Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1202−1211
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10531

Optimalisasi Hyperparameter Random Forest Menggunakan Random Search untuk Klasifikasi Risiko Stunting Pada Balita

Putri Windari*, Khairunnas, Irma Eryanti Putri

Fakultas Teknik dan Ilmu Komputer, Ilmu Komputer, Universitas Muhammadiyah Bima, Bima
Jln. Anggrek No.16, Ranggo Na'e, Kota Bima, Nusa Tenggara Barat, Indonesia
Email: [1,*]putrywindari@gmail.com, [2]assignment.khairunnas@gmail.com, [3]irerput1802@gmail.com
Email Penulis Korespondensi: putrywindari@gmail.com
Submitted: **30/06/2026;** Accepted: **21/07/2026;** Published: **21/07/2026**

Abstrak−Stunting merupakan kondisi gagal tumbuh anak yang terjadi akibat kurangnya asupan gizi kronis dalam jangka waktu yang lama. Kondisi ini tidak hanya menghambat pertumbuhan fisik anak, tetapi juga menghambat kemampuan belajar, serta meningkatkan risiko berbagai penyakit dimasa depan. Tingginya angka stunting di Kota Bima menunjukkan perlunya pemanfaatan metode Machine Learning untuk membantu menganalisis faktor risiko stunting dengan lebih akurat. Penelitian ini bertujuan mengoptimalkan kinerja algoritma Random Forest menggunakan metode Random Search dalam mengklasifikasikan risiko stunting pada balita di Kelurahan Sambinae Kota Bima. Dataset yang digunakan 1.162 data balita dengan 12 atribut yang diperoleh dari Puskesmas Mpunda Kota Bima. Tahapan penelitian meliputi pengumpulan data, preprocessing data, pembagian data, pembangunan model Random Forest sebagai model baseline, optimalisasi hyperparameter menggunakan Random Search, evaluasi menggunakan Confusion Matrix berdasarkan nilai Accuracy, Precision, Recall, dan F1-Score, serta analisis feature importance. Sebelum dilakukan optimalisasi, model Random Forest baseline menghasilkan accuracy sebesar 70,39%, precision sebesar 56,67%, recall sebesar 62,96%, dan f1-score sebesar 59,65%. Setelah dilakukan optimalisasi dengan Random Search kinerja model meningkat menjadi accuracy sebesar 71,67%, precision sebesar 58,43%, recall sebesar 64,20%, dan f1-score sebesar 61,18%. Hasil menunjukkan bahwa optimalisasi hyperparameter dengan Random Seach mampu meningkatkan peforma model Random Forest dalam mengklasifikasi risiko stunting. Kontribusi penelitian ini adalah menghasilkan model klasifikasi risiko stunting berbasis algoritma Random Forest yang dioptimalkan menggunakan Random Search sehingga diperoleh kombinasi hyperparameter yang lebih optimal dibandingkan parameter bawaan. Selain itu, penelitian ini memberikan analisis perbandingan performa model sebelum dan sesudah optimalisasi serta informasi mengenai variabel-variabel yang paling berpengaruh terhadap klasifikasi risiko stunting. Hasil penelitian ini diharapkan dapat mendukung tenaga kesehatan dan pemerintah daerah dalam mengidentifikasi risiko stunting secara lebih cepat dan akurat sebagai dasar penyusunan strategi pencegahan yang lebih tepat sasaran.

Kata Kunci: Stunting; Machine Learning; Random Forest; Random Search; Risiko Stunting.

Abstract—Stunting is a condition of impaired child growth resulting from chronic nutritional deficiency over an extended period. This condition not only hinders physical growth but also impedes learning abilities and increases the risk of various future diseases. The high prevalence of stunting in Bima City highlights the need to utilize Machine Learning methods to analyze stunting risk factors more accurately. This study aims to optimize the performance of the Random Forest algorithm using the Random Search method to classify stunting risk among children under five in Sambinae Urban Village, Bima City. The dataset comprises records for 1,162 children under five, featuring 20 attributes obtained from the Mpunda Community Health Center (Puskesmas) in Bima City. The research stages include data collection, preprocessing, data splitting, construction of a baseline Random Forest model, hyperparameter optimization using Random Search, evaluation via a Confusion Matrix (based on Accuracy, Precision, Recall, and F1-Score), and feature importance analysis. Prior to optimization, the baseline Random Forest model yielded an accuracy of 70,39%, precision of 56,67%, recall of 62,96%, and an F1-score of 59,65%. Following optimization with Random Search, model performance improved to an accuracy of 71,67%, precision of 58,43%, recall of 64,20%, and an F1-score of 61,18%. The results demonstrate that hyperparameter optimization using Random Search effectively enhances the Random Forest model's performance in classifying stunting risk. The study contributes a stunting risk classification model based on the Random Forest algorithm, optimized via Random Search to achieve a more effective hyperparameter combination than the default settings. Furthermore, the study provides a comparative performance analysis before and after optimization, along with insights into the variables that most significantly influence stunting risk classification. These findings are expected to assist healthcare professionals and local government authorities in identifying stunting risks more rapidly and accurately, thereby serving as a foundation for formulating more targeted prevention strategies**.**

Keywords: Stunting; Machine Learning; Random Forest; Random Search; Stunting Risk

1. PENDAHULUAN

Stunting merupakan salah satu permasalahan kesehatan yang masih menjadi perhatian utama di Indonesia karena dapat memengaruhi kualitas sumber daya manusia. Stunting terjadi ketika pertumbuhan anak tidak sesuai dengan usianya akibat kekurangan gizi yang berlangsung dalam waktu yang lama. Kondisi ini umumnya terjadi sejak masa kehamilan hingga anak berusia dua tahun, yang dikenal sebagai periode 1.000 Hari Pertama Kehidupan (HPK). Masa tersebut merupakan tahap yang sangat penting karena berpengaruh terhadap pertumbuhan fisik, perkembangan otak, dan kesehatan anak [1]. Kondisi tersebut berdampak terhadap kemampuan belajar, serta meningkatkan risiko berbagai penyakit di masa depan [2]. Berdasarkan hasil Survei Status Gizi Indonesia (SSGI), angka prevalensi stunting di Indonesia menunjukkan adanya penurunan. Pada tahun 2022 prevalensi stunting tercatat sebesar 30,8%, kemudian turun menjadi 29,6% pada tahun 2023. Walaupun mengalami penurunan, angka tersebut masih tergolong tinggi karena belum mencapai target nasional yang ditetapkan pemerintah, yaitu sebesar

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1202−1211
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10531

14%. Di Kota Bima, jumlah balita yang mengalami stunting mencapai 1.145 jiwa dengan prevalensi sebesar 11,3% [3]. Kondisi ini menunjukkan bahwa stunting masih menjadi permasalahan yang memerlukan perhatian serius dan upaya yang mampu mengidentifikasi faktor risiko stunting secara cepat dan tepat. Oleh karena itu, diperlukan suatu pendekatann yang mampu membantu mengidentifikasi risiko stunting secara cepat dan tepat, sebagai dasar dalam penyusunan strategi pencegahan dan penanganan yang lebih efektif [4].

Seiring dengan meningkatnya kasus stunting, perkembangan teknologi menjadi solusi yang layak untuk mendukung proses identifikasi risiko stunting melalui proses analisis data kesehatan balita. Salah satu algoritma klasifikasi yang banyak digunakan adalah algoritma Random Forest, yang merupakan algoritma yang bekerja dengan membangun banyak pohon keputusan menggunakan teknik Bootstrap Aggregating (bagging) serta memilih fitur secara acak pada setiap proses pembentukan pohon. Algoritma ini dikenal memiliki kemampuan yang baik, tahan terhadap overfitting, serta mampu menentukan tingkat kepentingan setiap variabel [5]. Meskipun demikian, peforma model Random Forest sangat dipengaruhi oleh nilai hyperparameter yang digunakan. Penggunaan parameter bawaan (default parameter) belum tentu menghasilkan peforma model yang optimal. Oleh karena itu penelitian ini menerapkan tahap optimalisasi terhadap Random Forest dengan menggunakan Random Search yang bekerja dengan memilih kombinasi nilai parameter yang dipilih secara acak pada ruang pencarian (search space) yang telah ditentukan sehingga hasil lebih efisien dalam menentukan kombinasi parameter terbaik dibandingkan mencoba seluruh kemungkinan parameter [6]. Berbagai penelitian sebelumnya telah menerapkan berbagai pendekatan dalam mengidentifikasi maupun mengklasifikasikan kasus stunting. Penelitian yang dilakukan oleh Ratnasari dkk, membandingkan beberapa algoritma klasifikasi, yaitu Random Forest, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), dan Naïve Bayes untuk deteksi dini stunting. Hasil penelitian menunjukkan bahwa Random Forest mampu memberikan performa klasifikasi yang lebih baik dibandingkan algoritma lainnya. Namun, penelitian tersebut masih menggunakan parameter bawaan (default parameter) sehingga belum mengevaluasi pengaruh optimalisasi hyperparameter terhadap peningkatan performa model [7]. Selanjutnya penelitian yang dilakukan Shen et al, mengembangkan model prediksi stunting menggunakan beberapa algoritma machine learning, seperti Logistic Regression, Support Vector Machine (SVM), Conditional Decision Tree, dan Extreme Gradient Boosting (XGBoost), dengan menerapkan teknik feature selection menggunakan LASSO dan Random Forest Recursive Feature Elimination (RF-RFE) serta hyperparameter tuning. Penelitian tersebut menghasilkan model LASSO-XGBoost sebagai model dengan performa terbaik. Namun, penelitian tersebut belum mengevaluasi algoritma Random Forest sebagai model utama yang dioptimalkan menggunakan metode Random Search sehingga potensi peningkatan performa Random Forest pada klasifikasi risiko stunting masih belum diketahui [8]. Mhd Adi Setiawan dkk, menerapkan algoritma Metaheuristik untuk mengoptimalkan parameter pada algoritma K-Nearest Neighbor (KNN) dalam pengelompokan faktor stunting. Penelitian tersebut menunjukkan bahwa proses optimalisasi mampu meningkatkan performa algoritma KNN. Meskipun demikian, penelitian tersebut belum menerapkan optimalisasi pada algoritma Random Forest yang memiliki karakteristik berbeda dalam proses pembentukan model klasifikasi [9]. Penelitian oleh Anggoro dan Mukti yang membandingkan metode Grid Search dan Random Search untuk optimalisasi hyperparameter pada algoritma Extreme Gradient Boosting (XGBoost). Hasil penelitian menunjukkan bahwa Random Search mampu memperoleh kombinasi hyperparameter yang lebih efisien dibandingkan Grid Search dengan performa model yang kompetitif. Akan tetapi, penelitian tersebut diterapkan pada algoritma XGBoost dalam kasus prediksi gagal ginjal kronis sehingga belum membahas penerapan metode Random Search pada algoritma Random Forest untuk klasifikasi risiko stunting [10]. Penelitian yang dilakukan oleh Chola et al, membandingkan beberapa algoritma machine learning, yaitu Logistic Regression, Random Forest, Support Vector Classification, dan Extreme Gradient Boosting (XGBoost) untuk mengklasifikasikan stunting pada balita di Zambia. Hasil penelitian menunjukkan bahwa algoritma Random Forest memberikan performa klasifikasi terbaik dibandingkan algoritma lainnya berdasarkan nilai accuracy, f1-score, precision, dan recall. Selain itu, penelitian tersebut juga mengidentifikasi variabel-variabel yang memiliki pengaruh penting terhadap terjadinya stunting. Meskipun demikian, penelitian tersebut masih berfokus pada perbandingan performa beberapa algoritma klasifikasi dan belum melakukan optimalisasi hyperparameter pada algoritma Random Forest. Oleh karena itu, pengaruh optimalisasi hyperparameter terhadap peningkatan performa Random Forest belum dapat diketahui secara komprehensif [11].

Meskipun telah ada beberapa penelitian sebelumnya, mengenai pengembangan model klasifikasi terhadap risiko stunting. Namun, penelitian-penelitian tersebut umumnya masih berfokus pada metode yang berbeda, sementara kajian mengenai optimalisasi hyperparameter pada algoritma Random Forest dengan menggunakan Random Search pada klasifikasi risiko stunting masih terbatas. Selain itu, sebagian besar penelitian belum membandingkan secara eksplisit peforma model Random Forest sebelum dan sesudah dilakukan optimalisasi, sehingga pengaruh optimalisasi terhadap peningkatan kinerja model belum dapat dievaluasi secara komprehensif. Penelitian ini bertujuan untuk mengisi kesenjangan dari penelitian tersebut dengan mengoptimalkan hyperparameter Random Forest menggunakan metode Random Search untuk klasifikasi risiko stunting pada balita. Kontribusi penelitian ini adalah menghasilkan model klasifikasi risiko stunting berbasis algoritma Random Forest yang dioptimalkan menggunakan metode Random Search untuk memperoleh kombinasi hyperparameter terbaik. Selain itu, penelitian ini menyajikan analisis komparatif performa model sebelum dan sesudah optimasi berdasarkan nilai accuracy, precision, recall, dan F1-score, serta mengidentifikasi variabel-variabel yang memiliki pengaruh paling besar terhadap klasifikasi risiko stunting melalui analisis feature importance. Hasil dari penelitian

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1202−1211
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10531

ini diharapkan dapat menjadi informasi pendukung bagi tenaga kesehatan maupun pemerintah dalam upaya pencegahan serta penanganan stunting secara lebih tepat di Indonesia, khususnya di Kelurahan Sambinae Kota Bima.

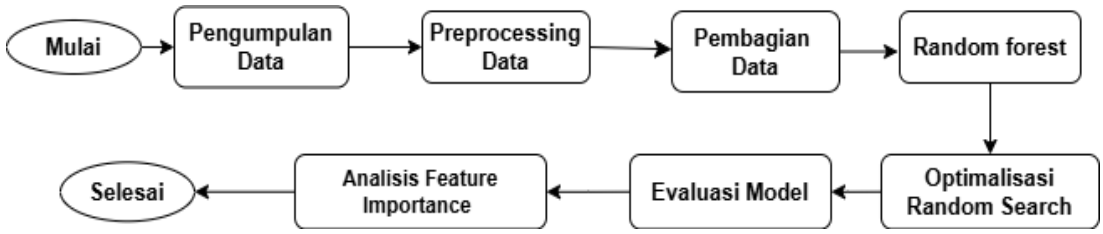
2. METODOLOGI PENELITIAN

2.1 Objek Penelitian

Objek penelitian ini menggunakan pendekatan kuantitatif dengan memanfaatkan data sekunder berupa data balita yang diperoleh dari Puskesmas Mpunda Kota Bima dalam format Microsoft Excel. Dataset yang digunakan terdiri dari 1.162 data yang telah melalui proses seleksi berdasarkan kelengkapan atribut sehingga layak digunakan. Dataset berisi informasi mengenai kondisi pertumbuhan dan status gizi balita yang berpotensi memengaruhi kejadian stunting. Atribut yang tersedia dalam dataset meliputi Nama, Jenis Kelamin (JK), Puskesmas, Desa/Kel, RT, RW, Tanggal Lahir, Usia Saat Ukur, Tanggal Pengukuran, BB, TB, Cara Ukur, Lila, BB/U, ZS BB/U, TB/U, ZS TB/U, BB/TB, ZS BB/TB, serta Kategori. Pada tahap preprocessing, atribut yang tidak berpengaruh terhadap proses klasifikasi, seperti identitas dan lokasi, dihapus sehingga variabel yang digunakan dalam pemodelan terdiri atas JK, BB, TB, LiLA, BB/U, ZS BB/U, TB/U, ZS TB/U, BB/TB, ZS B/TB, Usia_Bulan, dan Kategori. Populasi penelitian merupakan seluruh data balita yang tercatat pada Puskesmas Mpunda Kota Bima, sedangkan sampel penelitian adalah data balita yang memenuhi kriteria kelengkapan atribut sehingga dapat digunakan dalam proses analisis.

2.2 Tahapan Penelitian

Berikut tahapan penelitian seperti yang terdapat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Tahapan penelitian yang digunakan pada tahap penelitian ini ditunjukkan pada Gambar 1. Setiap tahapan saling berkaitan yang terdiri dari pengumpulan data, preprocessing data, pembagian data, pemodelan algoritma Random Forest, optimalisasi dengan Random Search, evaluasi serta analisis feature importance sebagai dasar evaluasi model.

2.2.1 Pengumpulan Data

Penelitian ini menggunakan pendekatan kuantitatif dengan data sekunder yang diperoleh dari Puskesmas Mpunda Kota Bima dalam format Excel. Dataset terdiri dari 1.162 data,yang telah memenuhi kriteria kelengkapan data.

2.2.2 Pre-Procesing Data

Tahap preprocessing data bertujuan untuk meningkatkan kualitas data sehingga dapat digunakan dalam proses pemodelan machine learning dengan bekerja secara optimal [12]. Proses preprocessing data meliputi beberapa tahapan sebagai berikut:

a. Data Cleaning: Tahap ini dilakukan untuk membersihkan data dengan menghapus atribut yang tidak relevan terhadap proses klasifikasi, seperti Nama, Tanggal Lahir, Puskesmas, Desa/Kelurahan, RT, RW, Tanggal Pengukuran, dan Cara Ukur. Atribut-atribut tersebut dihapus karena bersifat sebagai identitas, informasi administratif, atau informasi pencatatan yang tidak memiliki hubungan langsung dengan penentuan status stunting. Selain itu, keberadaan atribut tersebut berpotensi menambah kompleksitas data tanpa memberikan kontribusi terhadap peningkatan akurasi model. Selanjutnya dilakukan pemeriksaan terhadap data yang hilang (*missing value*) maupun data duplikat untuk memastikan kualitas data sebelum diproses lebih lanjut..
b. Tranformasi Data: Tahap ini dilakukan dengan mengubah atribut Usia Saat Ukur menjadi Usia_Bulan agar memiliki bentuk numerik yang lebih representatif dalam menggambarkan usia balita. Selanjutnya atribut kategorikal seperti jenis kelamin, BB/U, TB/U, BB/TB dan kategori diubah menjadi bentuk numerik menggunakan teknik encoding agar diproses oleh algoritma.
c. Konversi Data Numerik: Tahap ini dilakukan terhadap atribut BB, TB, LILA, ZS BB/U, ZS TB/U, dan ZS BB/TB untuk memastikan seluruh atribut numerik memiliki tipe data yang sesuai untuk diproses oleh algoritma.
d. Seleksi Fitur: Digunakan untuk memilih atribut yang relavan dalam proses klasifikasi risiko stunting. Berdasarkan hasil seleksi atribut yang digunakan terdiri dari JK, BB, TB, LiLA, BB/U, ZS BB/U, TB/U, ZS

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1202−1211
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10531

TB/U, BB/TB, ZS B/TB, Usia_Bulan yang digunakan sebagai variabel independen, sedangkan variabel Kategori digunakan sebagai variabel target.

2.2.3 Pembagian Data

Dataset dibagi menjadi data latih sebesar 80% dan data uji sebesar 20% menggunakan metode train-test split. Berdasarkan distribusi kelas, data terdiri atas 756 data stunting dan 406 data tidak stunting, sehingga terdapat ketidakseimbangan kelas (class imbalance) dengan rasio sekitar 1,86:1. Untuk mengatasi kondisi tersebut, teknik Synthetic Minority Over-sampling Technique (SMOTE) diterapkan pada data latih guna menghasilkan distribusi kelas yang lebih seimbang sebelum proses pelatihan model dilakukan [13].

2.2.4 Pemodelan Random Forest

Model Random Forest adalah algoritma klasifikasi yang dikembangkan dari metode Decision Tree yang bekerja dengan membangun sejumlah pohon keputusan menggunakan teknik Bootstrap Aggregating (bagging), yaitu pemilihan fitur secara acak pada setiap pohon. Selain memiliki kemampuan klasifikasi yang baik, Random Forest juga mampu menghitung tingkat kepentingan setiap variabel (fitur importance) [14].

2.2.5 Optimalisasi Random Search

Random Search bekerja dengan menguji sebagian kombinasi yang dipilih secara acak sehingga proses optimalisasi menjadi lebih efisien, terutama pada ruang pencarian yang besar. Meskipun tidak menjamin diperolehnya kombinasi parameter terbaik secara global, metode ini sering mampu menghasilkan konfigurasi hyperparameter yang memberikan kinerja model yang baik dengan waktu komputasi yang lebih singkat [15]. Random Search merupakan salah satu teknik optimasi hyperparameter yang mampu meningkatkan performa model machine learning dengan mengeksplorasi kombinasi parameter secara efisien tanpa harus menguji seluruh kemungkinan kombinasi parameter [16].

Proses optimalisasi hyperparameter pada algoritma Random Forest dilakukan menggunakan metode Random Search dengan mengeksplorasi beberapa kombinasi parameter yang berpengaruh terhadap kinerja model. Parameter yang dioptimalkan meliputi n_estimators, yaitu jumlah pohon keputusan (decision tree) yang dibangun dalam model; max_depth, yang menentukan kedalaman maksimum setiap pohon keputusan; min_samples_split, yaitu jumlah minimum sampel yang diperlukan untuk membagi suatu node menjadi cabang baru; min_samples_leaf, yang menunjukkan jumlah minimum sampel yang harus terdapat pada setiap node daun (leaf node); serta bootstrap, yaitu teknik pengambilan sampel secara acak dengan pengembalian (sampling with replacement) yang digunakan dalam proses pembentukan setiap pohon keputusan. Kombinasi dari hyperparameter tersebut dipilih secara acak oleh metode Random Search untuk memperoleh konfigurasi yang mampu menghasilkan performa klasifikasi terbaik [17].

2.2.6 Evaluasi Model

Tahap evaluasi model dilakukan untuk mengevaluasi model Random Forest sebelum dioptimalisasi dengan yang sudah dioptimalisasi. Tujuan dilakukan evaluasi ini untuk mengetahui seberapa baik kinerja model dari masing-masing algoritma yang digunakan. Terdapat beberapa confusion matrix seperti pada Tabel 1 berikut:

Tabel 1.Confusion Matrix

	Prediksi Positif	**Prediksi Negatif**
Aktual Positif	TP	FN
Aktual Negatif	FN	TN

Dari Tabel 1 dapat membantu menentukan peforma model ialah sebegai berikut:

a. Accuracy: yaitu mengukur untuk mengetahui seberapa baik model klasifikasi dalam memberikan prediksi yang benar terhadap seluruh data uji [18].

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FN+FP}} \quad (1)$$

b. Precision: merupakan ukuran yang digunakan untuk mengetahui seberapa banyak prediksi positif yang dihasilkan model benar-benar termasuk ke dalam kelas positif [19].

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (2)$$

c. Recall: merupakan ukuran yang digunakan untuk mengetahui kemampuan model dalam mengenali seluruh data yang sebenarnya termasuk ke dalam kelas positif [20].

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (3)$$

d. F1-Score: merupakan metrik evaluasi yang menggabungkan nilai presisi dan recall untuk memberikan gambaran mengenai kinerja model secara keseluruhan [21].

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1202−1211
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10531

$$F1 - Score = 2X\frac{Precision+Recall}{Precision+Recall} \quad (4)$$

2.2.7 Analisis Feature Importance

Analisis feature importance dilakukan untuk mengidentifikasi seberapa besar kontribusi setiap variabel terhadap model. Hasil evaluasi model baseline kemudian dibandingkan dengan model hasil optimalisasi Random Search untuk mengetahui pengaruh optimalisasi hyperparameter terhadap peningkatan kinerja algoritma Random Forest [22].

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Dataset yang digunakan pada penelitian ini merupakan data balita yang diperoleh dari Puskesmas Mpunda Kota Bima. Pengambilan data dilakukan pada tanggal 18 Desember 2025. Dataset terdiri atas 1.162 data balita yang memuat informasi mengenai kondisi pertumbuhan dan status gizi balita. Variabel yang yang digunakan yaitu JK, Usia_Bulan, BB, TB, LILA, BB/U, ZS BB/U, TB/U, ZS TB/U, BB/TB, ZS BB/U dan Kategori. Tabel 2 menyajikan data balita yang digunakan sebagai dataset penelitian.

Tabel 2. Dataset Stunting

JK	Usia Bulan	BB	TB	LiLA	BB/U	ZS BB/U	TB/U	ZS TB/U	BB/TB	ZS BB/TB	Kategori
L	23	7.5	76	14.3	Sangat kurang	-4.04	Sangat pendek	-3.54	Gizi Baik	-3.4	Stunting
L	43	12	86	16	Kurang	-2.15	Sangat pendek	-3.69	Gizi Baik	0.09	Tidak Stunting
L	53	13	96	12.3	Kurang	-2.19	Pendek	-2.34	Gizi Buruk	-1.2	Stunting
L	53	13.2	95.5	14.5	Kurang	-2.07	Pendek	-2.45	Gizi Buruk	-0.9	Stunting
P	40	11	90	14.5	Kurang	-2.33	Pendek	-2.07	Gizi Buruk	-1.61	Stunting

Berdasarkan Tabel 2, dataset memuat informasi antropometri dan status gizi balita yang digunakan sebagai dasar dalam proses klasifikasi risiko stunting. Seluruh data selanjutnya melalui tahap preprocessing sebelum digunakan dalam proses pembentukan model.

3.2 Preprocessing Data

Tahap preprocessing data dilakukan untuk meningkatkan kualitas data sebelum proses klasifikasi.Tahapan preprocesing meliputi data cleaning, tranformasi data, konversi data numerik, penanganan missing value, serta ecoding pada variabel kategorikal. Hasil preprocessing data ditampilkan pada Tabel 3 berikut:

Tabel 3. Data Setelah Preprocessing

JK	Usia Bulan	BB	TB	LiLA	BB/U	ZS BB/U	TB/U	ZS TB/U	BB/TB	ZS BB/TB	Kategori
0	23	7.5	76.0	14.3	0	-4.04	0	-3.54	0	-3.4	0
1	43	12.0	86.9	16.0	1	-2.15	0	-3.69	3	0.09	1
0	53	13.0	96.0	12.3	1	-2.19	1	-2.34	3	-1.2	0
0	53	13.2	95.5	14.5	1	-2.07	1	-2.45	3	-0.9	0
0	40	11.6	90.0	14.5	1	-2.33	1	-2.07	3	-1.61	0

Pada tahap encoding, beberapa variabel kategorikal dikonversi ke dalam bentuk numerik agar dapat diproses oleh algoritma Random Forest. Variabel Jenis Kelamin (JK) dikonnversikan menjadi 0 untuk laki-laki dan 1 untuk perempuan. Variabel Berat Badan menurut Umur (BB/U) dikonversi menjadi 0 untuk sangat kurang, 1 untuk kurang, dan 2 untuk normal. Variabel Tinggi Badan menurut Umur (TB/U) dikonversi menjadi 0 untuk sangat pendek dan 1 untuk pendek. Selanjutnya, variabel Berat Badan menurut Tinggi Badan (BB/TB) dikonversi menjadi 0 untuk gizi buruk, 1 untuk gizi kurang, 2 untuk normal, dan 3 untuk gizi baik. Adapun variabel target, yaitu Kategori, dikodekan menjadi 0 untuk stunting dan 1 untuk tidak stunting. Proses ini bertujuan untuk mengubah data kategorikal menjadi format numerik sehingga dapat diolah secara optimal oleh algoritma klasifikasi Random Forest.

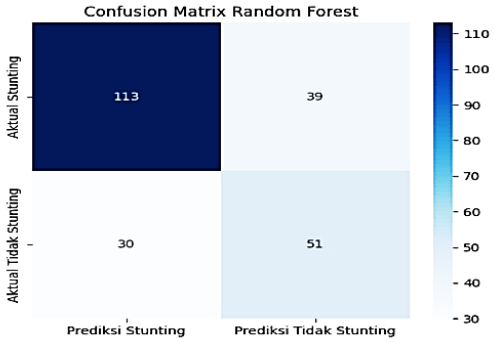
Berdasarkan Tabel 3, seluruh atribut telah berhasil dikonversi dalam format numerik sehingga dapat diproses oleh algoritma Random Forest. Variabel kategorikal telah diubah menjadi nilai numerik menggunakan teknik label encoding. Selain itu, atribut yang tidak relevan juga telah dihapus. Selanjutnya, dilakukan penanganan terhadap nilai yang hilang (missing value) sehingga dataset yang digunakan pada proses pemodelan tidak mengandung data kosong dan diperoleh dataset akhir yang siap digunakan pada tahap pembagian data, penyeimbangan kelas menggunakan SMOTE, dan pembangunan model klasifikasi.

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1202−1211
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10531

3.3 Pemodelan Random Forest

Setelah tahap preprocessing selesai, selanjutnya dibangun model klasifikasi dengan algoritma Random Forest pada parameter bawaan (default parameter) sebagai model baseline yang digunakan untuk mengetahui performa awal algoritma sebelum dilakukan proses optimalisasi hyperparameter. Hasil klasifikasi model baseline ditampilkan dalam bentuk Confusion matrix pada Gambar 2.

Berdasarkan Gambar 2, model mampu mengklasifikasikan data ke dalam kelas stunting dan tidak stunting. Hasil confusion matrix tersebut kemudian digunakan untuk menghitung nilai accuracy, precision, recall, dan F1-score. Hasil evaluasi model Random Forest ditampilkan pada Tabel 4.



Gambar 2. Confusion Matrix Random Forest

Tabel 4. Hasil Evaluasi Random Forest

Parameter	Random Forest
Accuracy	70,39%
Precision	56,67%
Recall	62,96%
F1-Score	59,65%

Berdasarkan Tabel 4, model Random Forest baseline memperoleh nilai accuracy sebesar 70,39%, yang menunjukkan bahwa sekitar 70% data uji berhasil diklasifikasikan dengan benar. Nilai precision sebesar 56,67% menunjukkan bahwa lebih dari setengah data yang diprediksi sebagai stunting merupakan prediksi yang benar. Sementara itu, nilai recall sebesar 62,96% mengindikasikan bahwa model mampu mengenali sekitar 63% data balita yang benar-benar mengalami stunting. Adapun F1-Score sebesar 59,65% menunjukkan keseimbangan antara nilai precision dan recall sehingga dapat digunakan sebagai gambaran umum performa model dalam melakukan klasifikasi. Hasil evaluasi ini selanjutnya dijadikan sebagai baseline untuk dibandingkan dengan model Random Forest yang telah dioptimasi menggunakan metode Random Search.

3.4 Pemodelan Random Forest dengan Random Search

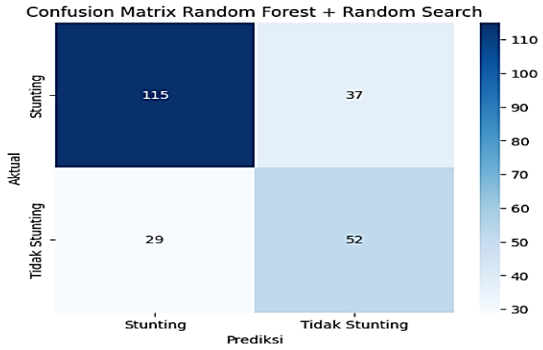
Setelah model Random Forest baseline dibangun, tahap berikutnya adalah melakukan optimalisasi hyperparameter menggunakan metode Random Search. Proses optimalisasi dilakukan menggunakan RandomizedSearchCV, yaitu metode yang memilih kombinasi hyperparameter secara acak dari ruang pencarian (search space) yang telah ditentukan sebelumnya. Setiap kombinasi parameter dievaluasi menggunakan teknik Cross Validation, sehingga konfigurasi yang diperoleh diharapkan mampu menghasilkan performa klasifikasi yang lebih baik dibandingkan model baseline. Berdasarkan proses optimalisasi tersebut diperoleh kombinasi hyperparameter terbaik yang digunakan untuk membangun model Random Forest. Nilai hyperparameter terbaik disajikan pada Tabel 5.

Tabel 5. Hyperparameter Random Search

Hyperparameter	Nilai
n_estimators	300
min_samples_split	2
min_samples_leaf	2
max_features	sqrt
max_depth	None

Selanjutnya dilakukan optimalisasi hyperparameter menggunakan metode Random Search di evaluasi menggunakan confusion matrix sebagaimana ditunjukkan pada Gambar 3.

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1202−1211
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10531



Gambar 3. Confusion Matrix Random Forest + Random Search

Berdasarkan Gambar 3, terlihat hasil perhitungan model setelah optimalisasi mengalami perbaikan dibandingkan model baseline. Perubahan tersebut kemudian dievaluasi menggunakan matrix accuracy, precision, recall, dan F1-score. Perbandingan hasil evaluasi ditampilkan pada Tabel 6.

Tabel 6. Evaluasi Random Search

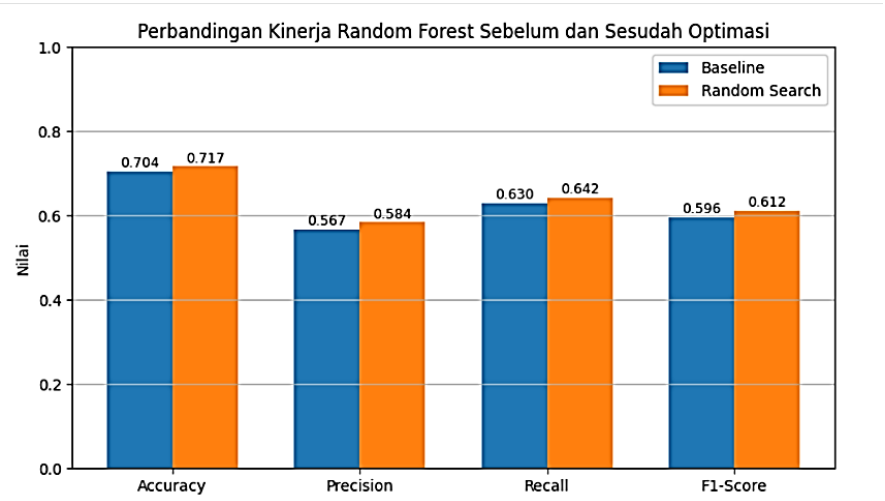
Parameter	Random Search
Accuracy	71,67%
Precision	58,43%
Recall	64,20%
F1-Score	61,18%

Berdasarkan Tabel 6, model Random Forest yang telah dioptimalisasi menggunakan Random Search memperoleh accuracy sebesar 71,67%, meningkat dibandingkan model baseline. Nilai precision meningkat menjadi 58,43%, yang menunjukkan bahwa prediksi terhadap kelas stunting menjadi lebih tepat. Selain itu, recall sebesar 64,20% menunjukkan bahwa model mampu mengenali lebih banyak data balita yang benar-benar mengalami stunting dibandingkan sebelum optimalisasi. Nilai F1-Score sebesar 61,18% juga mengalami peningkatan, yang menunjukkan keseimbangan antara precision dan recall menjadi lebih baik setelah dilakukan proses optimalisasi hyperparameter.

Tabel 7. Perbandingan Random Forest dan Random Search

Parameter	Random Forest	Random Search
Accuracy	70,39%	71,67%
Precision	56,67%	58,43%
Recall	62,96%	64,20%
F1-Score	59,65%	61,18%

Berdasarkan Tabel 7, penerapan Random Search memberikan peningkatan pada seluruh metrik evaluasi dibandingkan model Random Forest baseline. Nilai accuracy meningkat dari 70,39% menjadi 71,67% atau sebesar 1,28%. Nilai precision meningkat sebesar 1,76%, recall meningkat sebesar 1,24%, sedangkan F1-Score meningkat sebesar 1,53%.



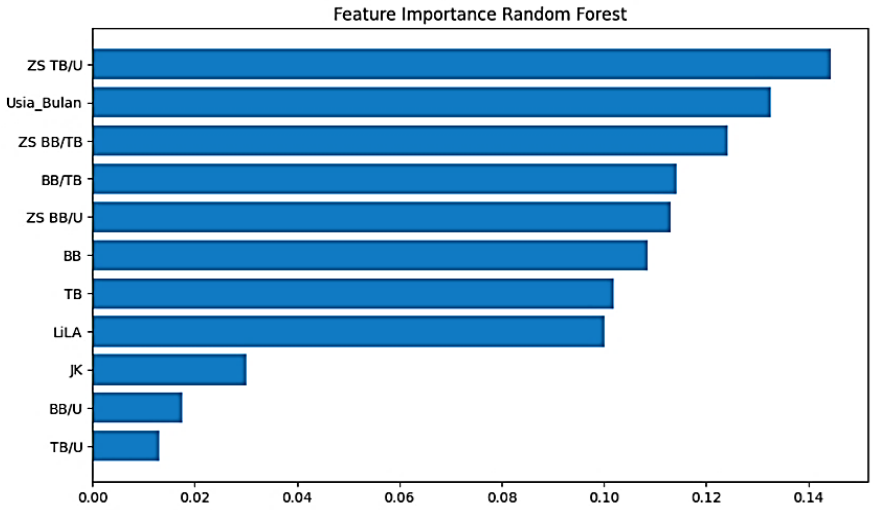
Gambar 4. Perbandingan Random Forest dan Random Search

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1202−1211
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10531

Berdasarkan Gambar 4, model Random Forest yang dioptimalisasi dengan Random Search memperoleh nilai yang lebih tinggi pada seluruh metrik evaluasi dibandingkan model baseline. Peningkatan paling besar terjadi pada nilai precision dan F1-Score, sedangkan accuracy dan recall juga menunjukkan kenaikan meskipun tidak terlalu besar. Hasil tersebut menunjukkan bahwa proses optimalisasi hyperparameter mampu memperbaiki kemampuan model dalam mengklasifikasikan data balita stunting dan tidak stunting. Meskipun peningkatan yang diperoleh tidak terlalu besar, hasil ini mengindikasikan bahwa kombinasi hyperparameter yang diperoleh melalui Random Search menghasilkan model yang lebih baik dibandingkan penggunaan parameter bawaan.

3.5 Analisis Fiture Importance

Berdasarkan Gambar 5, variabel ZS TB/U memiliki nilai feature importance tertinggi, sehingga menjadi variabel yang paling berpengaruh dalam proses klasifikasi stunting. Hasil ini menunjukkan bahwa indikator Z-Score Tinggi Badan menurut Umur (TB/U) merupakan faktor utama dalam membedakan balita stunting dan tidak stunting karena mencerminkan kondisi pertumbuhan linier yang menjadi dasar penentuan status stunting. Variabel Usia_Bulan, ZS BB/TB, BB/TB, dan ZS BB/U juga memberikan kontribusi yang cukup besar terhadap kinerja model. Sebaliknya, variabel JK, BB/U, dan TB/U memiliki nilai feature importance yang lebih rendah, sehingga pengaruhnya terhadap proses klasifikasi relatif kecil. Secara keseluruhan, hasil ini menunjukkan bahwa variabel berbasis Z-Score, khususnya ZS TB/U, memiliki peran yang lebih dominan dibandingkan variabel antropometri lainnya dalam mengidentifikasi risiko stunting.



Gambar 5. Analisis Fiture Importance

Temuan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa indikator tinggi badan menurut umur (TB/U) merupakan indikator utama dalam penentuan status stunting karena mampu menggambarkan gangguan pertumbuhan linier akibat kekurangan gizi kronis. Oleh karena itu, variabel ZS TB/U menjadi salah satu indikator yang paling berpengaruh dalam proses klasifikasi stunting.

3.6 Pembahasan

Hasil penelitian menunjukkan bahwa optimalisasi hyperparameter menggunakan Random Search mampu meningkatkan performa algoritma Random Forest pada klasifikasi risiko stunting. Hal ini ditunjukkan dengan meningkatnya nilai accuracy dari 70,39% menjadi 71,67%, precision dari 56,67% menjadi 58,43%, recall dari 62,96% menjadi 64,20%, serta F1-Score dari 59,65% menjadi 61,18%. Meskipun peningkatan yang diperoleh relatif kecil, hasil tersebut menunjukkan bahwa proses optimalisasi hyperparameter mampu menghasilkan konfigurasi model yang lebih baik dibandingkan penggunaan parameter bawaan.

Temuan ini sejalan dengan penelitian Wahidin dan Andika, yang menunjukkan bahwa algoritma Random Forest memiliki kemampuan yang baik dalam klasifikasi stunting [7]. Selain itu, hasil penelitian ini juga memperkuat penelitian Anggoro dan Mukti, yang menyatakan bahwa metode Random Search efektif dalam memperoleh kombinasi hyperparameter yang mampu meningkatkan performa model. Berbeda dengan penelitian tersebut yang menerapkan Random Search pada algoritma XGBoost, penelitian ini menunjukkan bahwa metode yang sama juga mampu meningkatkan kinerja Random Forest pada klasifikasi risiko stunting. Hasil analisis feature importance juga menunjukkan bahwa variabel ZS TB/U merupakan faktor yang paling berpengaruh dalam klasifikasi risiko stunting. Temuan ini sejalan dengan penelitian Chola et al, yang menyatakan bahwa indikator tinggi badan menurut umur merupakan variabel utama dalam identifikasi stunting [11]. Perbedaan nilai performa dengan beberapa penelitian terdahulu kemungkinan dipengaruhi oleh perbedaan karakteristik dataset, jumlah sampel, distribusi kelas, serta ruang pencarian hyperparameter yang digunakan.

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1202−1211
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10531

4. KESIMPULAN

Berdasarkan hasil penelitian algoritma Random Forest mampu digunakan untuk klasifikasi risiko stunting pada balita menggunakan data antropometri yang diperoleh dari Puskesmas Mpunda Kota Bima. Penerapan Random Search sebagai metode optimasi hyperparameter menghasilkan peningkatan kinerja model pada seluruh metrik evaluasi dibandingkan model baseline, sehingga menunjukkan bahwa optimalisasi hyperparameter dapat meningkatkan kemampuan model dalam mengklasifikasikan data stunting. Hasil analisis feature importance menunjukkan bahwa ZS TB/U merupakan variabel yang memberikan kontribusi terbesar terhadap proses klasifikasi, diikuti oleh Usia_Bulan, ZS BB/TB, BB/TB, dan ZS BB/U, sedangkan variabel Jenis Kelamin (JK) memiliki kontribusi paling rendah. Temuan ini menunjukkan bahwa indikator antropometri, khususnya Z-Score Tinggi Badan menurut Umur (ZS TB/U), memiliki peran yang lebih dominan dalam identifikasi risiko stunting. Kontribusi penelitian ini adalah menghasilkan model klasifikasi risiko stunting berbasis algoritma Random Forest yang dioptimalkan menggunakan metode Random Search sehingga diperoleh kombinasi hyperparameter yang lebih optimal dibandingkan penggunaan parameter bawaan (default parameter). Selain itu, penelitian ini menyajikan analisis komparatif performa model sebelum dan sesudah proses optimasi serta mengidentifikasi variabel-variabel yang paling berpengaruh terhadap klasifikasi risiko stunting. Hasil penelitian ini diharapkan dapat menjadi referensi dalam pengembangan penerapan machine learning di bidang kesehatan serta mendukung pengambilan keputusan bagi tenaga kesehatan dan pemerintah dalam upaya identifikasi dini, pencegahan, dan penanganan stunting secara lebih tepat sasaran.Untuk penelitian selanjutnya, disarankan menggunakan dataset dengan cakupan wilayah dan jumlah sampel yang lebih besar serta membandingkan metode optimasi hyperparameter lainnya, seperti Grid Search, Bayesian Optimization, Genetic Algorithm, atau Optuna, guna memperoleh model dengan kemampuan prediksi yang lebih baik.

REFERENCES

[1] A. Suryawan, “Malnutrition in early life and its neurodevelopmental and cognitive consequences : a scoping review,” Nutr. Res. Rev., vol. 35, pp. 136–149, 2022, doi: 10.1017/S0954422421000159.
[2] A. T. Mulyani, M. A. Khairinisa, and A. Khatib, “Understanding Stunting : Impact , Causes , and Strategy to Accelerate Stunting Reduction — A Narrative Review,” Nutrients, vol. 17, no. 9, p. 1493, 2025, doi: org/10.3390/nu17091493.
[3] S. Ramadoan, “Model Intervensi Terpadu dalam Mengatasi Prevalensi Stunting di Kota Bima,” JGLP, vol. Vol 6 No 2, pp. 229–239, 2024, doi: org/10.47650/jglp.v6i2.1562.
[4] N. F. Sahamony and H. Rianto, “Analysis of Performance Comparison of Machine Learning Models for Predicting Stunting Risk in Children ’ s Growth Analisis Perbandingan Kinerja Model Machine Learning untuk Memprediksi Risiko Stunting pada Pertumbuhan Anak,” MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 4, no. April, pp. 413–422, 2024, doi: org/10.57152/malcom.v4i2.1210.
[5] J. R. Khan, J. H. Tomal, and E. Raheem, “Model and variable selection using machine learning methods with applications to childhood stunting in Bangladesh,” Informatics Heal. Soc. Care, vol. 00, no. 00, pp. 1–18, 2021, doi: 10.1080/17538157.2021.1904938.
[6] M. K. Ayele, G. A. Baye, and S. H. Yesuf, “Predicting stunting status among under five children in ethiopia using ensemblemachine learning algorithms,” scientificreports, vol. 15, pp. 1–11, 2025, doi: org/10.1038/s41598-025-03206-1 3.
[7] A. J. Wahidin and T. H. Andika, “Deteksi Dini Stunting Pada Anak Berdasarkan Indikator Antropometri dengan Menggunakan Algoritma Machine Learning,” J. Algoritm., vol. 21 No. 2, pp. 378–387, 2024, doi: 10.33364/algoritma/v.21-2.2122.
[8] H. Z. and Y. J. Shen, Hao, “Machine Learning Algorithms for Predicting Stunting among Under-Five Children in Papua New Guinea,” Children, vol. 10, pp. 1–18, 2023, doi: org/10.3390/children10101638.
[9] M. Adi, S. Aritonang, M. A. Siregar, S. A. Arnomo, and F. Farasalsabila, “Penerapan Algoritma Metaheuristik untuk Optimasi Nilai K pada KNN Pengelompokan Faktor Stunting,” J. Pustaka AI, vol. Vol . 5 No, pp. 611–618, 2025, doi: doi.org/10.55382/jurnalpustakaai.v5i3.1367.
[10] D. A. Anggoro, “Performance Comparison of Grid Search and Random Search Methods for Hyperparameter Tuning in Extreme Gradient Boosting Algorithm to Predict Chronic Kidney Failure,” nternational J. Intell. Eng. Syst., vol. 14, no. 6, pp. 198–207, 2021, doi: 10.22266/ijies2021.1231.19.
[11] O. N. Chilyabanyama et al., “Performance of Machine Learning Classifiers in Classifying Stunting among Under-Five Children in Zambia,” Children, vol. 9, pp. 1–11, 2022, doi: 10.3390/children9071082.
[12] A. Tawakuli and T. Engel, “Make your data fair : A survey of data preprocessing techniques that address biases in data towards fair AI,” J. Eng. Res., vol. 13, no. 3, pp. 2362–2369, 2025, doi: 10.1016/j.jer.2024.06.016.
[13] T. N. Id, K. Mengersen, D. Sous, and B. Liquet, “SMOTE-CD : SMOTE for compositional data,” PLoS One, vol. 18, no. June, pp. 1–19, 2023, doi: 10.1371/journal.pone.0287705.
[14] H. A. Salman, A. Kalakech, and A. Steiti, “Random Forest Algorithm Overview,” Babylonian J. Mach. Learn., vol. 2024, pp. 69–79, 2024, doi: org/10.58496/BJML/2024/007.
[15] F. Hosseini, C. Prieto, and C. Alvarez, “Hyperparameter optimization of regional hydrological LSTMs by random search : A case study from Basque Country , Spain,” J. Hydrol. J., vol. 643, no. April, 2024, doi: 10.1016/j.jhydrol.2024.132003.
[16] J. Saputra, C. Lawrencya, J. M. Saini, and S. Suharjito, “Hyperparameter optimization for cardiovascular disease data - driven prognostic system,” Vis. Comput. Ind. Biomed. Art, vol. 6, 2023, doi: 10.1186/s42492-023-00143-6.
[17] R. Andonie and A. C. Florea, “CCC Publications Weighted Random Search for CNN Hyperparameter Optimization,” Int. J. Comput. Commun. Control, vol. 15, no. 2, p. 3868, 2020, doi: 10.15837/ijccc.2020.2.3868.
[18] D. Chicco, N. Tötsch, and G. Jurman, “The Matthews correlation coefficient (MCC) is more reliable than balanced

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1202−1211
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10531

accuracy , bookmaker informedness , and markedness in two-class confusion matrix evaluation," BioData Min., vol. 14, pp. 1–22, 2021, doi: org/10.1186/s13040-021-00244-z.

[19] F. B. Streamlit, D. Aprilianto, and E. Rizal, "Klasifikasi Penyakit Kanker Paru Menggunakan Algoritma Random Forest," METIK J., vol. 9 NOMOR.2, pp. 275–283, 2025, doi: 10.47002/metik.v9i2.1076.

[20] M. Algoritma, S. V. M. Dan, R. Forest, and F. N. Firzatullah, "Analisis Sentimen Pengguna Aplikasi Byond BSI Pada Google Play Store," METIK J., vol. 9 Nomor.2, pp. 356–363, 2025, doi: 10.47002/metik.v9i2.1089.

[21] E. Helmud, E. Helmud, and P. Romadiana, "Classification Comparison Performance of Supervised Machine Learning Random Forest and Decision Tree Algorithms Using Confusion Matrix," J. SISFOKOM (Sistem Inf. dan Komputer), vol. 13, pp. 92–97, 2024, doi: 10.32736/sisfokom.v13i1.1985.

[22] M. Reyhansyah, D. Putra, A. Winata, M. A. Fathir, and Y. W. Prabowo, "Analisa Kerentanan Web Application Menggunakan Metode OWASP," J. MEDIA Inform., vol. 7, no. 1, pp. 476–487, 2026, doi: doi.org/10.55338/jumin.v7i1.8273.