

Application Of The Random Forest Algorithm For Classification Of Stunting Factors In Toddlers

Penerapan Algoritma Random Forest Untuk Klasifikasi Faktor Stunting Pada Balita

Devia Zulvani¹, Fitri Ayuning Tyas², Ahmad Faizin³

^{1, 2, 3} Sistem Informasi, Fakultas Sains, Teknologi dan Kesehatan Universitas Muhammadiyah Brebes, Jl. Pangeran Diponegoro No. 184, Grengseng, Taraban, Kec. Paguyangan, Kabupaten Brebes, Jawa Tengah, 52276
Email: deviadevani02@mail.com, tyas_fa@stmikmpb.ac.id, faiz_ahmad@stmikmpb.ac.id

Abstract

Stunting is a type of growth failure caused by chronic malnutrition, a condition that causes children to experience stunted growth at a certain age. The high rate of stunting highlights the need for intensive and innovative efforts to address it, including through data-driven and technology-based approaches. This study aims to determine the accuracy of the Random Forest algorithm to be implemented for stunting risk classification in infants, using SMOTE for data balancing. Evaluation was conducted using accuracy, precision, recall, and F1-Score metrics. The results show that the Random Forest algorithm provides fairly high performance with an accuracy of 79.17%, precision of 90.54%, recall of 82.43%, and an F1-Score of 86.29%. Based on these results, it can be seen that proper preprocessing plays a significant role in improving classification performance. Based on this, it can be concluded that the Random Forest model is capable of building a classification model that is quite accurate in identifying stunting status in toddlers. With a total of 10,000 entries, this model successfully recognized patterns in the data with fairly good performance results.

Keywords: classification, stunting, data mining, random forest

Abstrak

Stunting adalah jenis kegagalan pertumbuhan yang disebabkan oleh kekurangan nutrisi yang berkelanjutan, kondisi ini menyebabkan anak-anak mengalami keterbelakangan pertumbuhan tinggi badan pada usia tertentu. Tingginya angka stunting ini menyoroti perlunya upaya intensif dan inovatif untuk menanganinya, termasuk melalui pendekatan berbasis data dan teknologi. Penelitian ini bertujuan untuk mengetahui keunggulan nilai akurasi dari algoritma Random Forest yang akan diimplementasikan untuk klasifikasi risiko stunting pada balita dengan SMOTE untuk menyeimbangkan data. Evaluasi dilakukan dengan metrik akurasi, precision, recall, dan F1-Score. Hasil penelitian menunjukkan bahwa algoritma Random Forest memberikan performa yang cukup tinggi yaitu dengan akurasi 79.17%, precision 90.54%, recall 82.43%, dan F1-Score sebesar 86.29%. Berdasarkan hasil tersebut, dapat diketahui bahwa preprocessing yang tepat berperan cukup besar dalam meningkatkan performa klasifikasi. Berdasarkan hal ini, dapat disimpulkan bahwa model Random Forest mampu membangun model klasifikasi yang cukup akurat dalam mengidentifikasi status stunting pada balita. Dengan jumlah data sebanyak 10.000 entri, model ini berhasil mengenali pola-pola yang terdapat pada data dengan hasil kinerja yang cukup baik.

Kata kunci: klasifikasi, stunting, data mining, random forest

1. PENDAHULUAN

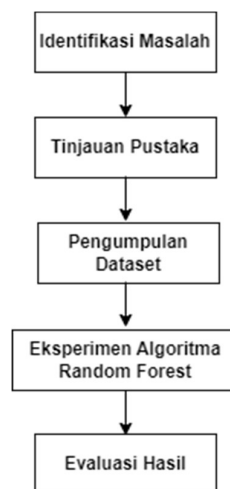
Stunting merupakan situasi di mana anak tidak berkembang sesuai dengan umur mereka akibat kekurangan nutrisi dalam waktu waktu yang lama. Akibatnya, tinggi badan anak menjadi lebih rendah daripada standar yang diharapkan [1]. Diperkirakan ada sekitar 149 juta anak balita di seluruh dunia yang mengalami *stunting* berdasarkan laporan UNICEF TAHUN 2023 [2], dan meskipun prevalensi di

Indonesia menurun, angkanya masih di atas target WHO [3]. Tingginya angka prevalensi ini menunjukkan bahwa diperlukan strategi yang lebih inovatif dan didukung oleh data, sementara identifikasi *stunting* saat ini biasanya dilakukan dengan mengukur antropometri seperti tinggi dan berat badan, meskipun metode ini efektif, metode ini memerlukan tenaga ahli, memakan waktu, dan beresiko mengalami kesalahan, terutama di wilayah yang memiliki sumber daya terbatas [4], maka diperlukan pendekatan yang lebih menyeluruh dengan data yang lebih lengkap [5], salah satunya dengan menerapkan *data mining* dan *machine learning* untuk mengidentifikasi dan memprediksi risiko *stunting*.

Penelitian ini mengusulkan penggunaan *Random Forest* untuk mengklasifikasikan faktor-faktor risiko *stunting* pada balita. *Random Forest* merupakan metode *ensemble learning* yang menggabungkan banyak pohon keputusan untuk menghasilkan prediksi yang lebih akurat dan stabil [6]. Keunggulan algoritma ini adalah kemampuannya mengatasi hubungan antar variabel yang kompleks dan mencegah *overfitting* [7]. Salah satu cara untuk mengatasi masalah ketidakseimbangan data—di mana jumlah kasus *stunting* lebih sedikit—penelitian ini akan menerapkan teknik *SMOTE* (*Synthetic Minority Oversampling Technique*) untuk menyeimbangkan dataset [8], sehingga meningkatkan akurasi model dalam mengklasifikasikan kasus *stunting*. Dengan menerapkan pendekatan ini, diharapkan dapat dihasilkan model klasifikasi yang lebih efektif dan akurat untuk mendukung upaya pencegahan *stunting*.

2. METODE PENELITIAN

Metode yang diterapkan pada penelitian ini ialah eksperimen, yakni suatu pendekatan yang dimaksudkan untuk memahami dampak suatu perlakuan pada variabel lain dalam situasi yang terkontrol. Metode eksperimen dilakukan dengan tahapan uji coba yang dirancang secara rapi agar bisa mendapatkan jawaban yang tepat dan tidak memihak [9]. Dalam penelitian ini, metode eksperimen digunakan untuk mengevaluasi sejauh mana algoritma *Random forest* mampu menunjukkan performa terbaik dalam mengklasifikasikan status *stunting* pada balita. Tahapan eksperimen mencakup proses memasukkan dataset, melakukan pra-pengolahan data (*data preprocessing*), menerapkan algoritma *Random forest*, serta mengevaluasi hasil model dengan memanfaatkan metrik *accuracy* dan *F1-Score* yang diukur melalui *confusion matrix*.



Gambar 1 Tahapan Penelitian

2.1 Identifikasi Masalah

Proses pengamatan terhadap permasalahan yang terjadi dalam masyarakat tentang tingginya angka *stunting* pada balita yang berfokus pada solusi yang relevan dan sesuai kebutuhan.

2.2 Tinjauan Pustaka

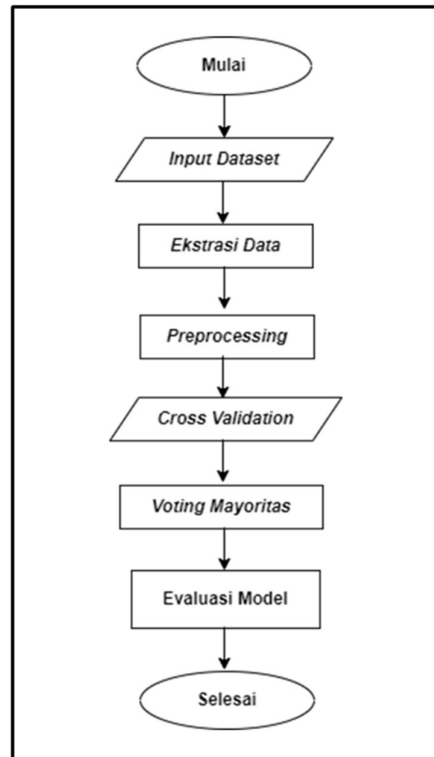
Mengumpulkan dan mempelajari penelitian sebelumnya serta referensi yang mendukung penelitian terdahulu.

2.3 Pengumpulan dan *Preprocessing* Data

Dataset yang dikumpulkan bersumber dari *web Kaggle* yang berjudul “faktor *stunting*” dan akan melalui tahap *preprocessing* terlebih dahulu untuk memastikan kualitas data.

2.4 Eksperiemn Algoritma *Random Forest*

Menerapkan algoritma klasifikasi *Random Forest* pada dataset yang sudah disiapkan dan memulai eksperimen.



Gambar 2 *Flowchart Random Forest*

2.4.1 Input Dataset

Memasukkan data yang ada setelah terkumpul. Dataset yang digunakan mencakup informasi antropometri anak berusia dibawah lima tahun dengan total sebanyak 10.000 data dari hasil penelitian Harnelia tahun 2024.

2.4.2 Ekstraksi Data

Bertujuan untuk mengambil data mentah dari sumber, dalam penelitian ini yaitu dataset berjudul faktor *stunting* sebelum dibersihkan untuk proses selanjutnya.

2.4.3 *Preprocessing*

Bertujuan untuk membersihkan serta mempersiapkan data sebelum diolah agar sesuai kebutuhan, juga dilakukan penyeimbangan data dengan *SMOTE*.

2.4.4 *Cross Validation*

Data yang sudah diproses kemudian dipecah menjadi data *k fold* untuk *training* dan 1 *fold* untuk *testing*. Proses ini diulang sampai semua *fold* pernah jadi *testing* dan menghasilkan perhitungan perfoma.

2.4.5 Voting Mayoritas

Setiap pohon keputusan memberikan ramalan dan hasil akhir ditentukan berdasarkan suara (klasifikasi) mayoritas dari seluruh pohon yang dibangun.

2.4.6 Evaluasi Model

Penilaian terhadap model dilaksanakan pada hasil akhir dengan memanfaatkan metrik seperti akurasi, *precision*, *recall*, *F1-Score*, *Confusion Matrix* dan *ROC-AUC* guna menilai kemampuan model dan pengklasifikasian.

3. HASIL DAN PEMBAHASAN

Hasil penelitian yang dilakukan dalam penerapan algoritma *Random Forest* pada klasifikasi stunting pada balita dengan metode eksperimen memiliki tahapan sebagai berikut:

3.1 Pengumpulan Dataset

Eksperimen penelitian ini menggunakan dataset publik berjudul “faktor *stunting*” yang merupakan hasil penelitian Harnelia tahun 2024 dan bersumber dari Kaggle.

3.2 Preprocessing Data

Preprocessing data dilakukan untuk menghilangkan data yang tidak penting yang bisa memengaruhi kualitas hasil model.

3.3 Eksperimen

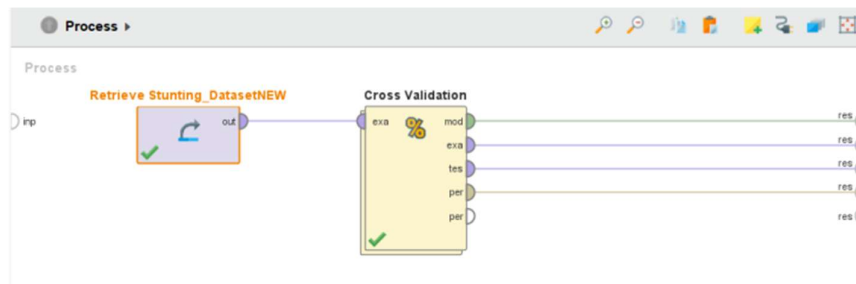
Proses tahapan eksperimen yang dilakukan:

3.3.1 Input Data

Memasukkan dataset yang akan digunakan ke *RapidMiner* dengan format csv, kemudian memastikan data sudah sesuai dengan kebutuhan untuk selanjutnya di proses ke tahap selanjutnya.

3.3.2 Eksperimen

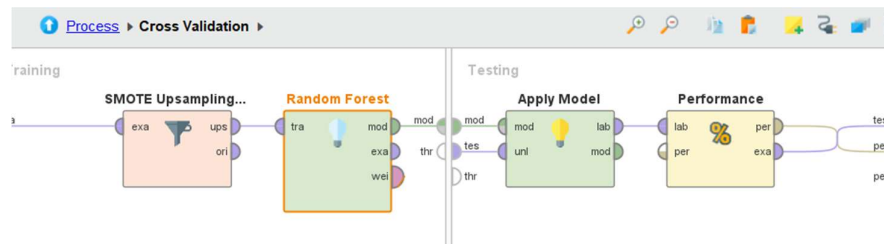
Menggunakan 2 operator utama yaitu *retrieve* untuk memuat dataset dan operator *cross-validation* untuk evaluasi kinerja model dengan membagi dataset menjadi *subset* tertentu.



Gambar 3 Proses Eksperimen

3.3.2.1 Cross-validation

Bertujuan untuk me,atih data pada *subset training* yang berisi operator SMOTE untuk menyeimbangkan data, dan operator *Random Forest* yang kemudian disambungkan ke operator Apply Model dan Perrformance (*Binominal Classification*) untuk mengukur kinerja model.



Gambar 4 Cross-Validation

3.3.3 Result Eksperimen

Result yang dihasilkan dari tahapan sebelumnya memaparkan performa akurasi *Random Forest* dalam klasifikasi *stunting* pada balita sebesar 79,17%, *precision* kelas

Yes 90.54%, *Precision* kelas No 549.31%, dan *Recall* kelas Yes 82.43%, serta *Recall* kelas No 66.50%

☒ Table View
 ☐ Plot View

accuracy: 79.17% +/- 0.90% (micro average: 79.17%)

	true No	true Yes	class precision
pred. No	1360	1398	49.31%
pred. Yes	685	6557	90.54%
class recall	66.50%	82.43%	

Gambar 5 Result Random Forest

3.3.4 Hasil dan Evaluasi

Eksperimen dilakukan dengan aplikasi *RapidMiner* yang kemudian pada prosesnya dilakukan uji coba akurasi dengan *10-Fold Cross Validation* dan *Confusion Matrix* untuk melihat performanya . Hasil dari validasi yang sudah dilaksanakan melalui analisis hasil *confussion matrix* adalah untuk memperoleh nilai *True Positive (TP)*, *False Positive (FP)*, *False Negative (FN)*, dan *True Negative*.

```

PerformanceVector:
accuracy: 79.17% +/- 0.90% (micro average: 79.17%)
ConfusionMatrix:
True:   No    Yes
No:    1360   1398
Yes:    685   6557
AUC (optimistic): 0.830 +/- 0.015 (micro average: 0.830) (positive class: Yes)
AUC: 0.823 +/- 0.017 (micro average: 0.823) (positive class: Yes)
AUC (pessimistic): 0.817 +/- 0.018 (micro average: 0.817) (positive class: Yes)
precision: 90.55% +/- 0.80% (micro average: 90.54%) (positive class: Yes)
ConfusionMatrix:
True:   No    Yes
No:    1360   1398
Yes:    685   6557
recall: 82.43% +/- 1.28% (micro average: 82.43%) (positive class: Yes)
ConfusionMatrix:
True:   No    Yes
No:    1360   1398
Yes:    685   6557
f_measure: 86.29% +/- 0.65% (micro average: 86.29%) (positive class: Yes)
ConfusionMatrix:
True:   No    Yes
No:    1360   1398
Yes:    685   6557
    
```

Gambar 6 Hasil Akurasi Penelitian

Hasil matrix evaluasi menunjukkan *accuracy* sebesar 79.17%, nilai *precision* sebesar 90.54% , nilai *recall* sebesar 82.43%, dan *f_measure/ FI-Score* sebesar 86.29 %. Hasil ini memaparkan bahwa model memiliki kinerja yang cukup baik, dan evaluasi membantu memahami kinerja model lebih dalam untuk menunjukkan area perbaikan lebih lanjut kedepannya.

Berikut adalah salah satu tampilan satu model pohon keputusan:

Tree

```

Birth Length > 49.500
|   Body Length = 49
|   |   Body Weight = 10: Yes {No=0, Yes=5}
|   |   Body Weight = 10.5: No {No=1, Yes=0}
|   |   Body Weight = 2.9
|   |   |   Birth Weight = 2.3: No {No=4, Yes=0}
|   |   |   Birth Weight = 2.8: No {No=2, Yes=1}
|   |   |   Birth Weight = 2.9: Yes {No=0, Yes=2}
|   |   Body Weight = 6.2: Yes {No=0, Yes=3}
|   |   Body Weight = 6.4
|   |   |   Age > 16.500: Yes {No=0, Yes=10}
|   |   |   Age ≤ 16.500: No {No=32, Yes=7}
|   |   Body Weight = 7
|   |   |   Birth Weight = 2.3: Yes {No=0, Yes=1}
|   |   |   Birth Weight = 2.8: No {No=3, Yes=0}
|   |   |   Birth Weight = 3: Yes {No=0, Yes=2}
|   |   Body Weight = 7.1
|   |   |   Gender = Female: No {No=1, Yes=0}
|   |   |   Gender = Male: Yes {No=0, Yes=3}
|   |   Body Weight = 7.2: Yes {No=0, Yes=3}
|   |   Body Weight = 7.7: No {No=13, Yes=0}
|   |   Body Weight = 8: Yes {No=0, Yes=1}
|   |   Body Weight = 8.4: Yes {No=0, Yes=1}
|   |   Body Weight = 8.5
|   |   |   Age > 12: No {No=5, Yes=1}
|   |   |   Age ≤ 12: Yes {No=0, Yes=7}
|   |   Body Weight = 9
|   |   |   Age > 14.500: Yes {No=0, Yes=6}
|   |   |   Age ≤ 14.500: No {No=20, Yes=12}
|   Body Length = 63
|   |   Body Weight = 10
|   |   |   Gender = Female: No {No=2, Yes=2}
|   |   |   Gender = Male: Yes {No=0, Yes=3}

```

Gambar 7 Model Pohon Keputusan

4. KESIMPULAN

Penelitian ini menguji performa algoritma *Random Forest* untuk mendeteksi *stunting* pada balita menggunakan data sebanyak 10.000 entri, didukung oleh teknik *SMOTE* untuk mengatasi ketidakseimbangan kelas dan *Cross-Validation* untuk validasi model. Hasilnya menunjukkan model memiliki kinerja yang kuat dengan akurasi 79,17%, presisi 90,54%, dan *recall* 82,43%, menghasilkan *F1-Score* 86,28%. Meskipun model berhasil mendeteksi sebagian besar kasus, masih ada 1.938 kasus *stunting* yang tidak terdeteksi (*false negative*), menunjukkan perlunya peningkatan sensitivitas model di masa depan.

DAFTAR PUSTAKA

- [1] Maula Hidayat, F., Kusrini, & Ainul Yaqin. (2024). Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Status Gizi Stunting Pada Balita. *Dielektrika*, 11(2), 107–118. <https://doi.org/10.29303/dielektrika.v11i2.384>

- [2] Ray, S., Das, J., Pande, R., & Nithya, A. (2024). *Swati Ray 1 , Joyati Das 2* , Ranjana Pande 3 , and A. Nithya 2. 12(2)*, 195–222. <https://doi.org/10.1201/9781032622408-13>
- [3] Maula Hidayat, F., Kusriani, & Ainul Yaqin. (2024). Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Status Gizi Stunting Pada Balita. *Dielektrika*, 11(2), 107–118. <https://doi.org/10.29303/dielektrika.v11i2.384>
- [4] Ray, S., Das, J., Pande, R., & Nithya, A. (2024). *Swati Ray 1 , Joyati Das 2* , Ranjana Pande 3 , and A. Nithya 2. 12(2)*, 195–222. <https://doi.org/10.1201/9781032622408-13>
- [5] Putri, I. P., Terttiaavini, T., & Arminarahmah, N. (2024). Analisis Perbandingan Algoritma Machine Learning untuk Prediksi Stunting pada Anak. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(1), 257–265. <https://doi.org/10.57152/malcom.v4i1.1078>
- [6] Rayadin, M. A., Musaruddin, M., Saputra, R. A., & Isnawaty, I. (2024). Implementasi Ensemble Learning Metode XGBoost dan Random Forest untuk Prediksi Waktu Penggantian Baterai Aki. *BIOS: Jurnal Teknologi Informasi Dan Rekayasa Komputer*, 5(2), 111–119.
- [7] Setyawan, N. H., & Wakhidah, N. (2025). *Analisis perbandingan metode logistic regression, random forest, gradient boosting untuk prediksi diabetes. 10(1)*, 150–162.
- [8] Erlin, E., Desnelita, Y., Nasution, N., Suryati, L., & Zoromi, F. (2022). Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 21(3), 677–690. <https://doi.org/10.30812/matrik.v21i3.1726>
- [9] Fauzan, R., Vitianingsih, A. V., & Cahyono, D. (2025). *Application of Classification Algorithms in Machine Learning for Phishing Detection Penerapan Algoritma Klasifikasi pada Machine Learning untuk Deteksi Phishing. 5(April)*, 531–540.