Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1537−1546
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6128

Analisa Optimasi Grid Search pada Algoritma Random Forest dan Decision Tree untuk Klasifikasi Stunting

Ririt Sheila Tina Rahmayani*, Fikri Budiman

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia
Email: [1,*]111202113802@mhs.dinus.ac.id, [2]fikri.budiman@dsn.dinus.ac.id
Email Penulis Korespondensi: 111202113802@mhs.dinus.ac.id
Submitted: **31/10/2024**; Accepted: **01/12/2024**; Published**: 03/12/2024**

Abstrak−Stunting merupakan masalah serius yang menjadi perhatian global karena dampaknya yang signifikan terhadap kesehatan dan pertumbuhan balita. Kondisi ini terjadi akibat kekurangan gizi dalam waktu lama. Di Indonesia, permasalahan gizi masih umum ditemukan, termasuk stunting yang memengaruhi pertumbuhan serta perkembangan anak. Berkaitan dengan hal tersebut, data mining memiliki peran penting dalam menghadapi tantangan ini. Oleh karena itu, tujuan dari penelitian ini untuk mengoptimalkan klasifikasi stunting menggunakan algoritma Decision Tree dan Random Forest yang dioptimalkan dengan Grid Search. Pengoptimalan ini dilakukan untuk meningkatkan akurasi kedua algoritma serta mengidentifikasi algoritma yang lebih unggul dalam menentukan stunting. Dataset yang digunakan berjumlah 10.000 data balita dengan atribut penting terkait kondisi kesehatan. Hasil analisis menunjukkan bahwa model Decision Tree awal memiliki akurasi sebesar 70,2%. Setelah dilakukan optimasi menggunakan Grid Search, akurasi model Decision Tree meningkat signifikan menjadi 82,8%. Sementara itu, model Random Forest awal mencapai akurasi 77,9%, dan setelah dioptimasi dengan Grid Search, akurasinya meningkat lebih tinggi dibandingkan dengan Decision Tree, yaitu sebesar 84,1%. Peningkatan ini mencerminkan efektivitas optimasi dalam meningkatkan kemampuan model dalam mengklasifikasikan stunting secara lebih akurat. Penelitian ini memberikan wawasan penting mengenai efektivitas kedua algoritma dalam mengidentifikasi stunting serta menekankan pentingnya optimasi untuk meningkatkan akurasi klasifikasi, yang dapat mendukung intervensi tepat bagi kesejahteraan generasi mendatang.

Kata Kunci: Decision Tree; Grid Search; Optimasi; Random Forest; Stunting

Abstract−Stunting is a serious problem that is of global concern because of its significant impact on the health and growth of children under five. This condition occurs due to long-term malnutrition. In Indonesia, nutritional problems are still common, including stunting which affects children's growth and development. In this regard, data mining has an important role in facing this challenge. Therefore, the aim of this research is to optimize stunting classification using Decision Tree and Random Forest algorithms optimized with Grid Search. This optimization was carried out to increase the accuracy of the two algorithms and identify algorithms that are superior in determining stunting. The dataset used consists of 10,000 toddler data with important attributes related to health conditions. The analysis results show that the initial Decision Tree model has an accuracy of 70.2%. After optimization using Grid Search, the accuracy of the Decision Tree model increased significantly to 82.8%. Meanwhile, the initial Random Forest model achieved an accuracy of 77.9%, and after optimization with Grid Search, its accuracy increased even higher compared to Decision Tree, namely 84.1%. This increase reflects the effectiveness of optimization in increasing the model's ability to classify stunting more accurately. This research provides important insights into the effectiveness of both algorithms in identifying stunting and emphasizes the importance of optimization to improve classification accuracy, which can support appropriate interventions for the well-being of future generations.

Keywords: Decision Trees; Grid Search; Optimization; Random Forest; Stunting

1. PENDAHULUAN

Stunting adalah masalah yang menunjukkan gangguan perkembangan pada anak yang dikarenakan kurangnya gizi dalam waktu panjang yang masih menjadi masalah besar di Indonesia [1][2]. Kondisi tersebut menjadi salah satu permasalahan serius dalam kesehatan anak balita yang telah menjadi fokus perhatian global. Menurut data dari situs paudpedia.kemdikbud.go.id, angka *stunting* di Indonesia tercatat tinggi, yaitu mencapai 24,4% yang dimana melebihi standar toleransi WHO yaitu di bawah 20% [3]. Fenomena *stunting* menciptakan dampak yang signifikan terutama bagi pertumbuhan dan kesehatan anak. *Stunting* menunjukkan adanya defisiensi asupan gizi jangka panjang selama fase pertumbuhan dan perkembangan di awal kehidupan. Berdasarkan standar pertumbuhan WHO, stunting dievaluasi melalui pengukuran panjang atau tinggi badan relatif terhadap usia (TB/U), di mana nilainya berada di bawah minus dua standar deviasi (SD) [4]. Secara global, sekitar seperempat balita mengalami *stunting* [5]. WHO memperkirakan bahwa sekitar 22,2% atau sekitar 149,2 juta balita di seluruh dunia mengalami kondisi ini [6]. Angka *stunting* di Indonesia yang mencapai 21,6% masih dianggap tinggi karena melebihi 20%, sesuai dengan standar yang ditetapkan oleh WHO [7][8].

Mengingat betapa seriusnya dampak *stunting* serta tingginya prevalensi kasus di Indonesia, penanganan masalah ini menjadi sangat penting demi kesejahteraan generasi mendatang. Permasalahan *stunting* semakin kompleks karena kurangnya kesadaran dan kendala identifikasi yang tidak akurat. Evaluasi *stunting* memerlukan pendekatan komprehensif yang meliputi pengukuran antropometri, evaluasi kesehatan, pemeriksaan gizi, asesmen psikososial, dan riwayat perkembangan anak [9][10]. Namun, ketidaklengkapan data sering menghambat deteksi pola dan memperlambat intervensi. Untuk mengatasi masalah ini secara efektif, diperlukan metode pengumpulan dan analisis data yang lebih canggih serta infrastruktur yang lebih baik untuk menilai *stunting*. Penggunaan data mining sebagai model klasifikasi otomatis dapat mengurangi subjektivitas dan meminimalkan kemungkinan salah dalam pengambilan keputusan secara manual. Hal ini akan meningkatkan akurasi dan efisiensi dalam mendeteksi kasus *stunting.*

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1537−1546
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6128

Data mining merupakan proses yang digunakan untuk menemukan informasi dalam dataset yang telah dipilih, dengan menerapkan pendekatan atau strategi tertentu [11][12]. Teknik data mining sangat penting dalam perkembangan teknologi saat ini karena kemampuannya menghasilkan keputusan yang lebih baik dan mengekstraksi wawasan baru dari dataset mentah [13]. Salah satu teknik analisis dalam data mining adalah klasifikasi. Dengan menggunakan klasifikasi, data terkait *stunting* akan tersusun dengan lebih baik dan diberikan label yang mencerminkan karakteristiknya. Hal ini dapat mempermudah pengolahan data lebih lanjut serta meningkatkan efisiensi dan akurasi identifikasi kasus, sehingga penanganan dapat dilakukan dengan lebih tepat.

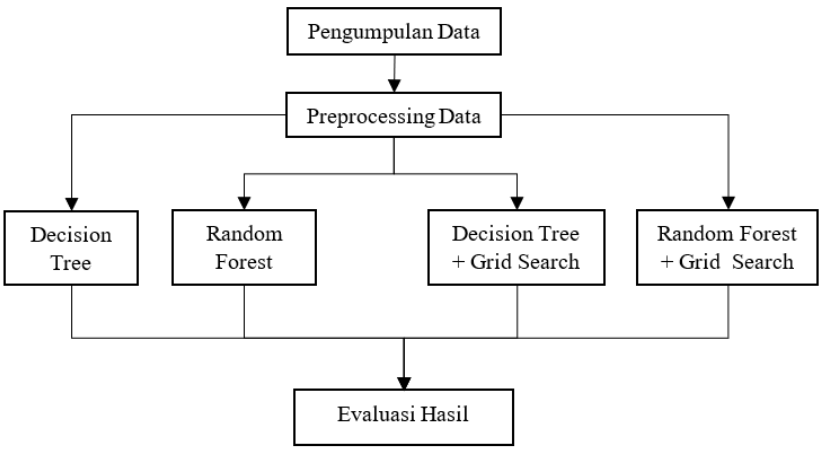
Dalam mengembangkan model ini, dilakukan percobaan guna meningkatkan kinerja klasifikasi pada dataset *stunting* menggunakan dua algoritma, yaitu *Decision Tree* serta *Random Forest*, yang dioptimalkan melalui *Grid Search*. Perbandingan antara kedua algoritma yang dioptimasi menggunakan *Grid Search* penting untuk mengetahui kelebihan dan kekurangan dari setiap algoritma, karena keduanya menerapkan pendekatan yang berbeda dalam melakukan klasifikasi. Kedua algoritma ini relevan untuk klasifikasi *stunting* karena mampu menangani data kesehatan yang kompleks. *Decision Tree* memberikan interpretasi yang jelas pada model, sehingga memudahkan pemahaman keputusan berbasis fitur. Sementara *Random Forest* dapat meningkatkan akurasi dengan menggabungkan prediksi dari beberapa pohon keputusan untuk mengurangi overfitting [14]. *Grid Search* digunakan untuk menemukan kombinasi hyperparameter terbaik, seperti jumlah pohon pada *Random Forest* atau kedalaman maksimal *Decision Tree*, yang dapat meningkatkan akurasi model dan memastikan algoritma bekerja optimal [15]. Perbandingan ini memberikan pemahaman yang lebih mendalam mengenai kinerja kedua algoritma, membantu memahami keunggulan serta kelemahan masing-masing dalam klasifikasi *stunting*, dan mendukung intervensi yang lebih efektif [16].

Studi penelitian terdahulu tentang klasifikasi telah diimplementasikan dengan berbagai jenis algoritma. Penelitian yang dilakukan oleh Putri Handayani, Abd.Charis Fauzan dan Harliana menggunakan Algoritma *Random Forest* yang menghasilkan akurasi 88,6% [17]. Penelitian lain yang membandingkan prediksi *stunting* dengan algoritma *Support Vektor Machine* dan *Random Forest* dilakukan oleh Yunada Wiratama dan RZ Abdul Aziz dengan hasil akurasi *Random Forest* lebih tinggi daripada SVM yaitu sebesar 0,9997 untuk *Random Forest* dan SVM sebesar 0,9951 [18]. Selain itu terdapat penelitian yang dilakukan oleh Mario Utomo dan Rastri Prathivi yang melakukan perbandingan Algoritma *Support Vector Machine* dan *Decision Tree* dengan hasil akurasi SVM sebesar 86,67% dan untuk *Decision Tree* sebesar 93,33% [19]. Penelitian lain yang dilakukan oleh Hery menghasilkan akurasi sebesar 8l,85% dengan menggunakan algoritma *Naive Bayes* [20]. Penelitian lainnya dilakukan oleh Syahrani menggunakan *K-Nearest Neighbor* dengan *Backward Elimination* menghasilkan akurasi sebesar 92,20% [21].

Berdasarkan berbagai penelitian, beberapa algoritma diketahui mampu menghasilkan akurasi tinggi dalam klasifikasi *stunting* pada balita. Namun, sampai saat ini belum ada yang menerapkan model *Grid Search* untuk mengoptimalkan dua algoritma, yaitu *Decision Tree dan Random Forest*, dengan tujuan meningkatkan akurasi pada kedua algoritma. Penelitian sebelumnya hanya menggunakan algoritma secara individu tanpa proses optimasi. Sebagai lanjutan dari penelitian tersebut, studi ini menggunakan pendekatan lebih komprehensif dengan menerapkan dua algoritma serta melakukan optimasi untuk meningkatkan akurasi. Penelitian ini bertujuan untuk membangun model klasifikasi *stunting* menggunakan algoritma *Random Forest* dan *Decision Tree* yang dioptimalkan melalui *Grid Search* guna mencapai akurasi yang tinggi. Penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam mendeteksi stunting dengan menghasilkan kesimpulan yang lebih akurat.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian



Gambar 1. Alur Penelitian

Berdasarkan alur penelitian seperti Gambar 1 di atas, langkah awal yang dilakukan adalah mengumpulkan data relevan, kemudian preprocessing data untuk memastikan bahwa data yang digunakan memiliki kualitas yang siap

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1537−1546
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6128

untuk diolah. Selanjutnya, model klasifikasi dibangun dengan memanfaatkan algoritma Decision Tree dan Random Forest kemudian masing masing algoritma dioptimasi menggunakan metode Grid Search untuk menemukan kombinasi hyperparameter terbaik. Langkah terakhir yaitu evaluasi hasil penelitian dilakukan menggunakan confusion matrix untuk menghitung akurasi kinerja model. Berikut penjelasan mendetail dari setiap bagian alur penelitian.

2.2 Pengumpulan Data

Data yang digunakan adalah dataset publik tentang stunting pada balita yang diperoleh dari Kaggle, yang dapat diakses melalui www.kaggle.com. Dataset yang di ambil berjumlah 10.000 data dengan 8 atribut dan 2 kelas. Tabel 1 merupakan spesifikasi atribut beserta jenis datanya.

Tabel 1. Atribut dan Tipe Data

Atribut	Tipe Data
Gender	Teks
Age	Numerik
Birth Weight	Numerik
Birth Length	Numerik
Body Weight	Numerik
Body Length	Numerik
Breastfeeding	Teks
Stunting	Teks

Dalam dataset yang di ambil terdapat informasi data yang digunakan yaitu, atribut Gender menunjukkan jenis kelamin anak, "Male" berarti laki-laki dan "Female" berarti perempuan. Age menunjukkan umur dalam rentang 6 hingga 48 bulan pada saat pengukuran dilakukan. Kemudian Birth Weight menunjukkan berat badan anak saat lahir, yang berkisar antara 2 hingga 3.1 kg. Birth Length menunjukkan panjang badan bayi saat lahir, berkisar antara 48 hingga 50 cm. Body Weight menunjukkan berat badan berkisar antara 2.9 hingga 10.5 kg. Body Length menunjukkan panjang badan berkisar antara 49 hingga 92.7 cm. Breastfeeding menunjukkan apakah anak mendapatkan ASI atau tidak, dengan "Yes" jika anak menerima ASI dan "No" jika tidak. Atribut terakhir yaitu status stunting anak, dengan "Yes" menunjukkan anak mengalami stunting, dan "No" berarti anak tidak mengalami stunting. Klasifikasi ini menjadi label target dalam klasifikasi stunting.

2.3 Preprocessing Data

Dataset yang digunakan perlu melalui tahap preprocessing data, yang merupakan proses awal dalam pengolahan data untuk mempersiapkan data mentah menjadi format yang cocok untuk digunakan dalam algoritma data mining. Tahapan dalam preprocessing data meliputi :

a. Data Selection
   Pada tahap ini, data relevan dipilih dari kumpulan data yang lebih besar untuk analisis atau pemodelan, dengan tujuan memastikan bahwa data terpilih sesuai dengan tujuan analisis atau permasalahan yang ingin diselesaikan.
b. Pembersihan Data (Data Cleaning)
   Mengatasi masalah seperti nilai yang hilang (missing values), data duplikat, dan outlier. Nilai rata-rata, median, atau nilai lain yang sesuai digunakan untuk mengisi data yang hilang.
c. Mapping
   Tahapan mapping mengubah data kategorikal menjadi data numerik, seperti menggunakan One-Hot Encoding atau Label Encoding. Pada data penelitian ini fitur Gender, Breastfeeding dan Stunting bertipe teks, maka perlu di ubah menjadi numerik. Misal, gender yang semula 'laki-laki' dan 'perempuan' diubah menjadi nilai numerik (0 dan 1), sedangkan breastfeeding dan stunting bertipe teks yaitu 'yes', 'no' maka diubah juga menjadi 0 dan 1.
d. Class Balancing
   Pada tahap ini dilakukan penyetaraan kelas untuk mengatasi distribusi data yang tidak seimbang. Kondisi ini terjadi ketika satu kategori memiliki lebih banyak sampel dibandingkan kategori lainnya, model cenderung lebih akurat dalam memprediksi kelas mayoritas sementara kelas minoritas sering terabaikan. Untuk mengatasi masalah tersebut dan memastikan model berkinerja baik pada kedua kelas, digunakan teknik oversampling dengan smote.
e. Standard Scaler
   Standard Scaler adalah teknik transformasi data yang digunakan untuk menstandarisasi fitur sehingga memiliki rata-rata sebesar nol dan penyimpangan standar sebesar satu.

2.4 Klasifikasi Decision Tree

Decision Tree membentuk model prediksi dalam bentuk struktur pohon yang terdiri dari simpul (node) dan cabang (branch). Pada setiap simpul, data dibagi berdasarkan suatu fitur atau atribut, sedangkan cabang menunjukkan hasil dari kondisi yang diterapkan pada fitur tersebut. Setiap simpul daun (leaf node) di pohon mewakili hasil akhir dari prediksi atau kategori yang dipilih. Langkah pertama yang perlu dilakukan pada Decision Tree adalah menghitung Entropy. Berikut persamaan dari Entropy :

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1537−1546
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6128

$$Entropy(S) = -\sum_{i=1}^{n} pi * \log_2 pi \quad (1)$$

Rumus persamaan (2) di mana S adalah himpunan data di simpul yang sedang dievaluasi. n adalah jumlah kelas dalam dataset. Dan pi adalah probabilitas kemunculan kelas ke-iii dalam dataset, dihitung sebagai rasio jumlah sampel dalam kelas tersebut terhadap total jumlah sampel di simpul. Kemudian setelah menghitung entropy, langkah selanjutnya menentukan Information Gain, yang mengukur seberapa besar pengurangan entropy ketika dataset dibagi berdasarkan suatu atribut. Persamaan Information Gain adalah sebagai berikut :

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

Pada persamaan (3) Entropy (S) merupakan entropy awal dari dataset S, yang mengukur ketidakpastian atau keacakan dalam himpunan data sebelum dibagi berdasarkan atribut A. Semakin tinggi nilai entropy, semakin tidak teratur data tersebut. Jika semua kasus dalam S termasuk dalam satu kelas, maka entropy adalah 0, menandakan bahwa data homogen. Untuk n adalah jumlah subset atau partisi yang terbentuk dari pembagian dataset S berdasarkan atribut A. Setiap subset Si mengandung data yang memiliki nilai tertentu dari atribut A. Kemudian |Si| merupakan jumlah kasus dalam subset Si, dan |S| sebagai jumlah total kasus dalam dataset S. Entropy (Si) digunakan untuk menghitung entropy pada subset Si, yang mengukur ketidakpastian dalam subset tersebut. Information Gain dihitung dengan mengurangkan entropy gabungan dari subset setelah pemisahan dari entropy awal. Rumus ini mengukur seberapa besar pengurangan ketidakpastian dalam dataset S ketika dibagi berdasarkan atribut A.

2.5 Klasifikasi Random Forest

Prinsip utama Random Forest adalah membuat sejumlah pohon keputusan dan mengombinasikan hasilnya guna meningkatkan ketepatan estimasi dan mengurangi risiko overfitting. Metode ini bekerja dengan membuat sejumlah pohon keputusan yang berbeda kemudian keputusan dari setiap pohon digabung untuk menentukan hasil final. Tahapan dari Random Forest yaitu mengambil beberapa subset acak dari dataset asli. Setiap subset dapat berisi pengambilan sampel yang berulang, sehingga satu data dapat muncul lebih dari sekali dalam subset. Untuk setiap subset data yang diambil, algoritma membangun pohon keputusan. Selama proses pembentukan pohon, ketika memilih fitur untuk membagi data di setiap simpul, algoritma memilih subset acak dari semua fitur yang tersedia. Pendekatan ini menciptakan variasi antar pohon dan mencegah pohon-pohon dari overfitting terhadap data yang sama. Terakhir yaitu penggabungan hasil prediksi, untuk tugas klasifikasi, Random Forest mengumpulkan prediksi dari semua pohon dan menentukan hasil akhir dengan menggunakan voting mayoritas. Kelas yang paling sering diprediksi oleh pohon-pohon keputusan akan menjadi hasil akhir.

2.6 Grid Search

Grid Search merupakan teknik yang digunakan untuk mengoptimalkan hyperparameter model dalam pembelajaran mesin. Keunggulan dari Grid Search adalah kesederhanaannya dan kemampuannya untuk menemukan kombinasi hyperparameter yang optimal secara menyeluruh. Pada penelitian ini Grid Search digunakan untuk optimasi algoritma Decision Tree dan Random Forest.

2.6.1. Optimasi Decision Tree

Pada penelitian ini, peningkatan akurasi model Decision Tree dilakukan menggunakan Grid Search guna menemukan kombinasi parameter terbaik. Grid Search berfungsi dengan menguji secara sistematis berbagai kemungkinan nilai parameter, sehingga dapat menghasilkan konfigurasi terbaik bagi algoritma tersebut. Dengan mengeksplorasi ruang parameter ini, Grid Search membantu dalam mengidentifikasi set parameter yang memberikan kinerja terbaik pada model Decision Tree, sehingga meningkatkan akurasi dan kemampuan generalisasi model terhadap data baru. Berikut Tabel 2 adalah parameter untuk Decision Tree.

Tabel 2. Parameter Decision Tree

Parameter	Value
criterion	['gini', 'entropy']
max_depth	[10, 20, 30]
min_samples_split	[2, 5, 10]
min_samples_leaf	[1, 2, 4]

2.6.2. Optimasi Random Forest

Metode Grid Search juga digunakan untuk mencari kombinasi parameter yang paling optimal untuk algoritma Random Forest guna meningkatkan akurasi model. Grid Search akan menguji secara sistematis berbagai kombinasi nilai parameter yang dibutuhkan untuk memecah node. Berikut Tabel 3 adalah parameter untuk Random Forest.

Tabel 3. Parameter Random Forest

Parameter	Value

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1537−1546
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6128

n_estimators	[100, 200, 300]
max_depth	[10, 20, 30]
min_samples_split	[2, 5, 10]
min_samples_leaf	[1, 2, 4]

2.7 Evaluasi Hasil

Evaluasi hasil dilakukan dengan memanfaatkan Confusion Matrix, yang berupa tabel guna mengevaluasi kinerja model melalui perbandingan antara hasil prediksi dan nilai asli. Confusion Matrix ini memberikan gambaran yang lebih rinci tentang bagaimana model memprediksi setiap kelas, sehingga membantu mengidentifikasi jenis kesalahan yang terjadi. Confusion matrix disajikan seperti pada Tabel 4 berikut.

Tabel 4. Confusion Matrix

Predicted	**Actual**	
	Positive	**Negative**
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

Dengan menggunakan informasi dari Tabel 4 di atas, dapat dimanfaatkan untuk perhitungan beragam metrik evaluasi kinerja model, yaitu:

a. $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$ (3)

Rumus ini mengukur akurasi atau tingkat ketepatan model secara keseluruhan, dengan membandingkan jumlah prediksi benar (baik positif maupun negatif) terhadap total keseluruhan prediksi.

b. $Precision = \frac{TP}{TP+FP}$ (4)

Precision mengukur ketepatan prediksi positif model, yaitu persentase prediksi positif yang sesuai dibandingkan dengan total prediksi positif yang dibuat. Precision yang tinggi menunjukkan sedikitnya kesalahan prediksi positif. Precision sangat penting dalam situasi di mana dampak dari kesalahan positif (false positive) cukup besar.

c. $Recall = \frac{TP}{TP+FN}$ (5)

Recall menunjukkan proporsi prediksi positif yang benar dari keseluruhan kasus positif yang sesungguhnya. Recall yang tinggi menandakan bahwa model dapat mengidentifikasi hampir seluruh contoh positif yang sebenarnya, yang sangat penting dalam situasi di mana kesalahan negatif (false negative) harus dihindari.

d. $F1-Score = \frac{2 * Precision * Recall}{Precision+Recall}$ (6)

F1-Score berguna untuk menyeimbangkan precision serta recall. Nilai F1 yang tinggi menunjukkan model memiliki precision dan recall yang baik. F1-Score sangat bermanfaat terutama ketika dataset tidak seimbang, karena memberikan evaluasi yang lebih adil terhadap kinerja model dengan mempertimbangkan kedua metrik tersebut (precision dan recall).

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Tahap awal penelitian ini adalah pengumpulan data relevan untuk klasifikasi stunting. Dataset yang digunakan terdiri dari 10.000 data, yang masing-masing memiliki 8 fitur utama yang berperan penting dalam analisis stunting. Fitur-fitur tersebut meliputi Gender, Age, Birth Weight, Birth Length, Body Weight, Body Length, Breastfeeding dan Stunting. Setiap fitur ini memiliki peran dalam membantu model klasifikasi memahami faktor-faktor yang memengaruhi stunting. Data pada fitur Stunting berfungsi sebagai label atau target yang menunjukkan apakah seorang anak mengalami stunting atau tidak. Adapun proses pengumpulan data ini sangat penting untuk memastikan kelengkapan dan keakuratan informasi yang akan digunakan dalam tahap preprocessing dan pengembangan model klasifikasi. Berikut Gambar 2 adalah isi dataset dalam penelitian ini.

	Gender	Age	Birth Weight	Birth Length	Body Weight	Body Length	Breastfeeding	Stunting
0	Male	17	3.0	49	10.0	72.2	No	No
1	Female	11	2.9	49	2.9	65.0	No	Yes
2	Male	16	2.9	49	8.5	72.2	No	Yes
3	Male	31	2.8	49	6.4	63.0	No	Yes
4	Male	15	3.1	49	10.5	49.0	No	Yes
...	...	...	...	...	...	...	...	...
9995	Male	15	3.0	49	9.0	63.0	No	Yes
9996	Female	12	2.8	48	7.7	63.0	No	No
9997	Male	16	2.8	49	7.7	49.0	No	No
9998	Male	14	2.8	49	10.0	69.0	No	Yes
9999	Female	10	3.0	49	7.7	80.0	No	Yes

10000 rows × 8 columns

Gambar 2. Informasi Dataset

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1537−1546
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6128

3.2 Preprocessing Data

Langkah berikutnya adalah melakukan preprocessing data, yaitu mempersiapkan dan membersihkan data agar dapat digunakan secara optimal dalam proses pemodelan. Proses preprocessing data terdapat beberapa langkah yaitu sebagai berikut :

3.2.1. Data Selection

Dalam tahap ini, fitur "breastfeeding" dihapus dari dataset sebagai bagian dari proses preprocessing data. Fitur yang digunakan berjumlah 8, namun setelah penghapusan fitur berubah menjadi 7 fitur. Penghapusan ini dilakukan karena fitur tersebut dianggap kurang relevan atau tidak memiliki kontribusi yang signifikan terhadap kinerja model. Langkah ini diambil untuk mengurangi kompleksitas data dan meningkatkan efisiensi serta akurasi model dalam memprediksi hasil stunting.

3.2.2. Pembersihan Data (Data Cleaning)

Tahapan data cleaning merupakan rangkaian langkah untuk membersihkan dan mempersiapkan data sebelum dilakukan analisis atau pemodelan. Data Cleaning dilakukan untuk mengatasi nilai yang hilang (missing values), menghapus data duplikat, serta penanganan outlier. Berikut Gambar 3 adalah hasil dari pembersihan data yang telah di lakukan.

```
Gender            0
Age               0
Birth Weight      0
Birth Length      0
Body Weight       0
Body Length       0
Stunting          0
dtype: int64
Jumlah data asli : 10000
Jumlah data yang duplikat : 2427
Jumlah data setelah menghapus duplikat: 7573
```

	Gender	Age	Birth Weight	Birth Length	Body Weight	Body Length	Stunting
0	Male	17	3.0	49	10.0	72.2	No
1	Female	11	2.9	49	2.9	65.0	Yes
2	Male	16	2.9	49	8.5	72.2	Yes
3	Male	31	2.8	49	6.4	63.0	Yes
4	Male	15	3.1	49	10.5	49.0	Yes
...	...	...	...	...	...	...	...
9992	Male	11	2.8	48	10.5	73.5	No
9994	Male	15	2.8	49	2.9	71.0	Yes
9996	Female	12	2.8	48	7.7	63.0	No
9997	Male	16	2.8	49	7.7	49.0	No
9999	Female	10	3.0	49	7.7	80.0	Yes

7573 rows × 7 columns

Gambar 3. Hasil Data Cleaning

Dari hasil data cleaning yang telah dilakukan seperti pada Gambar 3 di atas, dapat diperoleh informasi bahwa tidak ada nilai yang hilang dalam dataset. Kemudian terdapat data duplikat sejumlah 2.427 data dari data asli yang berjumlah 10.000. Setelah penghapusan data duplikat mendapatkan hasil sejumlah 7.573 data.

3.2.3. Mapping

Tahapan ini dilakukan transformasi nilai atribut berupa teks menjadi numerik yang dapat digunakan dalam pemodelan lebih lanjut. Beberapa atribut yang masih berisi nilai teks adalah atribut Gender dan Stunting, sebagaimana ditunjukkan pada Tabel 5 berikut.

Tabel 5. Data Sebelum Mapping

Gender	Age	Birth Weight	Birth Length	Body Weight	Body Length	Stunting
Male	14	2.8	48	2.9	49	No
Female	16	3	49	7	49	Yes

Berdasarkan data di Tabel 5, atribut Gender dan Stunting awalnya memiliki tipe data teks yang menggambarkan jenis kelamin dan kategori. Melalui proses mapping, nilai pada atribut Gender diubah dari ‘Male’ menjadi 1 dan ‘Female’ menjadi 0, sementara nilai atribut Stunting diubah dari ‘No’ menjadi 0 dan ‘Yes’ menjadi 1. Hasil transformasi ini disajikan dalam Tabel 6 berikut.

Tabel 6. Data Setelah Mapping

Gender	Age	Birth Weight	Birth Length	Body Weight	Body Length	Stunting
1	14	2.8	48	2.9	49	0
0	16	3	49	7	49	1

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1537−1546
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6128

3.2.4. Class Balancing

Tahap ini mencakup penyeimbangan kelas untuk mengatasi ketidakseimbangan dalam data, yang terjadi ketika jumlah data di satu kategori lebih banyak atau kurang dibandingkan dengan kategori lainnya. Metode oversampling diterapkan untuk mengatasi masalah ini, dengan tujuan untuk menyamakan jumlah data dengan cara menggandakan atau menambah data pada kategori minoritas sampai mendapatkan proporsi yang setara. Perbandingan hasil penyeimbangan kelas dapat dilihat dalam Tabel 7 berikut.

Tabel 7. Data Perbandingan Class Balancing

	Stunting	Tidak Stunting
Jumlah Data Sebelum Class Balancing	7.955	2.045
Jumlah Data Setelah Class Balancing	7.955	7.955

3.2.5. Standard Scaler

Pada proses Standard Scaler, dilakukan normalisasi data yang digunakan dalam preprocessing untuk memastikan bahwa fitur-fitur dalam dataset memiliki skala yang sama. Teknik ini mentransformasi data pada setiap fitur. Fitur dalam dataset sering kali memiliki rentang nilai yang berbeda. Tanpa normalisasi, fitur dengan skala yang lebih besar dapat mendominasi hasil model, yang mengakibatkan bias dan performa yang buruk. Berikut Tabel 8 adalah perbandingan data standard scaler.

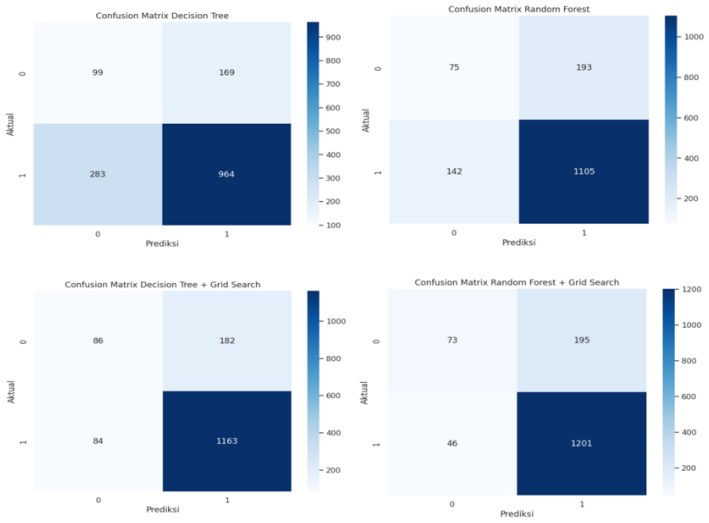
Tabel 8. Data Perbandingan Sebelum dan Setelah Standard Scaler

Atribut	Nilai Sebelum Standard Scaler	Nilai Setelah Standard Scaler
Gender	1	0.779355
Age	16	-0.046930
Birth Weight	2.9	0.151285
Birth Length	49	-0.269723
Body Weight	8.5	-0.359921
Body Length	72.2	-0.435891

Data dari Tabel 8 di atas menunjukkan beberapa atribut penting. Terdapat jarak yang signifikan antara nilai-nilai yang ada. Pada nilai Body Weight sebesar 8,5 jika dibandingkan dengan Body Length yang mencapai 72,2 dapat menjadi masalah dalam pemodelan, karena model akan menganggap nilai-nilai tersebut memiliki dampak yang berbeda, bahkan dapat mendominasi proses pelatihan. Oleh karena itu, langkah standar scaler menjadi penting untuk menyesuaikan skala dari nilai-nilai tersebut.

3.3 Evaluasi Hasil

Setelah dilakukan seluruh tahapan preprocessing data, kemudian dilakukan penghitungan evaluasi hasil untuk keempat model algoritma dengan confusion matrix. Keempat model tersebut yaitu tersiri dari Algoritma Decision Tree, Algoritma Random Forest, Algoritma Decision Tree yang dioptimasi dengan Grid Search dan Algoritma Random Forest yang dioptimasi dengan Grid Searh. Perhitungan evaluasi Confusion Matrix memberikan hasil seperti pada Gambar 4 di bawah.



Gambar 4. Hasil Confusion Matrix

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1537−1546
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6128

Berdasarkan data confusion matrix seperti pada Gambar 4 di atas, dapat dilakukan perhitungan untuk mengevaluasi kinerja model sebagaimana disajikan dalam Tabel 9 berikut.

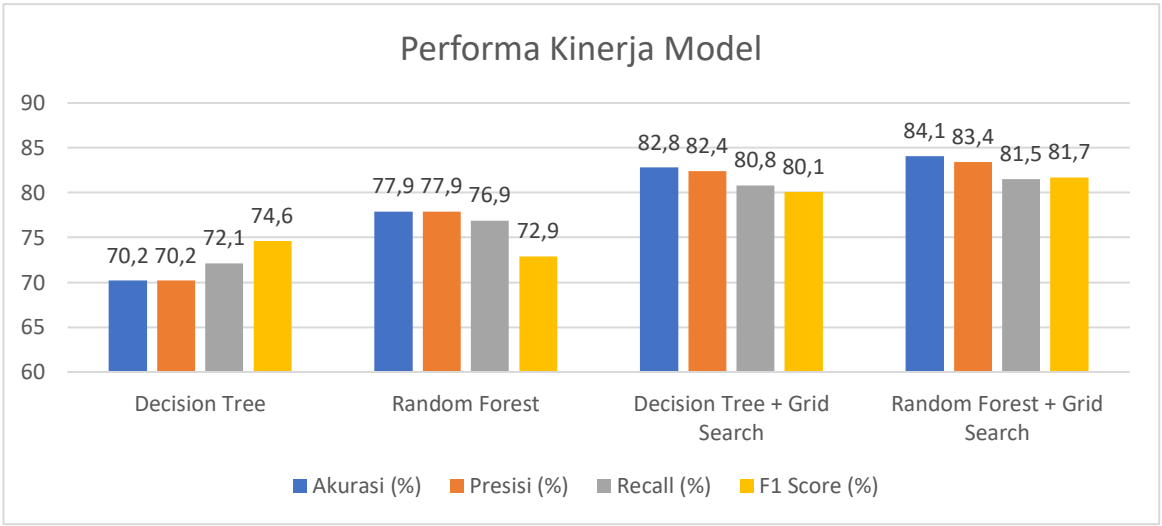
Tabel 9. Hasil evaluasi keseluruhan model

Metrik Evaluasi	Decision Tree	Random Forest	Decision Tree + Grid Search	Random Forest + Grid Search
Akurasi	70.2%	77.9%	82.8%	84.1%
Presisi	70.2%	77.9%	82.4%	83.4%
Recall	72.1%	76.9%	80.8%	81.5%
F1-score	74.6%	72.9%	80.1%	81.7%

Berdasarkan hasil evaluasi yang disajikan dalam Tabel 9 di atas, keempat model yang diuji menunjukkan kinerja yang baik dalam klasifikasi data. Model Decision Tree mencapai akurasi dan presisi sebesar 70,2%, dengan recall 72,1% serta F1-score 74,6%, menunjukkan kemampuan yang memadai dalam mengklasifikasikan data positif. Random Forest memberikan hasil yang lebih baik, dengan akurasi dan presisi 77,9% dan recall 76,9%, yang menunjukkan efektivitas dalam mengurangi kesalahan False Positive dan mendeteksi instance positif. Model Decision Tree yang dioptimasi dengan Grid Search menunjukkan akurasi tinggi sebesar 82,8%, presisi 82,4%, dan recall 80,8%, dengan F1-score mencapai 80,1%, menandakan presisi dan recall memiliki keseimbangan yang baik. Terakhir, model Random Forest yang juga dioptimalkan melalui Grid Search mencapai akurasi 84,1%, presisi 83,4%, dan recall 81,5%, dengan F1-score sebesar 81,7%, mencerminkan kemampuan yang sangat memuaskan dalam mengidentifikasi sampel positif dengan kesalahan yang minim. Secara keseluruhan, semua model menunjukkan peningkatan kinerja, terutama yang telah dioptimalkan dengan Grid Search, efektif dalam klasifikasi stunting.

3.4 Analisis Perbandingan Hasil Pengujian

Evaluasi kinerja model dilakukan melalui perbandingan beberapa algoritma klasifikasi, yaitu Decision Tree, Random Forest, serta versi dari kedua algoritma yang telah dioptimalkan menggunakan Grid Search. Metrik yang digunakan dalam evaluasi meliputi akurasi, presisi, recall, dan F1 Score, yang dinyatakan dalam persentase untuk memudahkan perbandingan performa antar model. Optimasi menggunakan Grid Search dilakukan untuk menemukan kombinasi hyperparameter terbaik sehingga model dapat mencapai hasil yang lebih maksimal. Untuk memvisualisasikan perbandingan kinerja setiap model, hasil evaluasi ditampilkan dalam diagram batang sebagaimana diperlihatkan pada Gambar 5. Diagram ini membantu dalam memahami seberapa efektif setiap algoritma dalam mengklasifikasikan data stunting, baik sebelum maupun setelah proses optimasi dilakukan.



Gambar 5. Diagram Performa Kinerja Model

Dari hasil pada Gambar 5, analisis performa keempat model untuk klasifikasi stunting dalam penelitian ini meliputi, Decision Tree, Random Forest, Decision Tree + Grid Search, dan Random Forest + Grid Search, menunjukkan hasil yang signifikan, terutama setelah proses optimasi. Model Decision Tree awal memiliki kinerja yang relatif baik dengan akurasi sebesar 70,2%, presisi 70,2%, recall 72,1%, dan F1-Score 74,6%. Meskipun cukup mampu mengklasifikasikan data stunting, terdapat ruang untuk peningkatan yang lebih besar. Sementara itu, model Random Forest awal menunjukkan hasil yang lebih baik, dengan akurasi 77,9%, presisi 77,9%, recall 76,9%, dan F1-Score 72,9%. Hal tersebut berarti bahwa Random Forest dengan kemampuannya lebih baik dalam membuat keputusan yang tepat jika dibandingkan dengan Decision Tree.

Setelah menerapkan optimasi melalui Grid Search, model Decision Tree menunjukkan peningkatan yang signifikan, dengan akurasi meningkat menjadi 82,8%, presisi 82,4%, recall 80,8%, dan F1-Score 80,1%. Peningkatan

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1537−1546
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6128

ini mencerminkan bahwa penyesuaian parameter model berhasil meningkatkan kinerja klasifikasi secara keseluruhan. Demikian pula, model Random Forest yang dioptimasi juga menunjukkan hasil yang mengesankan. Setelah optimasi, akurasi meningkat menjadi 84,1%, presisi 83,4%, recall 81,5%, dan F1-Score 81,7%. Peningkatan ini mengindikasikan bahwa penggunaan Grid Search telah efektif dalam meningkatkan kinerja model untuk mengklasifikasikan data dengan lebih baik, mengurangi kesalahan False Positive, dan meningkatkan deteksi kasus positif. Perbandingan keseluruhan menunjukkan bahwa optimasi menggunakan Grid Search pada kedua algoritma memberikan peningkatan kinerja yang signifikan. Kedua model tersebut terbukti dengan akurasi yang lebih tinggi dalam pengklasifikasian. Hasil ini menunjukkan bahwa optimasi parameter melalui Grid Search berdampak positif dalam meningkatkan kinerja model, dengan Random Forest yang dioptimasi menjadi model paling unggul dalam klasifikasi stunting. Peningkatan ini dapat berkontribusi pada upaya yang lebih efektif dalam mengidentifikasi dan menangani kasus stunting pada balita.

4. KESIMPULAN

Dalam penelitian ini, berbagai algoritma machine learning digunakan untuk menganalisis dan menentukan status stunting pada balita, dengan hasil yang memuaskan. Proses optimasi hyperparameter yang diterapkan pada model Decision Tree dan Random Forest berhasil meningkatkan kinerja keduanya. Model Decision Tree setelah optimasi dengan Grid Search menunjukkan peningkatan dengan akurasi sebesar 82,8%. Sedangkan untuk Random Forest setelah optimasi dengan Grid Search menunjukkan hasil terbaik, dengan akurasi mencapai 84,1%. Hasil menunjukkan bahwa optimasi dengan Grid Search memberikan peningkatan kinerja yang signifikan, terutama pada model Random Forest yang mencapai akurasi tertinggi di antara model lain. Perbandingan antara kedua model menunjukkan bahwa Random Forest yang dioptimalkan melalui Grid Search menjadi pilihan terbaik dalam mengidentifikasi kasus stunting, dengan akurasi lebih tinggi dan kemampuan untuk menghindari kesalahan klasifikasi yang signifikan. Temuan ini diharapkan dapat mendukung upaya pencegahan stunting pada balita melalui penyediaan metode klasifikasi yang lebih handal. Dengan pendekatan ini, diharapkan dapat berkontribusi dalam menyusun strategi intervensi yang lebih efektif, bertujuan menurunkan prevalensi stunting di negara berkembang dan meningkatkan kualitas hidup anak-anak. Selain itu, model ini mampu menyediakan solusi yang efektif untuk mengidentifikasi masalah stunting dengan tingkat ketepatan yang tinggi, sehingga mendukung proses pengambilan keputusan yang lebih akurat.

REFERENCES

[1] A. Arwansyah, A. F. Lewa, M. Muliani, S. Warnasih, A. Z. Mustopa, and A. R. Arif, “Molecular Recognition of Moringa oleifera Active Compounds for Stunted Growth Prevention Using Network Pharmacology and Molecular Modeling Approach,” *ACS Omega*, vol. 8, no. 46, pp. 44121–44138, Nov. 2023, doi: 10.1021/acsomega.3c06379.

[2] T. Mulyaningsih, I. Mohanty, V. Widyaningsih, T. A. Gebremedhin, R. Miranti, and V. H. Wiyono, “Beyond personal factors: Multilevel determinants of childhood stunting in Indonesia,” *PLOS ONE*, vol. 16, no. 11, p. e0260265, Nov. 2021, doi: 10.1371/journal.pone.0260265.

[3] “PAUDPEDIA - Prevalensi Stunting Tahun 2022 di Angka 21,6%, Protein Hewani Terbukti Cegah Stunting.” Accessed: Oct. 06, 2024. [Online]. Available: https://paudpedia.kemdikbud.go.id/kabar-paud/berita/prevalensi-stunting-tahun-2022-di-angka-216-protein-hewani-terbukti-cegah-stunting?do=MTQyMy1iNmNmMmYzZA==&ix=MTEtYmJkNjQ3YzA=

[4] M. Y. Titimeidara and W. Hadikurniawati, “IMPLEMENTASI METODE NAÏVE BAYES CLASSIFIER UNTUK KLASIFIKASI STATUS GIZI STUNTING PADA BALITA,” *J. Ilm. Inform.*, vol. 9, no. 01, Art. no. 01, Jun. 2021, doi: 10.33884/jif.v9i01.3741.

[5] F. Sulistyawati and N. P. Widarini, “Kejadian stunting masa pandemi covid-19,” *Med Respati J Ilm Kesehat*, vol. 17, no. 1, p. 37, 2022.

[6] B. Buenita, P. B. Chandra, and L. K. K. Z. Zendrato, “FAKTOR DETERMINAN KEJADIAN STUNTING PADA BALITA DI UPTD PUSKESMAS KECAMATAN GUNUNGSITOLI ALO’OA,” *J. Kesehat. Tambusai*, vol. 4, no. 3, pp. 3806–3818, Sep. 2023, doi: 10.31004/jkt.v4i3.18500.

[7] “World Health Statistics.” Accessed: Oct. 06, 2024. [Online]. Available: https://www.who.int/data/gho/publications/world-health-statistics

[8] D. S. P. M.Kes S. K. M., *Strategi Pencegahan Stunting Pada Usia Baduta (Bawah Dua Tahun)*. Deepublish, 2023.

[9] I. P. Putri, T. Terttiaavini, and N. Arminarahmah, “Analisis Perbandingan Algoritma Machine Learning untuk Prediksi Stunting pada Anak: Comparative Analysis of Machine Learning Algorithms for Predicting Child Stunting,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, Art. no. 1, Jan. 2024, doi: 10.57152/malcom.v4i1.1078.

[10] C. Serón-Arbeloa *et al.*, “Malnutrition Screening and Assessment,” *Nutrients*, vol. 14, no. 12, p. 2392, Jun. 2022, doi: 10.3390/nu14122392.

[11] I. W. M.Kom S. Si, B. N. M.Kom S. Si, and M. M.Si S. Si, *Data Mining Menggunakan Android, Weka, dan SPSS*. Airlangga University Press, 2020.

[12] M. A. Hossain, R. Ferdousi, S. A. Hossain, M. F. Alhamid, and A. E. Saddik, “A Novel Framework for Recommending Data Mining Algorithm in Dynamic IoT Environment,” *IEEE Access*, vol. 8, pp. 157333–157345, 2020, doi: 10.1109/ACCESS.2020.3019480.

[13] R. F. Putra *et al.*, *DATA MINING : Algoritma dan Penerapannya*. PT. Sonpedia Publishing Indonesia, 2023.

[14] L. Qadrini, H. Hikmah, and M. Megasari, “Oversampling, Undersampling, Smote SVM dan Random Forest pada Klasifikasi Penerima Bidikmisi Sejawa Timur Tahun 2017,” *J. Comput. Syst. Inform. JoSYC*, vol. 3, no. 4, Art. no. 4, Sep. 2022, doi: 10.47065/josyc.v3i4.2154.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1537−1546
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6128

[15] J. Rusman, B. Z. Haryati, and A. Michael, "Optimisasi Hiperparameter Tuning pada Metode Support Vector Machine untuk Klasifikasi Tingkat Kematangan Buah Kopi," *J-Icon J. Komput. Dan Inform.*, vol. 11, no. 2, Art. no. 2, Oct. 2023, doi: 10.35508/jicon.v11i2.12571.
[16] T. Hidayat *et al.*, "Performance Prediction Using Cross Validation (GridSearchCV) for Stunting Prevalence," in *2024 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)*, Feb. 2024, pp. 1–6. doi: 10.1109/AIMS61812.2024.10512657.
[17] P. Handayani, A. C. Fauzan, and H. Harliana, "Machine Learning Klasifikasi Status Gizi Balita Menggunakan Algoritma Random Forest," *KLIK Kaji. Ilm. Inform. Dan Komput.*, vol. 4, no. 6, Art. no. 6, Jun. 2024, doi: 10.30865/klik.v4i6.1909.
[18] Y. Wiratama and R. A. Aziz, "Perbandingan Prediksi Penyakit Stunting Balita Menggunakan Algoritma Support Vektor Machine dan Random Forest," *Build. Inform. Technol. Sci. BITS*, vol. 6, no. 2, Art. no. 2, Sep. 2024, doi: 10.47065/bits.v6i2.5543.
[19] M. Utomo and R. Prathivi, "Perbandingan Algoritma Support Vector Machine dan Decision Tree untuk Klasifikasi Performa Perusahaan," *Build. Inform. Technol. Sci. BITS*, vol. 6, no. 1, Art. no. 1, Jun. 2024, doi: 10.47065/bits.v6i1.5278.
[20] H. Kurniawan and A. Rahim, "Implementasi Algoritma Gaussian Naïve Bayes Dalam Klasifikasi Status Gizi Pada Balita," *Build. Inform. Technol. Sci. BITS*, vol. 6, no. 2, pp. 627–634, 2024.
[21] S. Lonang and D. Normawati, "Klasifikasi Status Stunting Pada Balita Menggunakan K-Nearest Neighbor Dengan Feature Selection Backward Elimination," *J. MEDIA Inform. BUDIDARMA*, vol. 6, no. 1, Art. no. 1, Jan. 2022, doi: 10.30865/mib.v6i1.3312.