



Implementasi *Random Forest* dalam Klasifikasi Kasus *Stunting* pada Balita dengan *Hyperparameter Tuning Grid Search*

Nida Faoziatun Khusna¹, Alfidha Rahmah², Rachmat Kahfiwan Nur³, Nasyiatul Izzah⁴, Kamilah Citra Chumairoh⁵, Fatkhurokhman Fauzi⁶

^{1,2,3,4,5}Statistika, Universitas Muhammadiyah Semarang

¹nidafaok@gmail.com, ²alfidhrhm@gmail.com, ³rachmatkahfi13@gmail.com, ⁴nasyiatulizzah255@gmail.com,

⁵kamilahcitra59@gmail.com, ⁶fatkhurokhmanf@unimus.ac.id

Corresponding author email: nidafaok@gmail.com

Abstract: *Stunting* is a global health issue that requires significant attention. It is closely linked to impaired growth and development caused by chronic malnutrition. Early detection of *stunting* is crucial for addressing this problem, which can be achieved by classifying the *stunting* status of young children. *Random Forest*, a classification method developed from *Decision Tree*, can be used to process children's health data and related factors to more accurately predict the risk of *stunting*. In this study, a dataset consisting of 10,000 records was used, with 80% allocated for training data and 20% for testing data. The *Random Forest* model achieved an accuracy of 0.7730. Subsequently, parameter optimization using *Grid Search* was employed, which successfully increased the model's accuracy to 0.9578

Keywords: *Stunting*, *Random Forest*, *Grid Search*

Abstrak: *Stunting* adalah isu kesehatan yang menjadi perhatian global. *Stunting* berkaitan erat dengan kondisi pertumbuhan dan perkembangan yang mengalami gangguan disebabkan oleh kekurangan gizi kronis. Untuk mengatasi masalah ini, diperlukan deteksi dini dengan mengklasifikasikan status *stunting* pada balita. *Random Forest* adalah sebuah metode klasifikasi yang dikembangkan dari *Decision Tree*, dapat digunakan untuk mengolah data kesehatan anak dan faktor-faktor terkait guna memprediksi risiko *stunting* dengan lebih akurat. Dalam penelitian ini, digunakan dataset yang terdiri dari 10.000 baris data, dengan 80% data sebagai data latih dan sebanyak 20% data uji. Diperoleh hasil pemodelan *Random Forest* dengan akurasi 0.7730. Selanjutnya digunakan optimasi parameter menggunakan *Grid Search* yang berhasil meningkatkan akurasi model menjadi 0.9578.

Kata kunci: *Stunting*, *Random Forest*, *Grid Search*

I. PENDAHULUAN

Stunting adalah salah satu krisis kesehatan dan kesejahteraan yang paling signifikan saat ini, dan termasuk salah satu isu kesehatan masyarakat yang paling krusial di dunia [1]. Berdasarkan data dari Perserikatan Bangsa-Bangsa (PBB) pada tahun 2020 lebih dari 149 juta yaitu sekitar 22% balita di dunia mengalami *stunting* [2]. *Stunting* didefinisikan sebagai kondisi dimana anak memiliki panjang atau tinggi badan yang pendek atau sangat pendek disebabkan buruknya asupan nutrisi dan infeksi berulang atau kronis selama 1000 hari pertama kehidupannya, yang menyebabkan kerusakan yang tidak dapat diperbaiki [3]. *Stunting* pada bayi dipengaruhi oleh beberapa faktor yaitu umur, panjang badan lahir, dan kecukupan pasokan zat gizi makro (karbohidrat, protein) dan zat gizi mikro (zat besi, vitamin A, kalsium, zinc) [4]. Pada penelitian Sutriyawan dan Nadhira pada tahun 2020 ditemukan bahwa faktor yang mempengaruhi terjadinya *stunting* meliputi pengetahuan ibu, pemberian ASI eksklusif, dan kebersihan dasar [5].

Kondisi *stunting* menyebabkan ketidaksesuaian usia dengan tinggi badan yang lebih rendah dari yang seharusnya. Anak dengan kondisi *stunting* cenderung mengalami penurunan tingkat kecerdasan, gangguan berbicara dan sulit menangkap pembelajaran dalam metode yang umum [6]. Berdasarkan hal tersebut kondisi *stunting* dapat memberikan dampak serius terhadap produktivitas di waktu dewasa, sehingga pencegahan dan penanganan masalah *stunting* menjadi prioritas kesehatan global [7].

Studi telah menunjukkan bahwa intervensi dini sangat penting untuk mencegah *stunting* dan efek buruk jangka panjangnya, seperti keterlambatan perkembangan fisik dan kognitif, yang dapat



mempengaruhi kualitas hidup anak hingga dewasa. Sayangnya, identifikasi anak yang berisiko masih sering dilakukan secara manual dan tidak efektif [8]. Maka dari itu, dibutuhkan suatu sistem yang cepat dan akurat dalam melakukan klasifikasi terhadap balita untuk mendeteksi status *stunting* atau tidak *stunting*. Berdasarkan hal tersebut, metode ilmiah dan teknologi seperti algoritma *machine learning* dapat memainkan peran yang signifikan.

Klasifikasi merupakan suatu proses menemukan sekumpulan pola atau fungsi-fungsi yang bertujuan untuk mendeskripsikan dan membagi beberapa kelas dari data [9]. Terdapat dua tahapan dalam proses klasifikasi, diantaranya yaitu tahap *learning* menggunakan data training dan tahap klasifikasi menggunakan data *testing* [10].

Salah satu algoritma *machine learning* yang dapat mengklasifikasikan *dataset* berukuran besar adalah *Random Forest*. Algoritma ini mampu menangani data dengan banyak dimensi dan berbagai skala, serta memiliki performa yang tinggi [11]. Algoritma ini merupakan bagian dari *ensemble learning* dengan memanfaatkan pohon keputusan (*decision trees*) dengan jumlah besar yang bekerja dalam melakukan prediksi atau klasifikasi secara bersamaan [12]. Penelitian menggunakan algoritma *Random Forest* sering kali diterapkan dalam metode prediksi, karena kemampuan algoritma ini untuk memprediksi nilai dengan cermat dan cepat. Selain itu, algoritma *Random Forest* juga dapat mencegah *overfitting* [13]. *Overfitting* merupakan kondisi ketika algoritma peramalan menggambarkan data sampel terlalu baik dengan hanya menghafal pola masa lalu dan demikian tidak benar-benar mempelajari pola [14].

Algoritma *Random Forest* telah banyak digunakan dalam penelitian untuk meningkatkan akurasi model klasifikasi. Sebagai contoh, penelitian oleh Riki Supriyadi et al (2020) mengenai penggunaan algoritma *Random Forest* dalam menentukan kualitas anggur merah memberikan hasil akurasi sebesar 74,68%. Selain itu, pada penelitian oleh Nugraha Listiana Hanum dan Achmad Udin Zailani (2020) diperoleh hasil akurasi dari penggunaan algoritma klasifikasi *Random Forest* dalam menentukan kelayakan pemberian kredit yaitu sebesar 87,88%. Penelitian oleh Nanik Wuryani dan Sarifah Agustani (2021) menunjukkan hasil akurasi *Random Forest* untuk mendeteksi penderita *Covid-19* berbasis Citra *CT Scan* yaitu sebesar 96,9%.

Dalam proses *machine learning* terutama klasifikasi, umumnya melibatkan pengujian berbagai model terhadap *dataset*, kemudian memilih model dengan performa terbaik. Oleh karena itu, untuk menemukan kombinasi *hyperparameter* terbaik secara cepat, penelitian ini akan menerapkan metode *Hyperparameter Tuning Grid Search*. Penelitian menggunakan metode ini pernah dilakukan oleh I Ketut Adian Jayaditya et al (2023) mengenai klasifikasi penyakit kardiovaskular. Dari penelitian tersebut menunjukkan hasil akurasi menggunakan *Grid Search* sebesar 73,06%. Selain itu, pada penelitian oleh Dimas Aryo Anggoro et al (2022) yaitu implementasi *Grid Search CV* dalam algoritma *Random Forest* pada data kanker payudara menunjukkan bahwa *Grid Search* dapat meningkatkan akurasi model dari 94,8% menjadi 95,45%.

Berdasarkan penjelasan tersebut, penelitian ini bertujuan untuk melakukan klasifikasi atau deteksi dini *stunting* pada balita menggunakan algoritma *Random Forest* dengan optimasi *Hyperparameter Tuning Grid Search*. Dalam penelitian ini juga akan membandingkan kinerja *Random Forest* sebelum dan sesudah penerapan *Hyperparameter Tuning Grid Search*. Dengan mengaplikasikan klasifikasi pada kasus *stunting*, diharapkan dapat menyediakan solusi deteksi dini yang cepat dan akurat. Temuan dari penelitian ini diharapkan dapat berkontribusi pada peningkatan kesehatan dan perkembangan anak-anak di seluruh dunia.

II. METODE PENELITIAN

II.1. Sumber Data

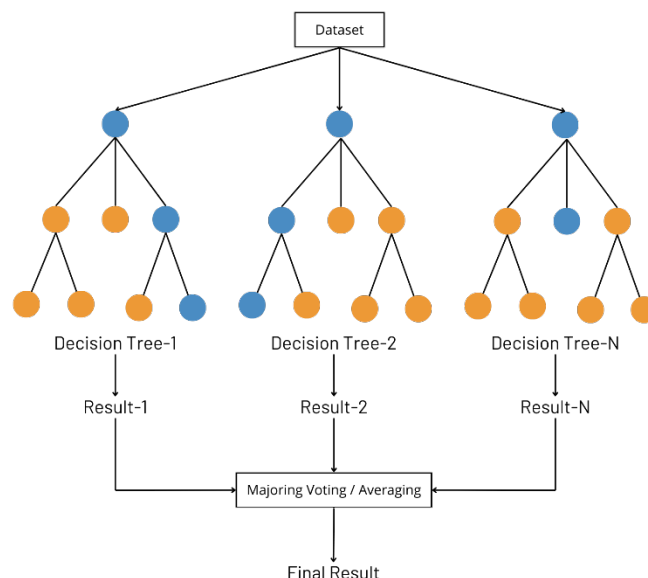
Penelitian ini adalah penelitian kuantitatif yang menggunakan data sekunder yang diperoleh dari website Kaggle. Data yang digunakan adalah *dataset* “*Stunting Balita Detection*” dengan variabel Umur (bulan), Jenis Kelamin, Tinggi Badan (cm), dan Status Gizi. Data ini mencakup informasi yang relevan untuk penelitian mengenai bagaimana mendeteksi *stunting* pada balita.

II.2. Preprocessing Data

Tahap awal yang perlu dilakukan sebelum melatih model, perlu dilakukan *preprocessing data* untuk memastikan kinerja model tidak terganggu. *Preprocessing data* dalam penelitian ini meliputi penghilangan data duplikat, penghilangan data pencilan, dan pemberian label *encoder* pada data kategorikal [15]. Kemudian setelah langkah ini selesai, diputuskan bahwa data dibagi menjadi 80% data *training* dan 20% data *testing*.

II.3. Random Forest

Random Forest adalah metode yang dikembangkan dari metode *Decision Tree*. Setiap pohon keputusan dilatih menggunakan sampel individu, dan masing-masing atribut dipecah pada pohon dipilih secara acak dari atribut subset [15]. Terdapat sejumlah kelebihan dari *Random Forest*, antara lain dapat meningkatkan akurasi hasil ketika menghadapi data yang hilang dan *resisting outliers*. Selain itu algoritma ini efisien dalam penyimpanan data. Kelebihan lainnya adalah kemampuan *Random Forest* untuk mengoptimalkan performa model klasifikasi melalui proses seleksi fitur yang efektif dalam memilih fitur terbaik [16]. Alur dalam pembuatan pohon keputusan pada *Random Forest* disajikan pada gambar 1[1].



Gambar 1. Visualisasi *Random Forest*

Rumus *Entropy*

$$Entropy(s) = \sum_{i=1}^n - p_i \log_2 p_i \quad (1)$$



Keterangan :

- S : Himpunan kasus
- A : Fitur
- n : Jumlah partisi S
- pi : Proporsi sampel untuk kelas i

Rumus Gain

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|Sv|}{|S|} * \text{Entropy}(Si) \text{Entropy}(Sv) \quad (2)$$

Keterangan :

- A : Atribut
- V : Menyatakan suatu nilai yang mungkin untuk atribut A
- |Si| : Jumlah kasus pada partisi ke $-i$
- |S| : Jumlah seluruh sampel data
- Entropy(Si) : Entropy untuk sampel-sampel yang memiliki nilai i

II.4. Grid Search

Grid Search adalah metode eksplorasi matematis untuk menemukan *hyperparameter* terbaik guna meningkatkan kinerja model klasifikasi. Dengan mencoba seluruh kombinasi *hyperparameter* yang telah ditentukan sebelumnya, metode ini dapat memilih kombinasi *hyperparameter optimal* yang menghasilkan performa terbaik dari model [17]. Proses ini sering dikombinasikan dengan *k-fold cross-validation* untuk menentukan *hyperparameter* terbaik dan kombinasi ini dikenal dengan sebutan *Grid Search Cross-Validation* atau *GridSearchCV* [18].

II.5. Confusion Matrix

Confusion Matrix merupakan alat evaluasi dalam machine learning yang digunakan untuk mengukur kinerja model klasifikasi. *Confusion Matrix* terdiri dari empat komponen utama: *True Positive (TP)*, *False Positive (FP)*, *True Negative (TN)*, dan *False Negative (FN)*. *Confusion matrix* memberikan dasar untuk berbagai metrik evaluasi, seperti akurasi, presisi, dan *recall*. Akurasi menggambarkan sejauh mana prediksi sesuai dengan nilai aktual. Presisi menunjukkan proporsi prediksi positif yang tepat, sementara *recall* mengukur proporsi nilai positif yang berhasil diprediksi dengan tepat. Semua metrik ini penting untuk menilai kualitas model klasifikasi [19].

Tabel 1. Confusion Matrix

Klasifikasi	Kelas Prediksi			
	Kelas 1	Kelas 2	Kelas 3	Kelas 4
Kelas 1	x_{11}	x_{12}	...	x_{1n}
Kelas 2	x_{21}	x_{22}	...	x_{2n}
...	x_{31}	x_{32}
Kelas 3	x_{n1}	x_{n2}	...	x_{nn}



Formula *confusion matrix* pada klasifikasi *multiclass* sebagai berikut [20].

$$TFN_i = \sum_{j=1, j \neq i}^n x_{ij} \quad (3)$$

$$TFP_i = \sum_{j=1, j \neq i}^n x_{ji} = 1, j \neq i \quad (4)$$

$$TTN_i = \sum_{j=1, j \neq i}^n \sum_{k=1, k \neq i}^n x_{jk} \quad (5)$$

$$TTP_{all} = \sum_j^n x_{jj} \quad (6)$$

Tiga indeks pengukuran yang digunakan pada confusion matrix adalah akurasi, sensitivitas, dan spesifisitas. Formula yang digunakan sebagai berikut [21] :

$$Akurasi_i = \frac{TTP_i + TTN_i}{Total\ Number\ of\ testing\ entries} \times 100\% \quad (7)$$

$$Sensitivitas_i = \frac{TTP_i}{TTP_i + TFP_i} \times 100\% \quad (8)$$

$$Spesifisitas_i = \frac{TTN_i}{TTN_i + TFP_i} \times 100\% \quad (9)$$

Keterangan:

Total *True Positive* (TTP) jumlah data yang diklasifikasikan dengan benar sebagai positif

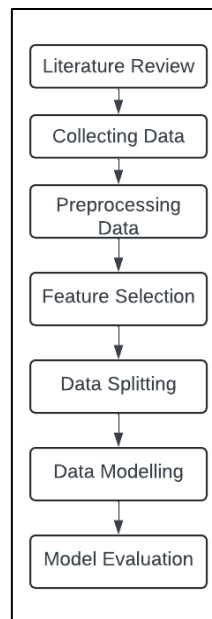
Total *True Negative* (TTN) jumlah data yang diklasifikasikan dengan benar sebagai negatif

Total *False Positive* (TFP) jumlah data yang diprediksi sebagai positif secara salah

Total *False Negative* (TFN) jumlah data yang diprediksi sebagai negatif secara salah

II.6. Tahapan Penelitian

Pada penelitian ini digunakan beberapa tahapan yang terlihat pada gambar 2 [1].



Gambar 2. Diagram Alir *Random Forest*

III. HASIL DAN PEMBAHASAN

III.1. *Preprocessing Data*

Pada tahap *Preprocessing* data dilakukan pembersihan data. Setelah data dibersihkan, terdapat 1.235 data yang terduplikasi sehingga data yang digunakan menjadi 8.765 data dengan variabel prediktor yaitu Umur (bulan), Jenis Kelamin dan Tinggi Badan (cm). Sedangkan variabel label yaitu Status Gizi.

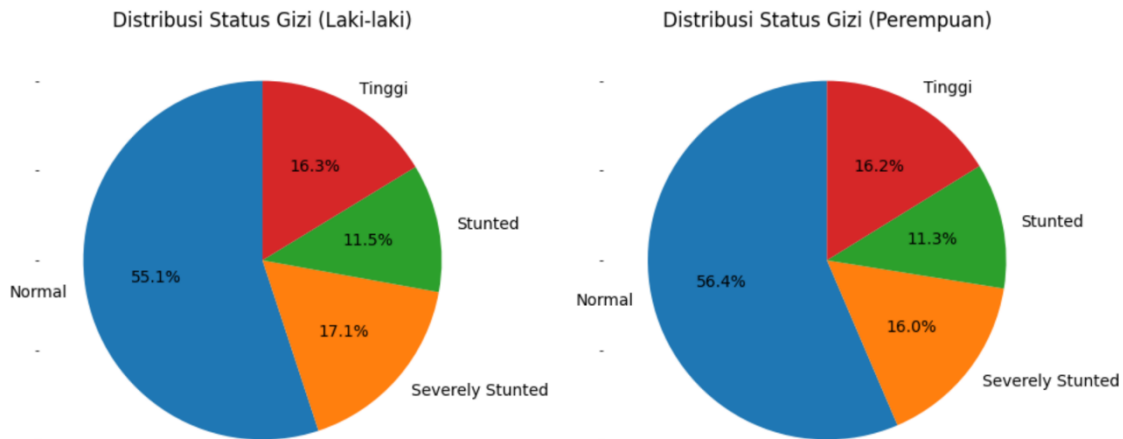
III.2. *Statistik Deskriptif*

Setelah dilakukan *preprocessing* data, diperlukan statistik deskriptif. Data yang digunakan terdiri dari 2 variabel numerik yaitu variabel Umur (bulan) & variabel Tinggi Badan (cm) dan 2 variabel kategorik yaitu variabel Status Gizi dan Jenis Kelamin.

Tabel 2. Statistik Deskriptif Variabel Umur dan Tinggi

Variabel	Mean	Standar Deviasi	Min	Max
Umur	29.924929	17.709007	0.000000	60.000000
Tinggi	88.375677	17.529683	40.058140	127.900000

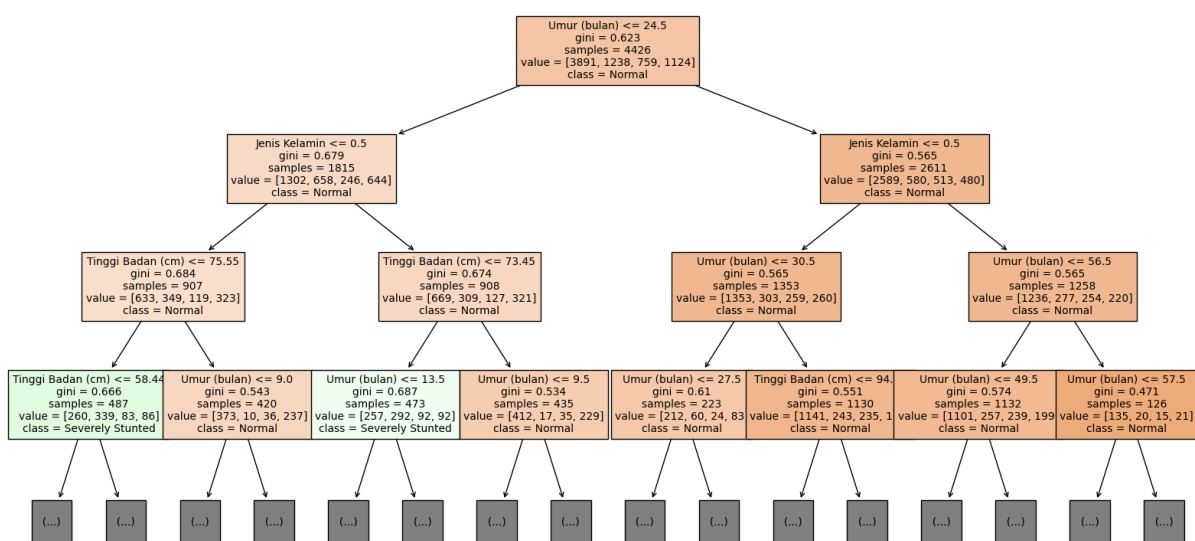
Dilihat pada tabel 2 diperoleh nilai *mean* variabel Umur (bulan) adalah 29,924929 dengan standar deviasi sebesar 17,709007. Nilai *mean* variabel Tinggi Badan (cm) adalah 88,375677 dengan standar deviasi sebesar 17,529683. Data variabel Tinggi Badan (cm) terkecil adalah 40,058140 dan terbesar sebesar 127,9 sedangkan nilai terkecil pada variabel Umur (bulan) adalah 0 atau bayi yang baru lahir dan terbesar adalah 60 bulan.



Gambar 3. Diagram Distribusi Variabel Kategorik

Distribusi data pada variabel kategorik Jenis Kelamin dan dan Status Gizi dapat dilihat pada gambar 3. Jumlah laki-laki adalah 4.364 dengan distribusi Status Gizi kategori Normal sebesar 55,1% dengan jumlah sebanyak 2.403, presentase Status Gizi kategori *Severely Stunted* sebesar 17,1% dengan jumlah sebanyak 746, presentase Status Gizi kategori *Stunted* sebesar 11,5% dengan jumlah sebanyak 504 dan presentase Status Gizi kategori Tinggi sebesar 16,3% dengan jumlah sebanyak 711. Sedangkan, jumlah perempuan adalah 4.401 dengan distribusi Status Gizi kategori normal sebesar 56,4% dengan jumlah sebanyak 2.484, presentase Status Gizi kategori *Severely Stunted* sebesar 16% dengan jumlah sebanyak 706, presentase Status Gizi kategori *Stunted* sebesar 11,3% dengan jumlah sebanyak 498 dan presentase Status Gizi kategori Tinggi sebesar 16,2% dengan jumlah sebanyak 713.

III.3. Klasifikasi Random Forest



Gambar 4. Visualisasi Klasifikasi Random Forest

Pada klasifikasi *Random Forest* ini menggunakan *hyperparameter* dengan *max_depth* = 5, *n_estimator* = 100, *min_sample_split* = 2 dan *min_sample_leaf* = 1. Didapatkan visualisasi klasifikasi



Random Forest pada gambar 4 menunjukkan bahwa *node* akar pertama kali membagi data berdasarkan variabel Umur.

Tabel 3. *Confusion Matrix* Klasifikasi *Random Forest*

Klasifikasi		Kelas Prediksi			
		Normal	<i>Severely Stunted</i>	<i>Stunted</i>	Tinggi
Kelas Aktual	Normal	961	1	3	4
	<i>Severely Stunted</i>	16	253	2	0
	<i>Stunted</i>	139	44	37	0
	Tinggi	189	0	0	104

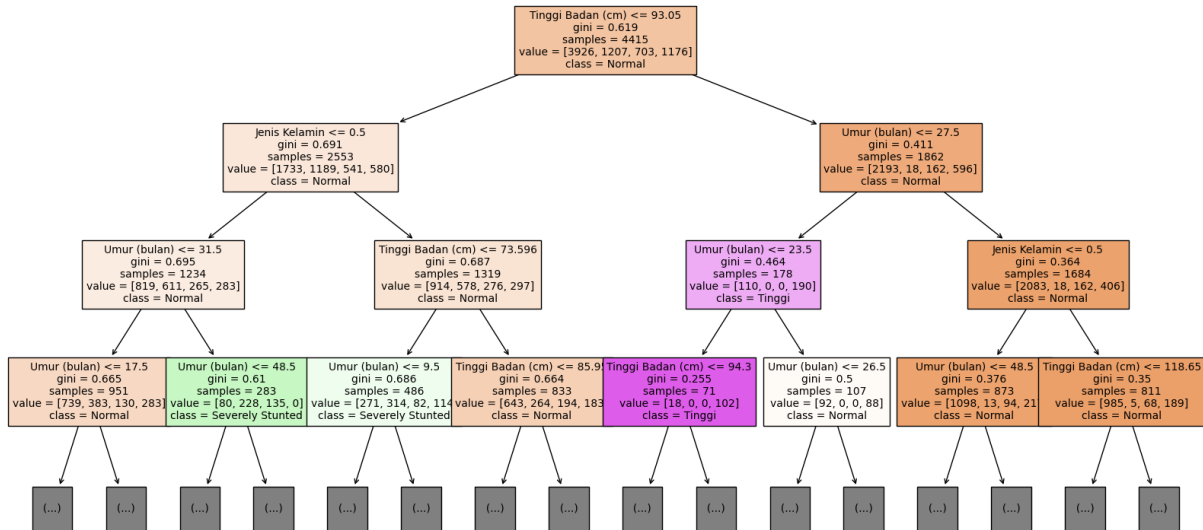
Tabel 3 menunjukkan hasil *confusion matrix* menggunakan data testing setelah model dilatih menggunakan klasifikasi *Random Forest*. Model ini berhasil mencapai tingkat akurasi sebesar 0.7730, dengan nilai *Out-of-Bag (oob) score* sebesar 0.7808.

III.4. Klasifikasi *Random Forest* Menggunakan *Hyperparameter Tuning Grid Search*

Dalam analisis ini, dilakukan penyetelan *hyperparameter* untuk menemukan nilai optimal untuk *max_depth*, *n_estimator*, *min_samples_split*, dan *min_samples_leaf* dalam model klasifikasi *Random Forest*. Berbagai nilai yang dicoba untuk setiap parameter antara lain *max_depth* dengan nilai 5, 10, 20, 30, dan 40, *n_estimator* dengan nilai 10, 30, 50, 100, dan 200, *min_samples_split* dengan nilai 2, 5, 10, 15, dan 20, serta *min_samples_leaf* dengan nilai 1, 2, 4, 6, dan 10. Hasil dari penyetelan *hyperparameter* ini menunjukkan nilai optimal yang dapat dilihat pada tabel 4.

Tabel 4. *Hyperparameter* Optimal

<i>max_depth</i>	<i>n_estimator</i>	<i>min_samples_split</i>	<i>min_samples_leaf</i>
20	50	2	1



Gambar 5. Visualisasi Klasifikasi *Random Forest* Menggunakan *Hyperparameter Tuning Grid Search*

Setelah mengidentifikasi *hyperparameter* optimal melalui penyetulan grid search, dilakukan pembangunan model klasifikasi *Random Forest*. Visualisasi dari model ini menggunakan hasil penyetulan *hyperparameter* dapat dilihat pada gambar 5, menunjukkan bahwa *node* akar pertama kali membagi data berdasarkan variabel Tinggi Badan.

Tabel 5. Confusion Matrix dengan Grid Search

Klasifikasi		Kelas Prediksi			
		Normal	Severely Stunted	Stunted	Tinggi
Kelas Aktual	Normal	942	0	15	12
	Severely Stunted	0	262	9	0
	Stunted	16	7	197	0
	Tinggi	15	0	0	278

Pada tabel 5, tertera *confusion matrix* yang menggambarkan performa model terhadap data *testing* setelah melalui penyetulan *hyperparameter*. Model ini berhasil mencapai tingkat akurasi sebesar 0.9578, yang menunjukkan peningkatan signifikan dibandingkan dengan akurasi model klasifikasi *Random Forest* sebelumnya.

II.5. Perbandingan Model Klasifikasi *Random Forest* dan *Random Forest* Menggunakan *Hyperparameter Tuning Grid Search*



Tabel 6. Perbandingan Model

Model	Akurasi	Presisi	Recall	F1 Score
Random Forest	0.7730	0.8573	0.6121	0.6339
Random Forest Menggunakan Hyperparameter Tuning Grid Search	0.9578	0.9480	0.9458	0.9469

Dari hasil analisis diatas dapat dilihat perbandingan akurasi, presisi, dan *recall* pada tabel 6 yang menunjukkan bahwa hasil pemodelan menggunakan *hyperparameter tuning grid search* menunjukkan nilai akurasi yang lebih tinggi dibandingkan dengan hasil pemodelan sebelumnya, sehingga dapat disimpulkan bahwa penggunaan *hyperparameter tuning grid search* pada klasifikasi *Random Forest* dapat memberikan model yang lebih baik.

IV. KESIMPULAN

Berdasarkan hasil analisis yang telah dilakukan diatas dapat disimpulkan bahwa pada dataset yang terdiri dari 10.000 baris data, dengan 80% digunakan sebagai data *training* dan 20% sebagai data *testing* dengan menggunakan klasifikasi *Random Forest* mendapat akurasi sebesar 0.7730 sedangkan model klasifikasi *Random Forest* menggunakan *Hyperparameter Tuning Grid Search* mendapat akurasi 0.9578. Dengan demikian, dapat dikatakan bahwa klasifikasi *Random Forest* menggunakan *Hyperparameter Tuning Grid Search* dapat memberikan model yang lebih baik.

UCAPAN TERIMA KASIH

Penulis mengucapkan terimakasih kepada kepada Laboratorium Riset Kesehatan dan Lingkungan, Jurusan S1 Statistika, Fakultas Sains dan Teknologi Pertanian, Universitas Muhammadiyah Semarang yang telah memfasilitasi pengolahan data serta pembuatan artikel ini. Selain itu, kami ucapkan terima kasih kepada Bapak Fatkhurokhman Fauzi, M.Stat yang telah menjadi dosen pembimbing penyusunan artikel ilmiah ini serta orang tua yang selalu mendoakan kami.

REFERENSI

1. M. S. Haris, M. Anshori, and A. N. Khudori, “Prediction of Stunting Prevalence in East Java Province With *Random Forest* Algorithm Prediksi Nilai Prevalensi Stunting Di Provinsi Jawa Timur Menggunakan Algoritma *Random Forest*,” *Jurnal Teknik Informatika (JUTIF)*, vol. 4, no. 1, pp. 11–13, 2023, [Online]. Available: <https://doi.org/10.20884/1.jutif.2023.4.1.614>
2. Eko, “149 Juta Anak di Dunia Alami Stunting Sebanyak 6,3 Juta di Indonesia, Wapres Minta Keluarga Prioritaskan Kebutuhan Gizi,” Sumber Siaran Pers Sekretariat Negara. [Online]. Available: <https://pauddpedia.kemdikbud.go.id/berita/149-juta-anak-di-dunia-alami-stunting-sebanyak-63-juta-di-indonesia-wapres-minta-keluarga-prioritaskan-kebutuhan-gizi?do=MTY2NC01YjRhOGZkNA==&ix=MTEtYmJkNjQ3YzA=>
3. D. F. Susanti, “Mengenal Apa Itu Stunting,” Direktorat Jenderal Pelayanan Kesehatan. [Online]. Available: https://yankes.kemkes.go.id/view_artikel/1388/mengenal-apa-itu-stunting
4. E. T. Siringoringo, A. Syauqy, B. Panunggal, R. Purwanti, and N. Widyastuti, “Karakteristik Keluarga Dan Tingkat Kecukupan Asupan Zat Gizi Sebagai Faktor Risiko Kejadian Stunting Pada Baduta,” *Journal of Nutrition College*, vol. 9, no. 1, pp. 54–62, 2020, doi: 10.14710/jnc.v9i1.26693.
5. A. Sutriyawan, R. D. Kurniawati, R. Hanjani, and S. Rahayu, “Prevalensi Stunting Dan Hubungannya Dengan Sosial Ekonomi,” *Jurnal Kesehatan*, vol. 11, no. 3, p. 351, 2021, doi: 10.35730/jk.v11i3.636.
6. I. Ali, D. Ade Kurnia, M. A. Pratama, and F. Al Ma’ruf, “Klasifikasi Status Stunting Balita Di Desa Slangit Menggunakan Metode K-Nearest Neighbor,” *Jurnal Ilmiah Manajemen Informatika dan Komputer*, vol. 05, no. 03, pp. 35–38, 2021, [Online]. Available: <http://jurnal.kopertipindonesia.or.id/index.php/kopertip>
7. F. A. Nita, E. Ernawati, F. Sari, J. J. Kristiari, and I. Purnamasari, “The influence of parenting on the incidence of stunting in toddlers aged 1-3 year,” *Jurnal Ilmiah Kesehatan Sandi Husada*, vol. 12, no. 2, pp. 399–405, 2023, doi: 10.35816/jiskh.v12i2.1107.



8. M. S. Haris, M. Anshori, and A. N. Khudori, “Prediction of Stunting Prevalence in East Java Province With *Random Forest* Algorithm Prediksi Nilai Prevalensi Stunting Di Provinsi Jawa Timur Menggunakan Algoritma *Random Forest*,” *Jurnal Teknik Informatika (JUTIF)*, vol. 4, no. 1, pp. 11–13, 2023, [Online]. Available: <https://doi.org/10.20884/1.jutif.2023.4.1.614>
9. P. B. N. Setio, D. R. S. Saputro, and Bowo Winarno, “Klasifikasi Dengan Pohon Keputusan Berbasis Algoritma C4.5,” *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 3, pp. 64–71, 2020.
10. S. Lonang and D. Normawati, “Klasifikasi Status Stunting Pada Balita Menggunakan K-Nearest Neighbor Dengan Feature Selection Backward Elimination,” *Jurnal Media Informatika Budidarma*, vol. 6, no. 1, p. 49, 2022, doi: 10.30865/mib.v6i1.3312.
11. A. Nugroho, I. Asror, and Y. F. A. Wibowo, “Klasifikasi Tingkat Kualitas Udara DKI Jakarta Berdasarkan Open Government Data Menggunakan Algoritma *Random Forest*,” *eProceedings of Engineering*, vol. 10, No. 2, no. 2, pp. 1824–1834, 2023, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/20030%0Ahttps://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/20030/19395>
12. I. P. Putri, T. Terttiaavini, and N. Arminarahmah, “Analisis Perbandingan Algoritma Machine Learning untuk Prediksi Stunting pada Anak,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 257–265, 2024, doi: 10.57152/malcom.v4i1.1078.
13. Fadellia Azzahra, N. Suarna, and Y. Arie Wijaya, “Penerapan Algoritma *Random Forest* Dan Cross Validation Untuk Prediksi Data Stunting,” *Kopertip : Jurnal Ilmiah Manajemen Informatika dan Komputer*, vol. 8, no. 1, pp. 1–6, 2024, doi: 10.32485/kopertip.v8i1.238.
14. M. Czajkowski and M. Kretowski, “Decision tree underfitting in mining of gene expression data. An evolutionary multi-test tree approach,” *Expert Syst Appl*, vol. 137, pp. 392–404, 2019, doi: <https://doi.org/10.1016/j.eswa.2019.07.019>.
15. H. Marlina, A. Zulus, and O. L. W. H. Wijaya, “Penerapan Algoritma *Random Forest* Dalam Klasifikasi Jurusan Di SMA Negeri Tugumulyo,” *BRAHMANA: Jurnal Penerapan Kecerdasan Buatan Terakreditasi*, vol. 4, no. 2, pp. 138–143, 2023, [Online]. Available: <https://tunasbangsa.ac.id/pkm/index.php/brahmana/article/view/188>
16. S. Devella, Y. Yohannes, and F. N. Rahmawati, “Implementasi *Random Forest* Untuk Klasifikasi Motif Songket Palembang Berdasarkan SIFT,” *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 7, no. 2, pp. 310–320, 2020, doi: 10.35957/jatisi.v7i2.289.
17. M. Fajri and A. Primajaya, “Komparasi Teknik Hyperparameter Optimization pada SVM untuk Permasalahan Klasifikasi dengan Menggunakan Grid Search dan Random Search,” *Journal of Applied Informatics and Computing*, vol. 7, no. 1, pp. 14–19, 2023, doi: 10.30871/jaic.v7i1.5004.
18. A. Toha, P. Purwono, and W. Gata, “Model Prediksi Kualitas Udara dengan Support Vector Machines dengan Optimasi Hyperparameter GridSearch CV,” *Buletin Ilmiah Sarjana Teknik Elektro*, vol. 4, no. 1, pp. 12–21, 2022, doi: 10.12928/biste.v4i1.6079.
19. A. A. M. Ramadani, S. Lestanti, and S. Kirom, “Klasifikasi Balita Stunting Menggunakan *Random Forest* Classifier Di Kabupaten Blitar,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 6, pp. 3846–3851, 2024, doi: 10.36040/jati.v7i6.7822.
20. D. Retno Mutia, N. Kumalasari, C. Pratiwi, and Y. N. Fu’adah, “Perancangan Identifikasi Penyakit Pada Retina Berbasis Pengolahan Citra Optical Coherence Tomography (Oct) Menggunakan Convolutional Neural Network (Cnn) Identification of Diseases in Retina Based on Optical Coherence Tomography (Oct) Image Using Convolut,” vol. 8, no. 6, pp. 11551–11559, 2021, [Online]. Available: www.kaggle.com,
21. Qorry Meidianingsih, D. E. Wardani, E. Salsabila, L. Nafisah, and A. N. Mutia, “Perbandingan Performa Metode Berbasis Support Vector Machine untuk Penanganan Klasifikasi Multi Kelas Tidak Seimbang,” *STATISTIKA Journal of Theoretical Statistics and Its Applications*, vol. 23, no. 1, pp. 8–18, 2023, doi: 10.29313/statistika.v23i1.1660.