Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 6, Nomor: 1, Juni 2025

ANALISIS KOMPARATIF KLASIFIKASI MACHINE LEARNING UNTUK MEMPREDIKSI STUNTING PADA ANAK USIA DI BAWAH LIMA TAHUN

Yudhi Fajar Saputra[1], Mahmoud Ahmad Al-Khasawneh[2], Milkhatun[3], Ni Wayan Wiwin Asthiningsih[4], Sitti Rahmah[5]

[1,5] Universitas Widya Gama Mahakam Samarinda / Samarinda, Indonesia
[2] School of Computing, Skyline University College/ Sharjah - Uni Emirat Arab
[3,4] Universitas Muhammadiyah Kalimantan Timur/ Samarinda, Indonesia

[1]fajaryudhi@uwgm.ac.id, [2]mahmoud@outlook.my, [3]mil668@umkt.ac.id, [4]nww131@umkt.ac.id, [5]sitti_rahmah@uwgm.ac.id

ABSTRAK

Stunting merupakan salah satu permasalahan kesehatan masyarakat yang bisa berdampak jangka panjang terhadap kualitas sumber daya manusia di Indonesia. Deteksi dini terhadap status stunting anak usia di bawah lima tahun menjadi langkah dalam mencegah gangguan pertumbuhan kronis akibat stunting, sehingga penelitian ini bertujuan untuk membangun model klasifikasi status stunting dengan memanfaatkan pendekatan data mining menggunakan algoritma Decision Tree dan Random Forest. Data yang digunakan diperoleh dari hasil survei terhadap ibu yang memiliki anak dibawah umur lima tahun dengan sejumlah 193 responden, data tersebut mencakup variabel antropometri dan sosial ekonomi, seperti tinggi badan, berat badan, usia anak, pendidikan orang tua, pendapatan keluarga, dan urutan kelahiran. data tersebut diproses melalui tahapan Knowledge Discovery in Databases (KDD) meliputi seleksi atribut, imputasi, encoding, dan klasifikasi melalui proses permodelan data mining, selanjutnya evaluasi dilakukan dengan metrik klasifikasi Classification Accuracy(CA) dan Area Under the Curve (AUC) dari kurva Receiver Operating Characteristic (ROC). Hasil penelitian menunjukkan bahwa model Random Forest memiliki performa lebih baik dibandingkan Decision Tree dengan nilai CA 71% dan AUC 0.74. dibandingkan Decision Tree dengan nilai CA 67% dan AUC 0.68. Peneliti berharap bahwa Model prdiksi ini berpotensial dapat digunakan sebagai sistem deteksi dini stunting berbasis data atau sebagai rujukan untuk penelitian berikutnya

***Kata Kunci*—Stunting, Machine Learning, Random Forest, Decision Tree, Classification Model, ROC Curve.**

ABSTRACT

Stunting is one of the public health issues that can have long-term impacts on the quality of human resources in Indonesia. Early detection of stunting status among children under five years of age is a critical step in preventing chronic growth disorders. Therefore, this study aims to develop a classification model for stunting status using a data mining approach with Decision Tree and Random Forest algorithms. The dataset was obtained from a survey of 193 mothers with children under five, encompassing anthropometric and socioeconomic variables such as height, weight, child's age, parental education, family income, and birth order. The data were processed through the stages of Knowledge Discovery in Databases (KDD), including attribute selection, imputation, encoding, and classification modeling. The model performance was evaluated using classification metrics: Classification Accuracy (CA) and the Area Under the Curve (AUC) from the Receiver Operating Characteristic (ROC) curve. The results show that the Random Forest model outperformed the Decision Tree, achieving a CA of 71% and an AUC of 0.74, compared to the Decision Tree with a CA of 67% and an AUC of 0.68. This predictive model is expected to be useful as a data-driven early detection system for stunting or serve as a reference for future research.

Keywords—Stunting, Machine Learning, Random Forest, Decision Tree, Classification Model, ROC Curve.

Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 6, Nomor: 1, Juni 2025

I. PENDAHULUAN

Stunting terus menjadi masalah kesehatan masyarakat yang signifikan di seluruh dunia, terutama di negara-negara berkembang, termasuk Indonesia. Berdasarkan laporan World Health Organization (WHO) Global Health Observatory, prevalensi stunting pada anak di bawah lima tahun global menurun dari 26,3 % pada tahun 2012 menjadi 22,3 %[1], namun masih berdampak terhadap 148,1 juta anak di dunia pada tahun 2021[2].

Di tingkat nasional, Indonesia juga menghadapi tantangan besar terkait stunting. Data dari Nutritional Status of Indonesian Toddlers survey tahun 2021 menunjukkan bahwa prevalensi stunting nasional berada pada angka 24,4 %, yaitu setara dengan ±5,33 juta anak usia bawah lima tahun termasuk di samarinda[3]. Meskipun terjadi penurunan dari rerata 36,4 % antara tahun 2005–2017, masalah ini masih mengakar kuat. Selain itu, estimasi dari World Bank mencatat bahwa prevalensi stunting turun dari 41,5 % pada tahun 2000 menjadi 31,8 % pada tahun 2020[4]. Data terbaru dari Kementerian Kesehatan Republik Indonesia mencatat bahwa pada tahun 2022, prevalensi stunting lanjut turun menjadi 21,6 %[5]. Fakta-fakta tersebut menegaskan bahwa stunting tetap menjadi krisis kesehatan yang serius di Indonesia, meskipun telah mengalami penurunan.

Berbagai faktor determinan seperti kemiskinan, pola makan tidak seimbang, sanitasi buruk, serta kurangnya akses layanan kesehatan dan pendidikan gizi memperkuat urgensi penanganan berbasis data yang akurat dan berkelanjutan [6]. Oleh karena itu, dibutuhkan sistem pendukung keputusan yang mampu mengidentifikasi kelompok berisiko stunting secara dini dan tepat sasaran

Eka Pratama et al. membandingkan beberapa algoritma machine learning pada dataset stunting nasional dengan pendekatan regresi, dan menemukan bahwa Random Forest Regression memberikan performa terbaik dimana penelitian ini menggunakan data prevalensi stunting di seluruh provinsi Indonesia dan mendapatkan keakuratan tinggi untuk prediksi geografis[7]. Penelitan dari Haris et al. Juga menerapkan algoritma Random Forest pada data stunting Provinsi Jawa Timur (n = 38), meskipun datanya relatif kecil, model tersebut mencapai nilai MAE sebesar 1,02 dan MSE sebesar 1,64[8]. Penelitian di tingkat internasional seperti di Bangladesh menggunakan data dari Demographic and Health Survey (DHS)–Bangladesh tahun 2014 (n = 7.886) menggunakan algoritma klassifikasi dengan hasil menunjukkan bahwa Random Forest mencapai akurasi 70,1 %–72,4 % dan nilai AUC sekitar 0,70 untuk klasifikasi stunting dan underweight[9].

Penelitian lain, seperti oleh Saragih et al. di Sumatera Utara, juga menggunakan Random Forest dalam memprediksi stunting dengan memberikan hasil error yang lebih rendah dibandingkan metode lain[10]. Selain itu, Juwariyem et al. berhasil menggunakan gabungan bagging + Random Forest pada dataset stunting balita (n $\approx$ 10.000), menghasilkan akurasi hingga 91,98 %, dengan nilai presisi dan recall yang tinggi untuk prediksi stunting dan non-stunting[11].

Disisi lain algoritma Decision tree dalam penelitian permodelan prediksi stunting juga dimplementasikan oleh Restu et al. (2023) yang menerapkan algoritma Decision Tree C4.5 pada data balita yang bersumber dari Puskesmas di Provinsi Lampung, Indonesia. Penelitian ini menunjukkan bahwa Decision Tree menghasilkan akurasi sebesar 81,5 % dalam mengklasifikasikan status gizi balita (stunting, normal, overweight)[12].

Selanjutnya, Kusuma et al. (2022) juga menggunakan algoritma Decision Tree CART dalam memprediksi status stunting berdasarkan data sosial ekonomi keluarga dan indikator kesehatan anak. Model tersebut menghasilkan akurasi sebesar 78,2 %[13]. Sementara itu, penelitian lain yang dilakukan oleh Sartika et al. di Jakarta menggunakan algoritma J48 (turunan C4.5) dalam sistem klasifikasi status gizi. Hasil penelitainnya menunjukkan bahwa Decision Tree J48 mampu mengidentifikasi pola yang memiliki hubungan dalam data dengan tingkat akurasi 83,1 %[14].

Namun demikian, penelitian-penelitian sebelumnya masih terbatas pada wilayah tertentu, variabel terbatas, serta kurang mengeksplorasi perbandingan kinerja antara Decision Tree dan Random Forest pada dataset yang lebih komprehensif.

Dengan prevalensi stunting yang masih tinggi di Indonesia, maka perlu dibutuhkan sebuah sistem deteksi dini berbasis data, maka dari itu penelitian ini bertujuan membangun dan membandingkan model klasifikasi status stunting pada anak balita menggunakan algoritma Decision Tree dan Random Forest, untuk mendapatkan model yang paling optimal sebagai alat bantu deteksi dini stunting. Hasil dari penelitian ini diharapkan dapat menjadi rekomendasi dalam membuat sistem atau alat bantu dalam mendeteksi dini yang lebih terarah dan efektif dalam menurunkan prevalensi stunting di Indonesia dan menjadi rujukan penelitian yang relevan dimasa mendatang

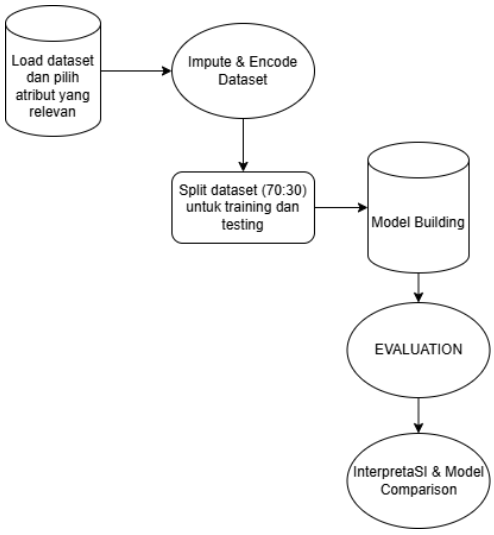
Meskipun beberapa studi menunjukkan performa Decision Tree yang cukup tinggi, sebagian besar penelitian menyatakan bahwa Random Forest memberikan performa yang lebih stabil dan akurat, terutama pada dataset yang besar dan kompleks[11]. Dengan demikian, baik Decision Tree maupun Random

Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 6, Nomor: 1, Juni 2025

Forest telah digunakan secara luas dalam penelitian terkait prediksi stunting, Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model klasifikasi status stunting pada anak di bawah lima tahun menggunakan algoritma klassifikasi machine learning khususnya decision tree dan random forest. Model ini diharapkan dapat menjadi alat bantu dalam mendeteksi dini dan merancang intervensi yang lebih tepat sasaran guna mempercepat upaya penurunan prevalensi stunting di Indonesia.

II. METODE PENELTIAN

Penelitian yang dilakukan merupakan penelitian *kuantitatif eksploratif* dengan pendekatan *data mining*, yang bertujuan membangun model prediktif klasifikasi status *stunting* pada anak usia di bawah lima tahun menggunakan algoritma *Decision Tree* dan *Random Forest*. Pendekatan data mining pada penelitian ini mengacu pada proses Knowledge Discovery in Databases (KDD). KDD merupakan suatu proses sistematis untuk mengekstraksi pengetahuan dari data, yang terdiri atas beberapa tahapan mulai dari pemilihan data hingga evaluasi hasil. Proses KDD digunakan secara luas dalam analisis data skala besar karena mendukung alur kerja yang terstruktur dan berulang [15],[16]. Proses utama dari KDD dalam penelitian ini meliputi *Selection* (Pemilihan Data), Pre-*processing* (Pra-pemrosesan Data), *transformation* (Transformasi Data), *Data Mining*, dan terakhir adalah *Interpretation/Evaluation* (Evaluasi dan Interpretasi).



Gambar 1. Workflow penelitian berdasarkan KDD

A. Pemilihan Data

Tahap ini bertujuan untuk memilih subset data yang relevan dari keseluruhan data yang tersedia. Dalam konteks penelitian ini, data yang dipilih meliputi karakteristik sosial-ekonomi keluarga dan data antropometri anak, termasuk tinggi badan, berat badan, usia, jenis kelamin, serta status stunting. Data-data tersebut akan dirangkum di bawah ini (Tabel 1). Pemilihan variabel pada peneltiasn ini mengacu pada kerangka konseptual UNICEF tentang gizi anak-anak [17]. Target Variabel dalam penelitian ini adalah status stunting yang diklasifikasikan berdasarkan kriteria dari WHO [18].

Tabel I
Atribut Dataset

Attribut	Description	Categories
Variabel terkait karakteristik anak		
Usia bayi	Usia bayi dalam bulan	1: 0 < 6, 2: 6 – 11, 3: 12 - 23 4: 24 - 35, 5: 36 - 47, 6: 48 - 59
Gender	Jenis kelamin pada bayi	1: Wanita 2: Laki-laki
Urutan bayi	Urutan kelahiran bayi	Numeric
Birthweight	Berat badan bayi saat lahir	Numeric
Birthsize	Ukuran bayi saat lahir	Numeric
Variabel terkait sosio-ekonomi		
Maternal age	Usia dari ibu yang melahirkan	1: < 18 th, 2: 19-35 th, 3: > 35 th,
Maternal Profession	Profesi ibu yang melahirkan	1: PNS, 2: Wira Swasta, 3: Pelajar 4: Lainnya
Maternal education	Pendidikan terakhir ibu yang melahirkan	1: SD, 2: SMP, 3: SMA, 4: S1, 5: S2
Marital status	Status dari ibu yang melahirkan	1: single, 2: Menikah, 3. Janda.
Maternal Income	Pendapatan keluarga dari ibu yang melahirkan	1: < 1 juta, 2: Rp 1.000.000 ≤ X ≤ Rp 2.500.000, 3: Rp 2.500.000 < X ≤ Rp 4.000.000, 4: Rp 4.000.000 < X ≤ Rp 6.000.000 5: Rp 6.000.000 < X ≤ Rp 10.000.000 6: > Rp 10.000.000
Status Stunting	Status ketertangan stunting dari anak	1: Sangat stunting 2: Stunting 3: Normal 4: Tinggi 5: Sangat Tinggi

Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 6, Nomor: 1, Juni 2025

B. Pra-pemrosesan Data

Prapemrosesan data adalah proses menyiapkan (membersihkan dan menyiapkan) data mentah sehingga dapat dipahami dan digunakan untuk analisis. Prapemrosesan data terdiri dari berbagai tahap, termasuk seleksi atribut, penanganan nilai kososng (Imputasi), Penyesuaian dan pengkategorian Atribut (encoding) [19], [20].

1) *Seleksi atribut:* atribut-atribut yang digunakan sebagai fitur (input) dalam proses klasifikasi antara lain: (1). *Maternal education*, (2). *Maternal Profession*, (3). *Maternal Income*, (5). *Birthsize*, (5). *Birthweight*, (6). Usia bayi, (7). Urutan anak, (8). *Gender*, Sementara itu, atribut target (output) yang digunakan adalah "Status stunting". Atribut identitas seperti ID_anak dan martial status tidak disertakan dalam pemodelan karena tidak relevan secara analitis.

2) *Imputasi:* Langkah berikutnya adalah proses imputasi untuk menangani nilai kosong (missing values). Atribut numerik yang memiliki nilai kosong diisi menggunakan nilai rata-rata (mean), sedangkann atribut kategorikal diisi menggunakan nilai modus (most frequent).

C. Transformasi Data

Setelah data melalui tahap pra-pemrosesan, langkah selanjutnya adalah transformasi data, transformasi data merupakan proses dalam analisis data mining yang bertujuan menyesuaikan format data dengan kebutuhan algoritma klasifikasi. Atribut kategorikal seperti Pendidikan orang tua, Pekerjaan, dan Jenis Kelamin dikodekan secara otomatis menjadi representasi numerik dalam Orange Data Mining melalui teknik label encoding[21], [22].

Selanjutnya, atribut target dikonversi dari klasifikasi multikelas menjadi empat kelas utama melalui label binarisasi untuk bisa diterapkan pada evaluasi berbasis ROC Curve dan AUC[23].

Adapun atribut numerik seperti Tinggi badan, dan Berat badan, digunakan tanpa transformasi karena algoritma Decision Tree dan Random Forest tidak sensitif terhadap skala data, dan tidak memerlukan proses normalisasi atau standardisasi seperti halnya algoritma berbasis jarak (misalnya k-NN atau SVM) [24], [25].

D. Data Mining

Data mining merupakan inti dari proses Knowledge Discovery in Databases (KDD), yang bertujuan untuk mengekstraksi pola, relasi, atau model yang bisa dikembangkan dari data yang telah diproses. Pada proses ini, algoritma machine learning digunakan untuk membangun model prediktif berdasarkan atribut-atribut input terhadap target output yang telah ditentukan sebelumnya[23], [22].

Dalam penelitian ini, proses data mining difokuskan pada pembangunan model klasifikasi status stunting menggunakan dua algoritma berbasis pohon keputusan, yaitu *Decision Tree* dan *Random Forest*. Kedua algoritma ini dipilih karena kemampuannya dalam menangani data campuran (numerik dan kategorikal), interpretabilitas yang tinggi, serta performa yang baik pada data berskala kecil hingga sedang [21], [26].

1) *Algoritma Decision Tree:* Decision Tree adalah algoritma klasifikasi yang menggunakan struktur pohon untuk memodelkan keputusan berdasarkan atribut-atribut input. Proses pemisahan (splitting) pada tiap node dilakukan berdasarkan metrik impuritas seperti Gini Index atau Entropy (Information Gain). Model ini menghasilkan representasi logika yang mudah diinterpretasi, menjadikannya populer dalam konteks pengambilan keputusan di bidang kesehatan masyarakat[25]. Pada penelitian ini, algoritma Decision Tree digunakan sebagai baseline model, di mana struktur pohon menghasilkan aturan klasifikasi yang secara langsung dapat mengelompokkan status stunting berdasarkan variabel-variabel seperti tinggi badan, usia anak, dan status ekonomi keluarga.

2) *Algoritma Random Forest:* *Random Forest* adalah pengembangan dari *Decision Tree* yang menggunakan pendekatan *ensemble learning*, yaitu membangun sejumlah pohon keputusan (trees) secara acak dan menggabungkan hasilnya melalui mekanisme voting mayoritas. Teknik ini secara signifikan dapat meningkatkan akurasi dan mengurangi risiko *overfitting* [24]. Dalam penelitian ini, *Random Forest* digunakan untuk mengevaluasi sejauh mana model klasifikasi dapat ditingkatkan dibandingkan Decision Tree tunggal. Parameter default dari Orange Data Mining digunakan, dengan opsi tuning seperti jumlah pohon (n_estimators) dan maksimum kedalaman (max_depth) dieksplorasi jika diperlukan.

Gambar 1. Komparasi antara decision tree dengan random forest

E. Evaluasi dan Interpretasi

Evaluasi dan interpretasi merupakan tahap akhir dalam proses Knowledge Discovery in Databases (KDD), yang bertujuan untuk menilai performa model yang telah dibangun serta menafsirkan hasilnya dalam konteks tujuan penelitian. Evaluasi yang baik

Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 6, Nomor: 1, Juni 2025

memungkinkan peneliti untuk memahami seberapa akurat model dalam memprediksi kelas target, serta untuk membandingkan kinerja antar algoritma yang digunakan[23].

Dalam penelitian ini, evaluasi dilakukan terhadap dua model klasifikasi, yaitu Decision Tree dan Random Forest, menggunakan grafik Receiver Operating Characteristic (ROC) dan nilai Area Under the Curve (AUC). Receiver Operating Characteristic (ROC) curve yang memvisualisasikan hubungan antara True Positive Rate (TPR) dan False Positive Rate (FPR) pada berbagai nilai threshold klasifikasi. Semakin dekat kurva ke sudut kiri atas, semakin baik performa model[27]. Sedangkan nilai Area Under the Curve (AUC) memberikan ukuran kuantitatif dari kualitas model; semakin mendekati nilai 1, semakin baik kemampuan model dalam membedakan antar variable target stunting.

III. HASIL DAN PEMBAHASAN

Beberapa item yang akan dipaparkan pada bab hasil meliputi deskripsi data, hasil pra-pemrosesan dan transformasi, hasil pemodelan dan evaluasi, dan perbandingan model *Decision Tree* vs *Random Forest*.

A. Deskripsi data

Dataset yang digunakan dalam penelitian ini terdiri dari 119 entri data anak usia di bawah lima tahun yang sudah dilakukan proses cleaning data, dimana responden pada penelitian ini berjumlah 192 responden yang mencakup variabel sosial-ekonomi orang tua, data antropometri anak, serta status stunting sebagai target klasifikasi. Untuk memperoleh gambaran awal terhadap struktur data, dilakukan analisis deskriptif terhadap beberapa kategori penting dan atribut numerik. Data deskripsi data disajikan pada tabel berikut

1) Distribusi Status Stunting: Status stunting dibagi menjadi beberapa kategori berdasarkan klasifikasi WHO, yaitu Sangat Stunting, Stunting, Normal, dan Tinggi. Sebagian besar anak dalam dataset berada pada kategori Normal, sementara sebagian lainnya mengalami Stunting atau Sangat Stunting. Detail jumlah per kategori disajikan dalam Tabel Distribusi Status Stunting.

Tabel 2
Deskripsi Status Stunting

No	Status stunting	Jumlah
1	Normal	79
2	Tinggi	18
3	Stunting	13
4	Sangat stunting	9

2) *Statistik Deskriptif Atribut:* empat atribut numerik yaitu Pendapatan orang tua, Tinggi badan, Berat badan, dan Usia anak disajikan pada tabel 3.

Tabel 3
Deskripsi Atribut Numerik

No	Atribut	Mean	Std
1	Pendapatan orang tua	3513903	2287246
2	Tinggi badan	85	24.14
3	Berat badan	12	9.38
4	Usia Anak	24	20.02

Berdasarkan Tabel Statistik Deskriptif Atribut Numerik, dapat diketahui bahwa:

- Pendapatan orang tua berkisar dari Rp 0 hingga Rp10.000.000, dengan rata-rata sebesar Rp3.513.903.
- Tinggi badan anak berkisar dari 7 cm hingga 175 cm, dengan rata-rata sekitar 85,7 cm.
- Berat badan anak memiliki nilai minimum 2 kg dan maksimum 75 kg, dengan median 10 kg.
- Usia anak dalam data berkisar antara 1 hingga 60 bulan, dengan rata-rata sekitar 24,5 bulan.

B. Hasil Pemodelan dan Evaluasi

Setelah data melalui tahap pra-pemrosesan dan transformasi, langkah berikutnya adalah membangun model klasifikasi untuk memprediksi status stunting pada anak usia di bawah lima tahun. Pada penelitian ini, digunakan dua algoritma berbasis pohon keputusan, yaitu *Decision Tree* dan *Random Forest*.

Evaluasi model dilakukan menggunakan metrik Area Under the Curve (AUC) dari grafik Receiver Operating Characteristic (ROC). Berikut adalah Tabel Perbandingan Evaluasi Model yang menyajikan performa antara model *Decision Tree* dan *Random Forest* berdasarkan metrik AUC.

Tabel 4
Tabel Perbandingan Evaluasi Model

No	Model Prediksi	Metrik AUC	CA (Classification Accuracy)
1	Decision Tree	0.67	0,68
2	Random Forest	0.74	0.71

Hasil ini menunjukkan bahwa Random Forest memiliki performa lebih baik dibandingkan Decision Tree, dengan nilai AUC sebesar 0.74 dan akurasi klasifikasi sebesar 71%.

Dengan analisis lebih lanjut terhadap model Random Forest mengungkap bahwa fitur-fitur seperti tinggi badan anak, usia, status ekonomi keluarga, dan berat badan merupakan variabel yang juga berkontribusi terhadap klasifikasi status stunting. Temuan ini lebih menguatkan bahwa faktor

Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 6, Nomor: 1, Juni 2025

antropometri dan sosial ekonomi adalah determinan utama risiko stunting.

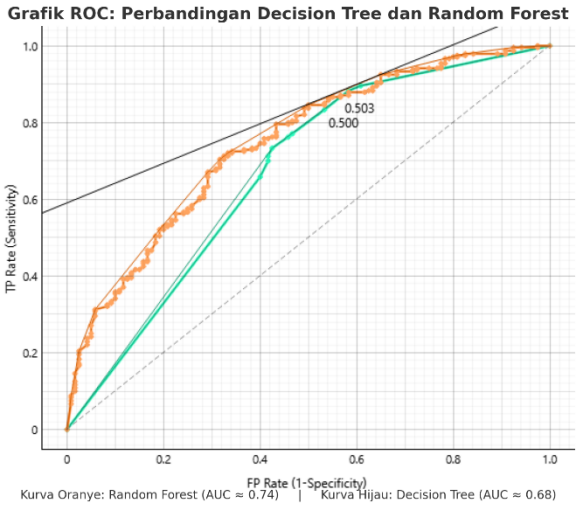
Penelitian selanjutnya disarankan untuk memperluas cakupan data, baik secara geografis maupun varaibel yang lebih luas tidak hanya *socio-economy*, serta mengeksplorasi teknik ensemble lanjutan atau deep learning ringan guna meningkatkan akurasi dan generalisasi model

Penelitian ini menghasilkan dua model klasifikasi berbasis pohon keputusan, yakni *Decision Tree* dan *Random Forest*, yang digunakan untuk memprediksi status stunting pada anak usia di bawah lima tahun. Berdasarkan hasil evaluasi yang telah disajikan sebelumnya, dapat diinterpretasikan bahwa kedua model mampu melakukan klasifikasi dengan tingkat akurasi yang cukup baik, namun menunjukkan perbedaan signifikan dalam hal stabilitas dan performa keseluruhan.

Model *Decision Tree* memperoleh nilai akurasi sebesar 68%, dengan AUC sebesar 0.67. Meskipun interpretasi model ini sangat jelas dan transparan karena dapat divisualisasikan dalam bentuk pohon keputusan, model ini cenderung mengalami overfitting terhadap data latih. Hal ini terlihat dari performanya yang tidak terlalu tinggi pada metrik evaluasi *classification accuracy*, Overfitting umum terjadi pada struktur pohon yang terlalu dalam, terutama ketika data pelatihan terbatas dan tidak representatif secara keseluruhan[25].

Sementara itu, model *Random Forest* menunjukkan kinerja yang lebih unggul, dengan akurasi sebesar 71% dan AUC sebesar 0.74. Keunggulan model ini terletak pada kemampuannya menggabungkan banyak pohon keputusan (ensemble learning), yang mampu mengurangi variansi dan meningkatkan generalisasi terhadap data uji[24]. Nilai *classification accuracy* yang lebih tinggi menunjukkan bahwa *Random Forest* lebih andal dalam mengidentifikasi anak-anak yang mengalami stunting dan mengurangi kesalahan klasifikasi (baik false positive maupun false negative).

Gambar 3 menampilkan kurva *Receiver Operating Characteristic* (ROC) yang dihasilkan dari evaluasi model Decision Tree dan Random Forest terhadap klasifikasi status stunting. Grafik ini merupakan alat visual yang umum digunakan untuk mengevaluasi performa model klasifikasi biner pada berbagai nilai ambang batas (threshold). Sumbu vertikal (True Positive Rate / Sensitivitas) diplot terhadap sumbu horizontal (False Positive Rate / 1 – Spesifisitas).



Gambar 3. Kurva Receiver Operating Characteristic (ROC) dari evaluasi model Decision Tree dan Random Forest

Dalam grafik tersebut, terdapat dua kurva ROC:

- Kurva berwarna oranye menunjukkan performa dari model Random Forest,
- Kurva berwarna hijau menunjukkan performa dari model Decision Tree.

Berdasarkan grafik tersebut, dapat dilihat bahwa kurva Random Forest cenderung lebih mendekati pojok kiri atas, yang mengindikasikan tingkat sensitivitas yang lebih tinggi pada berbagai threshold. Sementara itu, kurva Decision Tree berada di bawahnya, memperlihatkan bahwa model ini memiliki kemampuan klasifikasi yang lebih rendah secara keseluruhan.

Perbandingan antara kedua model ini juga diperkuat oleh nilai Area Under the Curve (AUC) yang diperoleh dari masing-masing model:

AUC Random Forest $\approx$ 0.74

AUC Decision Tree $\approx$ 0.67

Nilai AUC yang lebih tinggi pada Random Forest menunjukkan bahwa model ini lebih efektif dalam membedakan antara anak yang mengalami stunting dan yang tidak. Hal ini sesuai dengan karakteristik algoritma Random Forest sebagai metode ensemble yang lebih tahan terhadap overfitting dan memiliki kemampuan generalisasi yang lebih baik dibandingkan pohon keputusan tunggal seperti Decision Tree[24].

Dengan demikian, berdasarkan analisis kurva ROC dan AUC, Random Forest merupakan algoritma yang lebih direkomendasikan untuk digunakan dalam klasifikasi status stunting dalam penelitian ini, terutama ketika akurasi prediksi dan sensitivitas deteksi dini menjadi prioritas utama.

Secara keseluruhan, interpretasi hasil ini menunjukkan bahwa Random Forest lebih layak digunakan sebagai model klasifikasi untuk prediksi status stunting, terutama karena kombinasi antara

akurasi tinggi dan ketahanan terhadap overfitting. Namun demikian, model Decision Tree tetap memberikan nilai tambah dalam hal interpretasi yang lebih mudah, khususnya untuk keperluan komunikasi hasil dengan pemangku kebijakan di tingkat layanan primer seperti puskesmas atau kader posyandu.

Maka dari itu, penelitian ini memiliki tujuan untuk mengembangkan dan mengevaluasi model klasifikasi status stunting pada anak usia di bawah lima tahun menggunakan pendekatan data mining berbasis algoritma *Decision Tree* dan *Random Forest*. Berdasarkan hasil evaluasi model yang telah dilakukan, dapat disimpulkan bahwa tujuan penelitian ini telah tercapai dengan baik.

Model klasifikasi yang dibangun mampu mengelompokkan status stunting anak secara otomatis berdasarkan sejumlah atribut masukan, seperti tinggi badan, usia anak, pendapatan orang tua, serta latar belakang pendidikan dan pekerjaan orang tua. Kemampuan model dalam melakukan klasifikasi dibuktikan melalui nilai metrik evaluasi yang cukup tinggi, terutama pada model Random Forest yang memperoleh akurasi sebesar 71% dan AUC sebesar 0.74, menunjukkan performa yang baik dalam membedakan antara kasus stunting dan tidak stunting.

Hasil ini menunjukkan bahwa metode klasifikasi yang diterapkan dalam penelitian ini dapat mendukung identifikasi dini anak dengan risiko stunting, yang sejalan dengan tujuan utama penelitian, yaitu menghasilkan sistem pendukung keputusan berbasis data untuk mendeteksi status stunting. Model ini diharapkan dapat berfungsi sebagai alat bantu bagi tenaga kesehatan atau pengambil kebijakan dalam melakukan skrining awal terhadap kelompok balita yang berisiko tinggi mengalami masalah pertumbuhan.

Keterkaitan hasil ini juga relevan dengan rekomendasi dari World Health Organization (WHO), yang menyebutkan bahwa deteksi dini dan pemantauan pertumbuhan adalah komponen penting dalam pencegahan stunting, terutama pada periode 1.000 Hari Pertama Kehidupan (HPK)[28]. WHO menggarisbawahi bahwa stunting bukan hanya masalah antropometri, tetapi mencerminkan kondisi kronis akibat kurangnya asupan gizi, kesehatan lingkungan yang buruk, serta keterbatasan akses terhadap layanan kesehatan[29] .

Lebih jauh, arah penelitian ini juga mendukung agenda pemerintah Indonesia dalam menurunkan prevalensi stunting secara nasional. Melalui Peraturan Presiden Republik Indonesia Nomor 72 Tahun 2021 tentang Percepatan Penurunan Stunting, ditegaskan pentingnya penggunaan data dan teknologi untuk mendukung pemantauan dan perencanaan intervensi berbasis bukti[30]. Selain itu, dalam dokumen Rencana Aksi Nasional Percepatan Penurunan Stunting Indonesia (RAN-PASTI) 2021–2024, penguatan sistem informasi dan inovasi teknologi berbasis kecerdasan buatan menjadi salah satu strategi pendukung utama[31].

Dengan demikian, keseluruhan proses, mulai dari eksplorasi data, pembangunan model klasifikasi, hingga evaluasi performa, telah mendukung pencapaian tujuan penelitian, baik dari sisi keilmuan maupun kontribusi praktis di bidang kesehatan masyarakat.

KESIMPULAN

Penelitian ini bertujuan untuk mengembangkan model klasifikasi status stunting pada anak usia di bawah lima tahun menggunakan algoritma Decision Tree dan Random Forest berbasis pendekatan data mining. Melalui proses Knowledge Discovery in Databases (KDD), data dianalisis secara sistematis melalui tahapan seleksi atribut, pra-pemrosesan, transformasi data, pemodelan, dan evaluasi performa model.

Hasil penelitian menunjukkan bahwa kedua algoritma mampu melakukan klasifikasi status stunting dengan tingkat akurasi yang memadai. Model *Decision Tree* menghasilkan akurasi sebesar 68% dan AUC sebesar 0.67, sedangkan model Random Forest menunjukkan performa yang lebih unggul dengan akurasi 71% dan AUC sebesar 0.74. Analisis kurva ROC mengonfirmasi bahwa Random Forest memiliki sensitivitas dan spesifisitas yang lebih baik, menjadikannya pilihan model yang lebih andal untuk klasifikasi status stunting.

Dengan demikian, model klasifikasi yang dikembangkan dalam penelitian ini tidak hanya memenuhi tujuan keilmuan, tetapi juga memiliki potensi untuk dimanfaatkan sebagai alat bantu dalam sistem deteksi dini stunting di tingkat layanan kesehatan dasar. Penelitian ini memberikan kontribusi terhadap integrasi kecerdasan buatan dalam sistem informasi kesehatan, yang sejalan dengan agenda nasional percepatan penurunan stunting di Indonesia.

REFERENSI

[1] W. H. O. (WHO), “Child malnutrition: Stunting among children under 5 years of age,” 2024.
[2] UNICEF, WHO, and W. Bank, “Joint child malnutrition estimates 2021,” 2024.
[3] Unknown, “Parental stature..., 24.4% stunting in 2021,” *Asia Pac J Clin Nutr*, 2018.
[4] W. Bank, “Stunting prevalence in Indonesia 2000–2020,” 2020.
[5] K. et al., “Indonesia has the second highest stunting rate…,” *ScienceDirect*, 2024.
[6] Wikipedia, “Stunted growth,” 2025.

[7] A. Pratama and others, "Comparison of Machine Learning Algorithms for Predicting Stunting Prevalence in Indonesia," SISFOKOM J. Inf. Comput. Syst., vol. 13, no. 2, pp. 200–209, Jun. 2024.
[8] M. S. Haris, M. Anshori, and A. N. Khudori, "Prediction of Stunting Prevalence in East Java Province with Random Forest Algorithm," J. Teknol. Inf. (JUTIF), vol. 4, no. 1, pp. 11–13, Feb. 2023.
[9] S. A. Hemo and M. I. Rayhan, "Classification tree and Random Forest model to predict under-five malnutrition in Bangladesh," Biom. Biostat. Int. J., vol. 10, no. 3, pp. 116–123, 2021.
[10] V. R. Saragih and others, "Comparative analysis of supervised machine learning methods in predicting stunting in North Sumatra," J. OSCExp., Jun. 2025.
[11] Juwariyem, S. Sriyanto, S. Lestari, and C. Chairani, "Prediction of Stunting in Toddlers Using Bagging and Random Forest Algorithms," Sinkron, vol. 8, no. 2, pp. 947–956, Apr. 2024.
[12] A. D. Restu, Y. S. Putri, and R. P. Yanti, "Penerapan Algoritma Decision Tree C4.5 untuk Klasifikasi Status Gizi Balita," Jurnal Media Informatika Budidarma, vol. 7, no. 3, pp. 1567–1575, Sep. 2023.
[13] F. Kusuma, M. Izzati, and S. Rahmawati, "Prediksi Status Stunting Menggunakan Decision Tree CART," J. Inform. Kesehatan, vol. 4, no. 1, pp. 25–32, Jan. 2022.
[14] R. Sartika and M. A. Hakim, "Penerapan Decision Tree J48 dalam Sistem Klasifikasi Gizi Anak," Jurnal Informatika Medis Indonesia, vol. 9, no. 2, pp. 73–80, Dec. 2021.
[15] U. M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," AI Mag, vol. 17, no. 3, pp. 37–54, 1996.
[16] M. Fayyad, "The KDD process for extracting useful knowledge from volumes of data," Commun ACM, vol. 39, no. 11, pp. 27–34, 1996.
[17] UNICEF, Conceptual Framework Child Nutrition. New York, NY 10017, USA, 2021.
[18] W. H. Organization, Guideline: assessing and managing children at primary healthcare facilities to prevent overweight and obesity in the context of the double burden of malnutrition. New York, 2017.
[19] K. Al-Jabery, T. Obafemi-Ajayi, G. Olbricht, and others, Computational Learning Approaches to Data Analytics in Biomedical Applications. Academic Press, 2019.
[20] U. Pujianto, A. P. Wibawa, M. I. Akbar, and others, "K-nearest neighbour (k-nn) based missing data imputation," in 2019 5th International Conference on Science in Information Technology (ICSITech), IEEE, 2019, pp. 83–88.
[21] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, 4th ed. Burlington, MA: Morgan Kaufmann, 2016.
[22] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. San Francisco: Morgan Kaufmann, 2011.
[23] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases," AI Mag, vol. 17, no. 3, pp. 37–54, 1996.
[24] L. Breiman, "Random Forests," Mach Learn, vol. 45, no. 1, pp. 5–32, 2001.
[25] R. Quinlan, "Induction of decision trees," Mach Learn, vol. 1, no. 1, pp. 81–106, 1986.
[26] P. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining. Boston: Addison-Wesley, 2006.
[27] T. Fawcett, "An introduction to ROC analysis," Pattern Recognit Lett, vol. 27, no. 8, pp. 861–874, 2006.
[28] W. H. Organization, Essential Nutrition Actions: Mainstreaming Nutrition Through the Life-Course. Geneva: World Health Organization, 2019.
[29] W. H. Organization, Child Growth Standards: Length/Height-for-Age. Geneva: World Health Organization, 2006.
[30] P. R. Indonesia, "Peraturan Presiden Nomor 72 Tahun 2021 tentang Percepatan Penurunan Stunting," 2021, Jakarta, Indonesia.
[31] S. W. P. RI, "Rencana Aksi Nasional Percepatan Penurunan Stunting Indonesia (RAN-PASTI) 2021–2024," 2021, Jakarta.