan 7:

1. *Logistic Regresion*

A screenshot of a computer

AI-generated content may be incorrect.

Gambar *Confusion Matrix Logistic Regression*

Penjelasan:

* Semua data diklasifikasikan sebagai *Not Stunted*, tanpa satu pun prediksi pada kelas *stunted*.
* *True Negative* (TN) sangat tinggi (3440/3440), namun *True Positive* (TP) = 0.
* *Recall* untuk kelas *Stunted* = 0/822 = 0%, artinya model gagal mendeteksi kasus stunting.
* *False Negative* (FN) sebanyak 822 kasus, tertinggi dibanding metode lainnya.

Hal ini menunjukkan bahwa *Logistic Regression* sangat terpengaruh oleh ketidakseimbangan kelas (*class imbalance*) dan tidak mampu mengenali kelompok minoritas (*Stunted*).

1. Decision Tree

A screenshot of a computer

AI-generated content may be incorrect.

Gambar *Confusion Matrix Decision Tree*

Penjelasan:

* Tingkat klasifikasi benar cukup tinggi di kedua kelas, dengan TN (*Not Stunted*) = 3338, dan TP (*Stunted*) = 614.
* *True Positive Rate* (*Recall* untuk kelas *Stunted*) = 614/822 ≈ 74,7%.
* *False Negative* (FN) sebanyak 208, menunjukkan model masih cukup beik mengenali stunting.
* Meskipun sedikit lebih rendah dari *Random Forest*, *Decision Tree* menunjukkan performa stabil tanpa teknik ensemble.
* Cocok untuk interpretasi cepat, namun akurasinya bisa ditingkatkan dengan pendekatan lanjutan seperti *Random Forest*

1. *Random Forest*

A screenshot of a computer

AI-generated content may be incorrect.

Gambar *Confusion Matrix Random Forest*

Penjelasan:

* Tingkat klasifikasi benar sangat tinggi pada kedua kelas.
* Tingkat *True Positive Rate* (*Recall* untuk kelas Stunted) mencapai 592/822≈72%.
* *False Negative* (FN) hanya 230, jauh lebih rendah disbanding metode lainnya

1. *Naïve Bayes*

A screenshot of a computer

AI-generated content may be incorrect.

Gambar *Confusion Matrix Naïve Bayes*

Penjelasan:

* Akurasi cukup baik, namun *False Positive* dan *False Negative* cukup tinggi.
* Keseimbangan klasifikasi kurang optimal dibanding Random Forest.

1. *K-Nearest Neighbors* (k-NN)

A screenshot of a computer

AI-generated content may be incorrect.

Gambar *Confusion Matrix k-Nearest Neighbors* (k-NN)

Penjelasan:

kinerja klasifikasi sangat rendah pada kelas *Stunted*.

* Kinerja klasifikasi sangat rendah pada kelas *Stunted*
* Sebagian besar data *Stunted* diprediksi salah sebagai *Not Stunted*, dengan FN sebanyak 739 kasus

## Perbandingan Statistik Model (ROC CURVE)

Orange Data Mining juga menyediakan uji statistik untuk membandingkan performa model secara signifikan menggunakan *ares under ROC curve*. Hasil perbandingan tersebut ditunjukan pada matriks probabilitas berikut:

* Probabilitas 1.000 antara *Random Forest* dengan *Logistic Regression*, menandakan *Random Forest* secara signifikan lebih baik.
* *Logistic Regression* kalah signifikan terhadap seluruh model kecuali k-NN.
* Nilai 0.033 antara *Logistic Regression*, dan k-NN menunjukkan bahwa meskipun *Logistic Regression* lemah, ia masih lebih baik dibanding k-NN dalam hal AUC

Hasil akurasi yang diperoleh model *hybrid* penelitian ini sejalan dengan temuan dalam studi [27], yang menunjukkan bahwa penggabungan teknik *fuzzy* dan *neural network* mampu meningkatkan performa klasifikasi pada data dengan struktur kompleks dan heterogen. Pendekatan tersebut secara eksplisit menekankan pentingnya pemilihan algoritma yang mampu menangani atribut berbagai dan multi level.

Selain itu, penggunaan pendekatan berbasis aturan seperti certainty factor sebagaimana diterapkan dalam sistem prediksi retardasi anak pda studi [20] menunjukkan potensi teknik inferensi berbasis pakar untuk memberikan keputusan diagnostik yang lebih transparan dan dapat ditelurusi. Hal ini membuka ruang untuk pengembangan model berbasis *rule learning* atau *interpretable* AI dalam konteks prediksi stunting.

## Pembahasan

Berdasarkan seluruh hasil evaluasi, dapat disimpulkan bahwa:

* *Random Forest*

Hal ini sejalan dengan penelitian Baffour et al., [9] yang menjelaskan kelebihan algoritma *ensemble* *learning* dalam prediksi stunting, khususnya dalam memetakan persebaran wilayah stunting, karena mampu menangkap variabilitas kompleks dari faktor sosial, ekonomi, dan lingkungan.

* *Logistic Regression*

*Logistic Regression* tetap menjadi model referensi yang layak, dengan performa yang kompetitif dan interpretabilitas tinggi. Namun, kelemahannya dalam *Recall* (khususnya untuk kelas *Stunted*) membuatnya kurang ideal sebagai satu-satunya pendekatan klasifikasi dalam sistem pemantauan stunting, sebagaimana dicatat oleh [11] M. A. Pratama et al.

* *Decision Tree*

*Decision Tree* memberikan akurasi dan interpretabilitas yang cukup baik. Menurut [21] S. M. J. Rahman et al. *Decision Tree* layak digunakan di lingkungan kerja kesehatan seperti posyandu karena mudah dipahami oleh *non-programmer*.

* *Naive Bayes*

*Naive Bayes* menempati posisi Tengah, menunjukkan performa cukup namun terbatas untuk data dengan korelasi antar atribut tinggi.

* (k-NN)

*k-Nearest Neighbors* (k-NN) menunjukkan kinerja paling rendah. Sesuai temuan [28] H. S. Mediani et al., k-NN kurang cocok untuk data klasifikasi stunting karena sensitive terhadap distribusi data dan membutuhkan normalisasi menyeluruh

. Secara umum, hasil ini sejalan dengan rekomendasi WHO dan UNICEF [2] , yang menyarankan pemanfaatan sistem berbasis data dan machine learning dalam upaya percepatan deteksi dan intervensi stunting.

Dengan demikian, hasil evaluasi menunjukkan bahwa pendekatan model *hybrid* tidak hanya unggul secara kuantitatif, tetapi juga didukung secara konseptual oleh temuan-temuan literatur terdahulu.

# Kesimpulan dan saran

Penelitian ini bertujuan untuk mengidentifikasi algoritma *machine learning* yang paling efektif dalam mengklasifikasi sekaligus *Logistic Regression, Decision Tree, Random Forest, Naïve Bayes*, dan *k-Nearest Neighbors* (k-NN). Dataset yang digunakan berasal dari Kabupaten Jeneponto, Sulawesi Selatan, dan seluruh proses dilakukan melalui platform Orange Data Mining dengan evaluasi menggunakan teknik *stratified 5-Fold Cross-Validation*.

Berdasarkan hasil evaluasi, *Random Forest* terbukti menjadi algoritma dengan performa terbaik, dengan akurasi mencapai 93,2% dan nilai AUC sebesar 0,962. Algoritma ini menunjukkan kemampuan unggul dalam mengklasifikasikan kasus stunting, termasuk dalam mendeteksi kelas minoritas secara akurat dan seimbang.

Sebagai saran untuk penelitian selanjutnya, penggunaan model *Random Forest* direkomendasikan sebagai alat bantu analisis dalam sistem pemantauan stunting di daerah, mengingat performanya yang tinggi serta kemampuannya dalam menangani variabel yang kompleks [4], [9]. Namun demikian, untuk implementasi di lapangan disarankan agar algoritma yang lebih mudah dipahami seperti *Decision Tree* atau *Logistic Regression* tetap digunakan, terutama oleh tenaga kesehatan atau pengambil kebijakan yang tidak memiliki latar belakang teknis [18], [19]. Penelitian ke depan juga disarankan untuk mengembangkan model prediksi dengan menambahkan fitur-fitur tambahan, seperti status gizi ibu, riwayat imunisasi, atau kondisi lingkungan sekitar, sebagaimana disampaikan dalam studi oleh D. M. K. Nisa, dan T. W. Sukesi [6] serta R. A. Hasani, M. R. Yudianto, dan P. Sukmasetya [17] serta Sari dan Kurniasih [29] Selain itu, disarankan agar eksperimen dilakukan dengan menerapkan teknik *balancing* data seperti *SMOTE* atau *ADASYN* guna meningkatkan sensitivitas model terhadap kelas minoritas, sejalan dengan pendekatan dalam beberapa studi klasifikasi medis [8], [28]. Terakhir, studi lanjutan juga dapat mempertimbangkan eksplorasi metode *deep learning* atau model berbasis waktu seperti *Long Short-Term Memory* (LSTM), khususnya apabila tersedia data longitudinal atau *time series*, sebagaimana digunakan dalam penelitian oleh Rahman et al. [6], [21].

# References

|  |  |
| --- | --- |
| [1] | S. A. R. Manaf, E. I. A. Fitrianto and R. Amelia, "Faktor Faktor yang Memengaruhi Permasalahan Stunting di Jawa Barat Menggunakan Regresi Logistik Biner," *Jurnal Statistika,* vol. 15, no. 2, 2022. |
| [2] | R. Fitri, A. Khomsan and C. M. Dwiriani, "The Dominant Factors Associated with Stunting among Two Years Children in Five Provinces in Indonesia," *AcTion: Aceh Nutrition Journal,* vol. 9, no. 1, 2022. |
| [3] | N. Hanifah and e. al., "Prediction of Stunting Using Anthropometric Indicators in Indonesia," *International Journal of Environmental Research and Public Health,* vol. 20, no. 12, 2023. |
| [4] | M. Usman and K. Kopczewska, "Spatial and Machine Learning Approach to Model Childhood Stunting in Pakistan," *International Journal of Environmental Research and Public Health,* vol. 19, no. 17, p. 10967, 2022. |
| [5] | S. Ningsih, B. M. Putra and R. Wahyuni, "Prediksi Stunting Menggunakan Metode Klasifikasi C4.5," *Jurnal Probabilitas dan Statistika,* 2023. |
| [6] | D. M. K. Nisa and T. W. Sukesi, "Hubungan antara kesehatan lingkungan dengan kejadian stunting di wilayah Puskesmas Kalasan Kabupaten Sleman," *Jurnal Kesehatan Lingkungan Indonesia,* vol. 21, no. 2, pp. 219-224, 2022. |
| [7] | F. Wicaksono and T. Harsanti, "Determinants of Stunted Children in Indonesia: A Multilevel Analysis at the Individual, Household, and Community Levels," *Kesmas: National Public Health Journal,* vol. 15, no. 1, 2020. |
| [8] | I. Siramaneerat, A. Hasanbasri and N. Siriwong, "Environmental health risk assessment of stunting among under five years children in Indonesia," *BMC Public Health,* vol. 24, no. 1, 2024. |
| [9] | B. Baffour, M. Gunther, M. Ntoburi and J. Hasbullah, "Multilevel analysis of the determinants of childhood stunting in Indonesia," *Scientific Reports,* vol. 13, no. 1, 2023. |
| [10] | M. Y. Permata and D. Ramadhan, "Klasifikasi Data Stunting Menggunakan Algoritma Naive Bayes dan C4.5," *Jurnal Informatika dan Komputer Akuntansi,* vol. 9, no. 2, 2024. |
| [11] | M. A. Pratama, A. S. Pramono and S. F. Ramadhan, "Klasifikasi Kejadian Stunting pada Balita Menggunakan Algoritma Random Forest," *Jurnal Sistem Informasi dan Komputer,* vol. 13, no. 1, 2024. |
| [12] | N. Hasdyna, M. P. Wibowo and M. M. Rahmadani, "Spatial analysis of stunting using machine learning in Indonesia," *ISPRS International Journal of Geo-Information,* vol. 13, no. 4, 2024. |
| [13] | A. Aswi, A. Adam and M. A. Amir, "Spatial modeling and mapping of stunting among children in Indonesia," *Geospatial Health,* vol. 19, no. 1, 2024. |
| [14] | G. Nduwayezu, A. Mansourian, J. P. Bizimana and P. Pilesjö, "Hybridizing spatial machine learning to explore the fine scale heterogeneity between stunting prevalence and its associated risk determinants in Rwanda," *Geo Spatial Information Science,* 2025. |
| [15] | X. Tang, J. Wang and Y. Liu, "An interpretable XGBoost model to predict stunting in children," *Frontiers in Pediatrics,* vol. 10, 2022. |
| [16] | T. D. Y. Y. N. Novalina, "Analisis Prediksi Status Stunting dengan Metode Support Vector Machine," *Bulletin of Electrical Engineering and Informatics,* vol. 14, no. 2, 2025. |
| [17] | R. A. Hasani, D. Ramadhan and M. Y. Permata, "Implementasi Algoritma Random Forest untuk Prediksi Stunting," *E3S Web of Conferences,* vol. 444, 2024. |
| [18] | A. Khalim, N. L. Sari and A. Widodo, "Decision Tree Regression untuk Prediksi Prevalensi Stunting," *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi),* 2023. |
| [19] | A. N. Widya, R. A. D. Winanda, R. F. R. P. Putri and D. L. Irawan, "Predictors Influencing Stunting on the Island of Java and Mapping Distribution Using Geo Maps & Orange Data Mining," *Economics Development Analysis Journal,* vol. 13, no. 2, 2024. |
| [20] | A. Nursikuwagus, N. R. Andani and T. Hartono, "Construct and Design Prediction System of Level Children Retardation in Certainty Factor Model," *Journal of Engineering Science and Technology,* vol. 16, no. 2, p. 1539–1546, 2021. |
| [21] | S. M. J. Rahman, A. Hossain, M. A. Islam, M. A. Habib, T. Moni and T. Rahman, "Investigate the Risk Factors of Stunting and Wasting Among Under Five Children in Bangladesh: A Machine Learning Approach," *PLoS ONE,* vol. 16, no. 6, 2021. |
| [22] | O. N. Chilyabanyama, R. Chilengi, M. Simuyandi and e. al., "Performance of Machine Learning Classifiers in Classifying Stunting Among Under-Five Children in Zambia," *Children,* vol. 9, no. 7, p. Article No. 1082, 2022. |
| [23] | K. Y. Nugroho, A. K. Dewi, M. Z. Abdillah and N. Suciati, "Machine Learning for Childhood Nutritional Status Classification," *Procedia Computer Science,* vol. 207, p. 1003–1010, 2023. |
| [24] | S. Ndagijimana, I. H. Kabano, E. Masabo and J. M. Ntaganda, "Prediction of Stunting Among Under 5 Children in Rwanda Using Machine Learning Techniques," *Journal of Preventive Medicine and Public Health,* vol. 56, no. 1, pp. 41-49, 2023. |
| [25] | H. Shen, H. Zhao and Y. Jiang, "Machine Learning Algorithms for Predicting Stunting among Under Five Children in Papua New Guinea," *Children,* vol. 10, no. 10, 2023. |
| [26] | J. Koima and e. al., "Machine Learning Algorithms for Predicting Stunting among Under Five Children in Papua New Guinea," *Children,* vol. 10, no. 10, p. 1638, 2023. |
| [27] | A. Nursikuwagus, H. Purwanto and D. A. Dewi, "Feature Classification Based on Heterogenous Data Using Hybrid Machine Learning: A Review," *Journal of Theoretical and Applied Information Technology,* vol. 101, no. 17, p. 6991–7003, 2023. |
| [28] | H. S. Mediani and e. al., "Predictors of Stunting Among Children Under Five Years of Age in Indonesia: A Scoping Review," *Global Journal of Health Science,* vol. 12, no. 8, p. 83–97, 2020. |
| [29] | S. K. Sari and R. E. Kurniasih, "Hubungan status gizi ibu dan faktor lingkungan dengan kejadian stunting pada balita," *Jurnal Gizi dan Kesehatan,* vol. 9, no. 2, p. 102–110, 2021. |

Berikut hasil Implementasi dan Dataset yang digunakan, berupa file Orange Workflow dan Data Excel terkait status stunting: <https://github.com/chandra10522024-code/Status_Stunting.git>