

KINERJA RANDOM FOREST DAN XGBOOST DALAM KLASIFIKASI STUNTING PADA BALITA DI PULAU TALAUD

Ninka Natalia Br Singarimbun - 1301210471

S1 Informatika

2025



DAFTAR ISI

1. Pendahuluan

2. Metodologi

3. Hasil dan Pembahasan

4. Kesimpulan

PENDAHULUAN

LATAR BELAKANG



- Stunting adalah masalah pada pertumbuhan balita yang ditandai dengan tinggi badan balita yang terlalu pendek dibandingkan dengan balita seusianya [1].
- Di Indonesia, sekitar 6,3 juta anak balita mengalami stunting dari 149 juta anak di dunia yang terkena stunting (2022).
- Stunting meningkatkan risiko penyakit dan kematian, menunda kemampuan motorik, dan mengganggu perkembangan otak anak [3].
- Salah satu solusi yang diusulkan adalah pemanfaatan sistem pembelajaran mesin untuk memprediksi dan mencegah stunting pada anak, karena pembelajaran mesin telah berhasil diterapkan di bidang medis [4][5][6][7]

STUDI TERKAIT



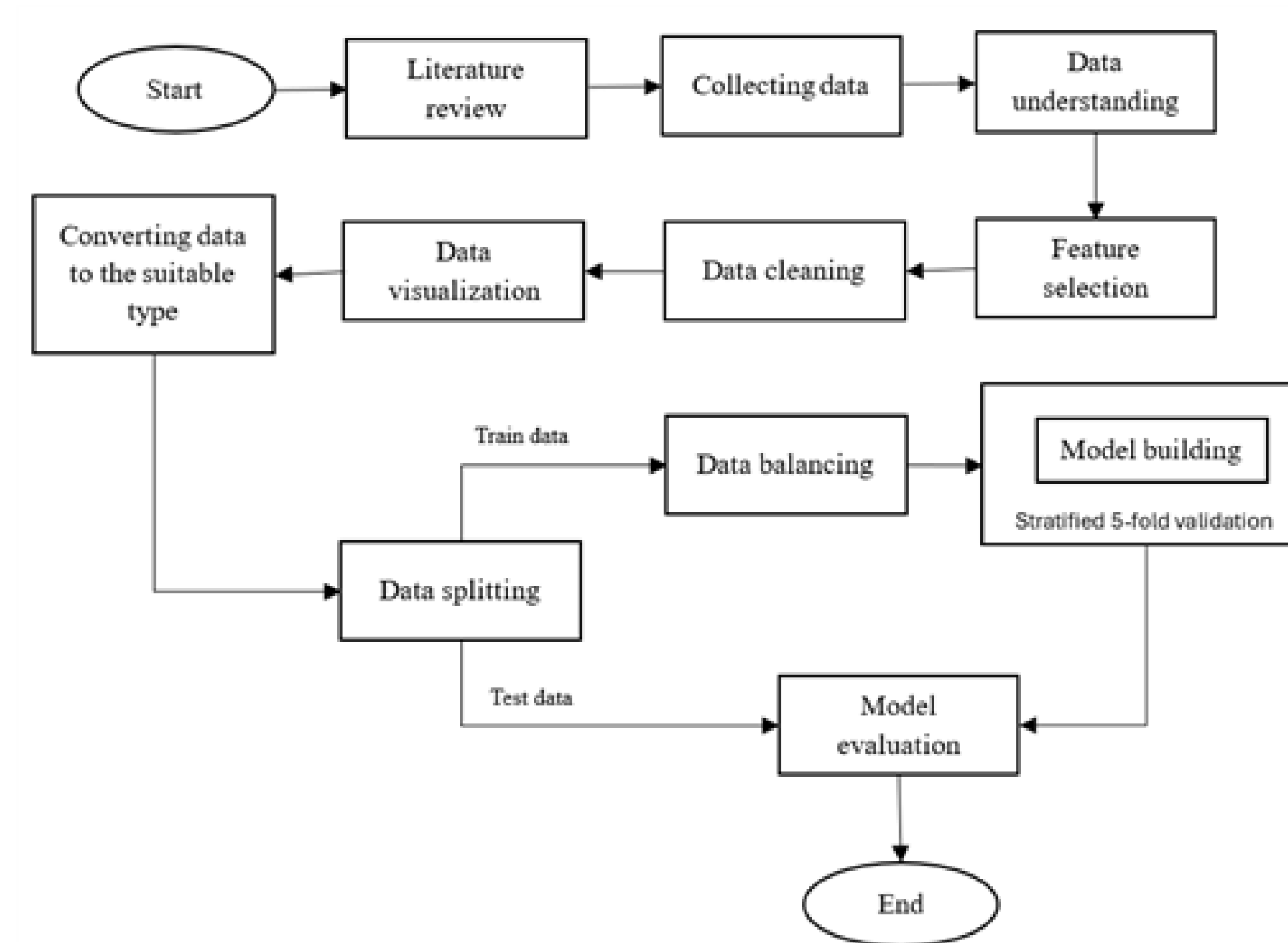
- Penelitian memprediksi stunting di Zambia dengan menggunakan metode Regresi Logistik, Random Forest, Naïve Bayes, Support Vector Machine (SVM), dan XGBoost. Hasilnya, model Random Forest mencapai akurasi data pelatihan dan pengujian yang tertinggi, dengan nilai 79,2% dan 61,6% [8]
- Studi pada tahun 2023 menganalisis penerapan metode machine learning untuk mengklasifikasikan stunting di Rwanda dengan menggunakan Gradient Boosting, Random Forest, SVM, XGBoost, dan Regresi Logistik. Penelitian tersebut menemukan bahwa XGBoost memiliki kinerja terbaik, dengan akurasi 79,13% [9].

TUJUAN



- Membandingkan kinerja algoritma Random Forest dan XGBoost, yang telah dikenal sebagai model terbaik di masing-masing pendekatan, pada data balita di Kepulauan Talaud yang memiliki karakteristik demografi dan sosio-ekonomi yang berbeda dengan daerah lain.
- Menggunakan metode penyeimbangan kelas berbasis rentang kuartil (Q1-Q3) untuk menciptakan data sintetis yang valid secara medis, memastikan relevansi lokal dan efektivitas model dalam mendeteksi stunting.

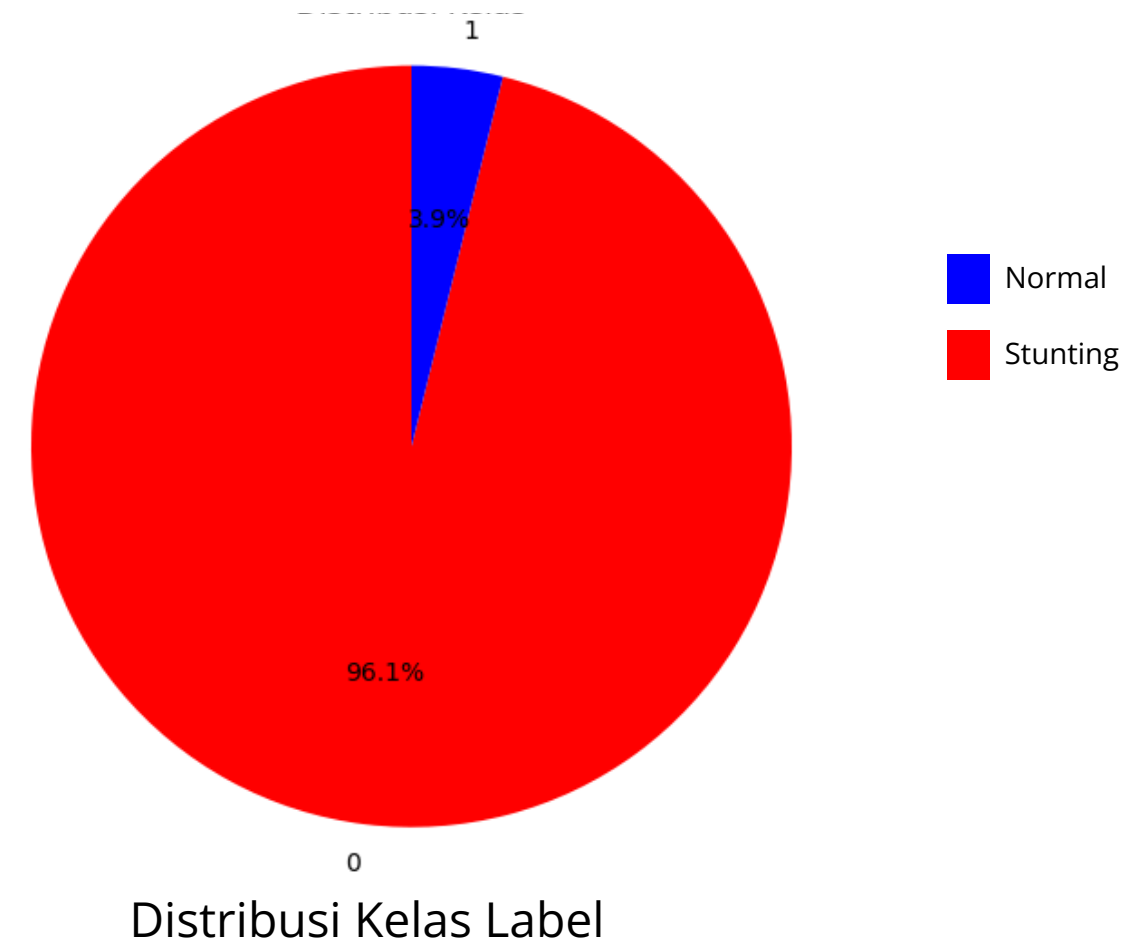
METODOLOGI



Dataset

- Data diperoleh dari Dinas Kesehatan Kabupaten Kepulauan Talaud, Sulawesi Utara berdasarkan pengukuran pada bulan Juni 2024 dengan jumlah data sebanyak 5.050 data (197 stunting dan 4.853 normal).
- Kelas label adalah TB/U karena kelas ini secara langsung menggambarkan apakah anak-anak mengalami stunting berdasarkan rasio tinggi badan dan usia mereka.

Fitur
1. Usia (U)
2. Jenis Kelamin
3. Berat Badan (BB)
4. Tinggi Badan (TB)
5. Z-Score
6. BB/TB
7. BB/U
8. TB/U



Data Preprocessing

TABLE II. DATA PREPROCESSING OVERVIEW		
Step	Description	Output Example
Feature Selection	Select the relevant features	Final Features: 8
Data Cleaning	Handle missing values by dropping incomplete rows. No duplicated data was identified	Total Data: 5,048
Feature Transformation	Age in string format 'X Years-Y Months-Z Days' converted to number of months	Age: 3 Years-11 Months-1 Day → 47 months
Feature Encoding	Features encoded into numerical categories	Nutritional status: 0: Normal, 1: Overnutrition, 2: Undernutrition
Outlier Handling	Remove outliers based on interquartile range (IQR)	Values outside Q1-Q3 replaced with median
Data Splitting	Split data into 80% training and 20% testing sets	Training: 4,038, Testing: 1,010
Oversampling	Quartile ranges for the oversampling method	Quartile: Q1 (low) to Q3 (high)

TABLE III. DATA BEFORE BALANCING		
	Normal (0)	Stunting (1)
Amount	4,853	195

TABLE IV. DATA AFTER BALANCING		
	Normal (0)	Stunting (1)
Amount	3,397	4,136

Random Forest

Random Forest adalah algoritma yang digunakan untuk mengklasifikasikan data yang bekerja dengan menggabungkan hasil beberapa pohon keputusan untuk mendapatkan satu hasil yang lebih akurat [11].

XGBoost

XGBoost adalah algoritma yang bekerja dengan cara meningkatkan pohon keputusan secara paralel dan terdistribusi [9].

Stratified K-Fold Cross Validation

Stratified K-Fold cross-validation hanya digunakan pada data latih (80%). Ini dilakukan untuk memperkirakan kinerja model dengan lebih akurat[16]

Metrik Evaluasi

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100\%$$

$$Precision = \frac{TP}{TP+FP} \times 100\%$$

$$Recall = \frac{TP}{TP+FN} \times 100\%$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\%$$

Referensi: [9] [10]

HASIL DAN PEMBAHASAN

PELATIHAN (TRAINING)

TABLE V. RANDOM FOREST PERFORMANCE FOR DATA TRAINING

Random Forest							
Fold	Precision		Recall		F1-Score		Accuracy
	0	1	0	1	0	1	
Fold 1	97.65%	98.19%	98.46%	97.26%	98.05%	97.72%	97.90%
Fold 2	97.13%	97.13%	97.53%	96.66%	97.33%	96.89%	97.13%
Fold 3	96.95%	97.71%	98.04%	96.42%	97.49%	97.06%	97.29%
Fold 4	97.20%	95.98%	96.49%	96.78%	96.84%	96.38%	96.63%
Fold 5	97.33%	97.48%	97.84%	96.90%	97.58%	97.19%	97.40%
Average	97.27%		97.24%		97.25%		97.26%

TABLE VI. XGBOOST PERFORMANCE FOR DATA TRAINING

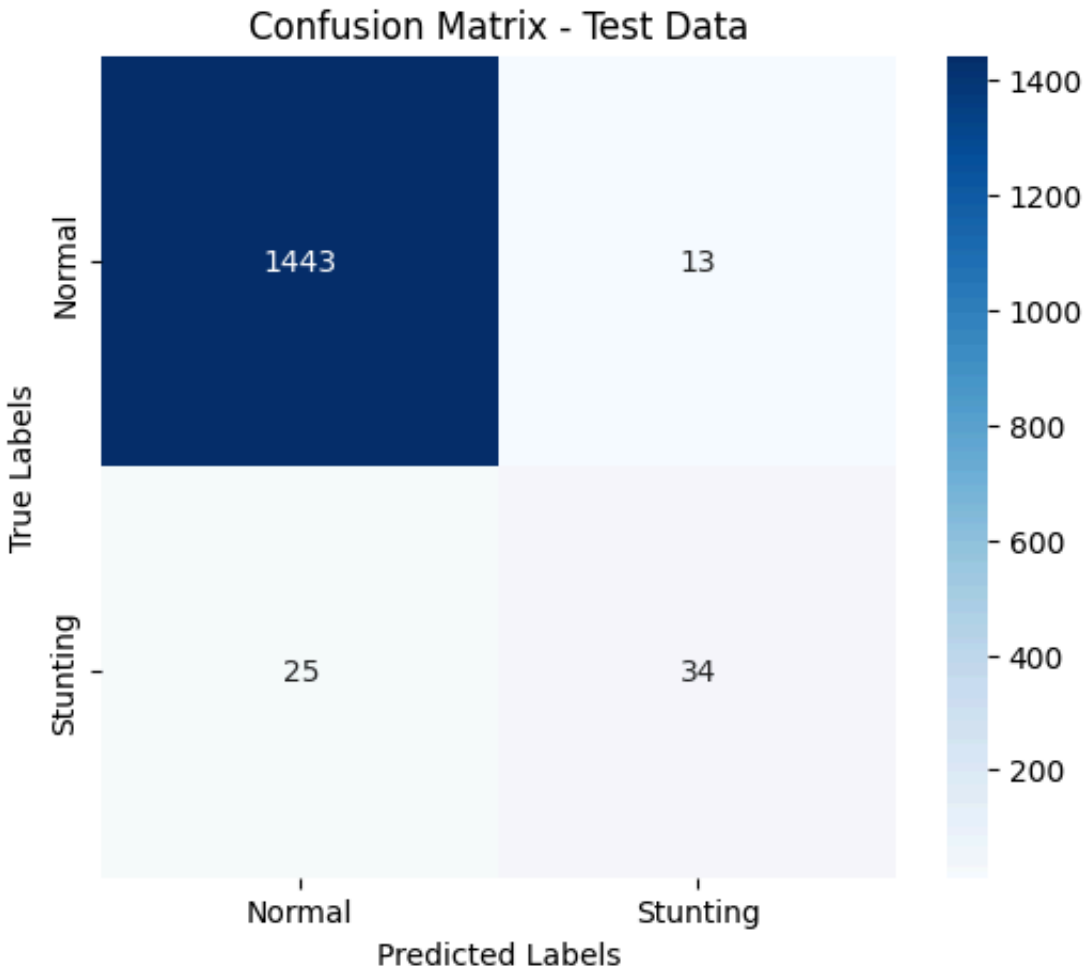
XGBoost							
Fold	Precision		Recall		F1-Score		Accuracy
	0	1	0	1	0	1	
Fold 1	98.07%	99.39%	99.49%	97.74%	98.77%	98.56%	98.67%
Fold 2	97.64%	97.84%	98.15%	97.26%	97.89%	97.55%	97.73%
Fold 3	96.75%	97.94%	98.25%	96.19%	97.50%	97.05%	97.29%
Fold 4	96.65%	97.70%	98.04%	96.07%	97.34%	96.88%	97.13%
Fold 5	97.24%	97.95%	98.25%	96.78%	97.74%	97.36%	97.57%
Average	97.71%		97.61%		97.66%		97.67%

PENGUJIAN

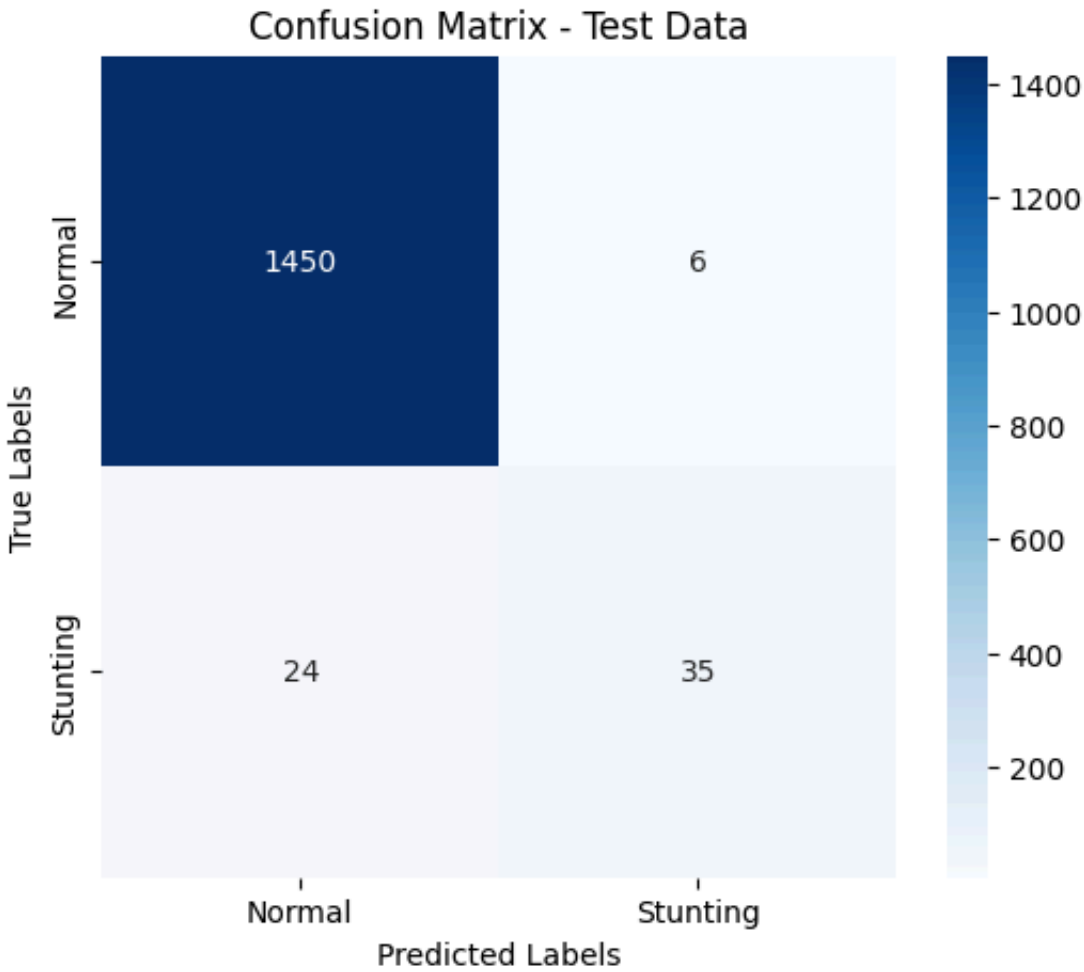
TABLE VII. MODELS PERFORMANCE FOR DATA TESTING

Model	Precision	Recall	F1-Score	Accuracy
Random Forest	85.32%	78.37%	81.43%	97.49%
XGBoost	91.87%	79.45%	84.49%	98.02%

Random Forest



XGBoost



KESIMPULAN





- Penelitian ini melatih dan menguji Random Forest dan XGBoost untuk klasifikasi stunting dengan menggunakan data yang seimbang.
- XGBoost mencapai performa pengujian tertinggi, yaitu akurasi (98,02%), presisi (91,87%), recall (79,45%), dan F1-Score (84,49%).
- Hasilnya menunjukkan bahwa XGBoost memiliki performa yang lebih baik karena mekanisme boosting-nya, yang memungkinkan model untuk menangani set data yang lebih signifikan dan kompleks.
- Random Forest menggunakan pendekatan bagging yang kurang responsif terhadap pola yang kompleks dibandingkan dengan boosting karena tidak belajar dari kesalahan prediksi sebelumnya.

REFERENSI

- [1] A. Fadilah, M. N. Pangestu, S. Lumbanbatu, and S. Defiyanti, "Pengelompokan Kabupaten/Kota Di Indonesia Berdasarkan Faktor Penyebab Stunting Pada Balita Menggunakan Algoritma KMeans," JIKO (Jurnal Informatika dan Komputer), vol. 6, no. 2, p. 223, 2022, doi: 10.26798/jiko.v6i2.581.
- [2] A. Bilqis, I. Putri, and S. A. Okdiyanti, "Analisis Kebijakan Pemerintah Kabupaten Jember Dalam Menangani Stunting Dengan Perspektif Problem Tree Analysis," JMA), vol. 2, no. 5, pp. 3031–5220, 2024, doi: 10.62281.
- [3] D. Agung Susanti, S. Patonah, F. Azizah, and S. Tinggi Ilmu Kesehatan Rajekwesi Bojonegoro, "Edukasi Orang Tua Tentang Pencegahan Stunting Dengan Perbaikan Pola Makan, Pola Asuh & Sanitasi Di Ra Nafisa Kendal Sidodadi Sukosewu," 2024.
- [4] M. A. Moreno-Ibarra, Y. Villuendas-Rey, M. D. Lytras, C. YáñezMárquez, and J. C. Salgado-Ramírez, "Classification of diseases using machine learning algorithms: A comparative study," Mathematics, vol. 9, no. 15, Aug. 2021, doi: 10.3390/math9151817.
- [5] S. Usabili and U. Indahyanti, "Pemodelan Deteksi Dini Diabetes Mellitus menggunakan Pendekatan Ensemble Learning," Manajemen Pelayanan Kesehatan, vol. 1, no. 1, p. 12, Jan. 2024, doi: 10.47134/mpk.v1i1.2924.
- [6] Y. Pratama, A. Prayitno, D. Azrian, N. Aini, Y. Rizki, and E. Rasywir, "Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma K-Nearest Neighbor," Bulletin of Computer Science Research, vol. 3, no. 1, pp. 52–56, Dec. 2022, doi: 10.47065/bulletincsr.v3i1.203.
- [7] P. Podder, S. Bharati, M. R. H. Mondal, and U. Kose, "Application of machine learning for the diagnosis of COVID-19," in Data Science for COVID-19 Volume 1: Computational Perspectives, Elsevier, 2021, pp. 175–194. doi: 10.1016/B978-0-12-824536-1.00008-3.
- [8] O. N. Chilyabanyama et al., "Performance of Machine Learning Classifiers in Classifying Stunting among Under-Five Children in Zambia," Children, vol. 9, no. 7, Jul. 2022, doi: 10.3390/children9071082.
- [9] S. Ndagijimana, I. H. Kabano, E. Masabo, and J. M. Ntaganda, "Prediction of Stunting among Under-5 Children in Rwanda Using Machine Learning Techniques," Journal of Preventive Medicine and Public Health, vol. 56, no. 1, pp. 41–49, Jan. 2023, doi: 10.3961/jpmpmh.22.388.

REFERENSI

- [10] M. I. Anugrah, J. Zeniarja, and D. S. Setiawan, "Peningkatan Performa Model Hard Voting Classifier dengan Teknik Oversampling ADASYN pada Penyakit Diabetes," *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 1, pp. 290–299, Jun. 2024, doi: 10.29408/edumatic.v8i1.25838.
- [11] Z. Jin, J. Shang, Q. Zhu, C. Ling, W. Xie, and B. Qiang, "RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Science and Business Media Deutschland GmbH, 2020, pp. 503–515. doi: 10.1007/978-3-030-62008-0_35.
- [12] G. Daffa and P. Gunawan, *Stunting Classification Analysis for Toddlers in Bojongsoang: A Data-Driven Approach*. IEEE, 2024.
- [13] R. Annisa, "Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung," *Jurnal Teknik Informatika Kaputama (JTIK)*, vol. 3, no. 1, 2019.
- [14] N. T. Hastuti and F. Budiman, "Edumatic: Jurnal Pendidikan Informatika Optimasi Klasifikasi Stunting Balita dengan Teknik Boosting pada Decision Tree," vol. 8, no. 2, pp. 655–664, 2024, doi: 10.29408/edumatic.v8i2.27913.
- [15] M. T, K. Vinoth, K. Dhilip, O. Geman, M. Margala, and M. Guduri, "The stratified K-folds cross-validation and classbalancing methods with high-performance ensemble classifiers for breast cancer classification," *Healthcare Analytics*, vol. 4, Dec. 2023, doi: 10.1016/j.health.2023.100247.
- [16] R. Rabiei, S. M. Ayyoubzadeh, S. Sohrabei, M. Esmaeili, and A. Atashi, "Prediction of Breast Cancer using Machine Learning Approaches," *J Biomed Phys Eng*, vol. 12, no. 3, pp. 297–308, Jun. 2022, doi: 10.31661/jbpe.v0i0.2109-1403.
- [17] B. Sahu, S. N. Mohanty, and S. K. Rout, "A Hybrid Approach for Breast Cancer Classification and Diagnosis," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 6, no. 20, 2019, doi: 10.4108/eai.19-12-2018.156086.
- [18] A. Vanacore, M. S. Pellegrino, and A. Ciardiello, "Fair evaluation of classifier predictive performance based on binary confusion matrix," *Comput Stat*, vol. 39, no. 1, pp. 363–383, Feb. 2024, doi: 10.1007/s00180-022-01301-9.
- [19] S. Lonang, A. Yudhana, and M. K. Biddinika, "Analisis Komparatif Kinerja Algoritma Machine Learning untuk Deteksi Stunting," *Jurnal Media Informatika Budidarma*, vol. 7, no. 4, p. 2109, Oct. 2023, doi: 10.30865/mib.v7i4.6553.



TERIMA KASIH

