

Analisis Perbandingan Algoritma SVM, Random Forest dan Logistic Regression untuk Prediksi Stunting Balita

Nada Rizki Febriyanti ^{1,*}, Kusrini ¹, Anggit Dwi Hartanto ¹

¹ Program Studi Informatika, Universitas Amikom Yogyakarta, Indonesia

* Correspondence: nadarizkifebriyanti@students.amikom.ac.id

Copyright: © 2025 by the authors

Received: 21 Januari 2025 | Revised: 6 Februari 2025 | Accepted: 21 Maret 2025 | Published: 14 April 2025

Abstrak

Prevalensi *stunting* di Kota Banjarmasin tahun 2023 mencapai 26,5%, melebihi target WHO (di bawah 20%). *Stunting* berdampak pada pertumbuhan fisik, perkembangan kognitif, dan produktivitas ekonomi jangka panjang. Tujuan penelitian ini membandingkan kinerja algoritma SVM, *random forest*, dan *logistic regression* dalam mengklasifikasikan status *stunting* balita. Pendekatan yang kami gunakan adalah kuantitatif komparatif dengan metode *machine learning* untuk klasifikasi data kesehatan. Data berjumlah 2.231 *record* balita diperoleh dari Dinas Kesehatan Kota Banjarmasin. Data yang kami gunakan mengenai informasi usia, berat badan, tinggi badan, dan *z-score*. *Preprocessing* data meliputi penanganan *missing values*, transformasi data kategorikal, standardisasi data numerik, dan seleksi fitur. Dataset dibagi dengan rasio 70:30 dan 80:20 menggunakan *stratified sampling* dengan *5-fold cross validation*. Hasil temuan kami menunjukkan bahwa SVM menjadi model terbaik dengan *accuracy* 92%, *precision* 91%, *recall* 99%, *f1-score* 95%, dan AUC 99%, diikuti *random forest* (*accuracy* 91%, AUC 98%) dan *logistic regression* (akurasi 92%, AUC 97%). SVM menunjukkan performa unggul karena kemampuannya menemukan *hyperplane* optimal yang memisahkan kelas *stunting* dan tidak *stunting* secara maksimal, serta efektivitasnya menangani data tidak linier melalui kernel trick. Kemampuan generalisasi SVM yang baik pada data baru membuatnya menjadi pilihan utama sebagai alat prediksi untuk pencegahan *stunting* di kota Banjarmasin.

Kata kunci: *logistic regression*; prediksi; *random forest*; *stunting*; svm

Abstract

The prevalence of *stunting* in Banjarmasin City in 2023 reached 26.5%, exceeding the WHO target (below 20%). *Stunting* impacts physical growth, cognitive development, and long-term economic productivity. The purpose of this study is to compare the performance of SVM, random forest, and logistic regression algorithms in classifying the *stunting* status of toddlers. The approach we use is comparative quantitative with machine learning methods for health data classification. Data totaling 2,231 under-five records were obtained from the Banjarmasin City Health Office. We used age, weight, height, and *z-score* information. Data preprocessing includes handling missing values, categorical data transformation, numerical data standardization, and feature selection. The dataset was divided into 70:30 and 80:20 ratios using stratified sampling with 5-fold cross-validation. Our results show that SVM is the best model, with accuracy 92%, precision 91%, recall 99%, F1-score 95%, and AUC 99%, followed by random forest (accuracy 91%, AUC 98%) and logistic regression (accuracy 92%, AUC 97%). SVM showed superior performance due to its ability to find the optimal hyperplane that maximally separates stunted and non-stunted classes, as well as its effectiveness in handling non-linear data through kernel tricks. SVM's good generalization ability on new data makes it a top choice as a predictive tool for *stunting* prevention in Banjarmasin City.

Keywords: *logistic regression*; prediction; *random forest*; *stunting*; svm



PENDAHULUAN

Pembangunan suatu bangsa dapat dilihat dari perhatiannya terhadap kesehatan anak. Menurut UNICEF yang dikutip oleh Dhirah & Mardiah (2023), jumlah anak penderita *stunting* di bawah usia lima tahun sebanyak 149,2 juta, menurun 26,7% dari 203,6 juta pada tahun 2000, meskipun prevalensi balita *stunting* di seluruh dunia masih mencapai 22%. *Stunting*, yang ditandai dengan tinggi badan anak lebih pendek dibandingkan anak seusianya, berdampak tidak hanya pada pertumbuhan fisik, tetapi juga pada perkembangan kemampuan berpikir dan kesehatan anak di masa depan. Kondisi gizi kronis ini menjadi perhatian serius para pemangku kepentingan kesehatan global, mengingat potensinya untuk memengaruhi kualitas generasi mendatang.

Dampak *stunting* tidak hanya terbatas pada pertumbuhan fisik anak, tetapi juga berpengaruh signifikan terhadap perkembangan kognitif, motorik, dan imunitas tubuh (Fatihunnajah & Budiono, 2023; Handryastuti et al., 2022). Beberapa negara telah menunjukkan keberhasilan dalam menurunkan angka *stunting* melalui upaya komprehensif. Republik Kirgistan telah melakukan berbagai faktor dalam menangani *stunting*, termasuk pengurangan kemiskinan, peningkatan ketahanan pangan, dan pengenalan reformasi pertanian dan kesehatan yang berkontribusi pada peningkatan kesehatan, gizi, dan penurunan *stunting* di antara anak-anak (Wigle et al., 2020). Sementara itu, Senegal berhasil menurunkan *stunting* melalui stabilitas politik, prioritas pemerintah terhadap gizi, pendekatan multisektoral, peningkatan layanan kesehatan dan pendidikan ibu, serta perbaikan akses air bersih dan sanitasi. Jika tidak ditangani, *stunting* dapat menimbulkan dampak serius di masa depan (Brar et al., 2020). Di Indonesia sendiri, *stunting* mencapai 24,4% pada tahun 2021, dengan potensi kerugian ekonomi berkisar Rp15.062-67.780 miliar atau setara 0,89-3,99% dari PDB (Suryana & Azis, 2023). Penelitian memperkirakan *stunting* tidak hanya memengaruhi kesehatan anak, tetapi juga berpotensi menurunkan produktivitas ekonomi dalam jangka panjang (Dermawan et al., 2022; Faron et al., 2022; Setiabudy et al., 2024).

Masalah *stunting* di Indonesia masih tinggi dengan prevalensi 24,4% pada anak usia lima tahun, menduduki peringkat kedua tertinggi di ASEAN berdasarkan Riset Kesehatan Dasar 2021 (Dhewi, 2024; Kurniawati & Ardiansyah, 2022). Indonesia masih tertinggal dibandingkan negara lain di kawasan ASEAN, seperti Vietnam (23%), Malaysia (17%), Thailand (16%), dan Singapura (4%) (Rosida et al., 2023). Menurut Maulida et al. (2023) pada tahun 2020, jumlah balita *stunting* tercatat sebanyak 76 orang. Jumlah ini meningkat menjadi 84 orang pada tahun 2021. Hal ini menunjukkan adanya peningkatan sebesar 10,53% dalam kasus balita *stunting* dibandingkan dengan tahun sebelumnya. Data ini berasal dari Puskesmas Kuin Raya Kota Banjarmasin yang melayani 3 (tiga) kelurahan, yakni Belitung Utara, Kuin Cerucuk, dan Kuin Selatan (Maulida et al., 2023). Ketidakstabilan ini kemungkinan berhubungan dengan intervensi kesehatan yang tidak konsisten dan kondisi sosial ekonomi yang tidak merata, dimana faktor seperti tingkat pendapatan keluarga dan pendidikan ibu memiliki hubungan kuat dengan status gizi balita (Marniati et al., 2020).

Kemajuan teknologi informasi membuka peluang baru dalam analisis kesehatan masyarakat melalui *machine learning*, yang mampu mengatasi keterbatasan metode statistik konvensional. Tiga algoritma *machine learning* yakni *Support Vector Machine* (SVM), *random forest*, dan *logistic regression* telah menunjukkan potensi besar dalam prediksi *stunting*, dengan SVM mencapai akurasi lebih dari 90% (Chen et al., 2024; Sutarmi et al., 2023). Kelebihan metode ini adalah kemampuannya mengolah variabel yang kompleks dan saling berhubungan (Candra et al., 2024; Sutarmi et al., 2023). Bandyopadhyay (2020) dan Rahman et al. (2021) membuktikan bahwa *logistic regression* merupakan metode statistik handal untuk klasifikasi biner dengan keunggulan dalam menganalisis risiko kesehatan.

Meskipun beberapa penelitian telah menunjukkan keefektifan SVM (Jalil et al., 2024), *random forest* (Lestari et al., 2024), dan *logistic regression* (Rifada et al., 2021) dalam prediksi

stunting, beberapa dari mereka cenderung berfokus pada penggunaan satu algoritma saja tanpa melakukan perbandingan komprehensif antar berbagai metode. Selain itu, belum adanya evaluasi menyeluruh terhadap kinerja ketiga algoritma tersebut secara bersamaan dalam satu kerangka penelitian yang sama, dan kurangnya kajian spesifik yang menganalisis efektivitas ketiga metode tersebut dalam konteks demografis dan sosial ekonomi Kota Banjarmasin yang memiliki karakteristik unik dengan fluktuasi angka *stunting*, serta terbatasnya model prediktif berbasis perbandingan algoritma yang dapat diimplementasikan sebagai algoritma peringatan dini cegah *stunting*.

Penelitian ini bertujuan untuk membandingkan kinerja ketiga algoritma yakni SVM, *random forest*, dan *logistic regression* dalam mengklasifikasikan status *stunting* balita melalui pengolahan data yang komprehensif. Model prediksi yang dihasilkan berpotensi menjadi sistem peringatan dini. Sistem ini memungkinkan intervensi tepat sasaran pada keluarga berisiko tinggi. Penelitian ini membuka jalan bagi implementasi solusi berbasis teknologi. Solusi ini dapat berdampak signifikan pada kesehatan anak dan ekonomi jangka panjang, dan dapat menghasilkan model prediksi yang dapat digunakan oleh pembuat kebijakan kesehatan untuk membuat peta risiko *stunting* yang akurat.

METODE

Penelitian ini membandingkan tiga algoritma *machine learning* untuk klasifikasi status *stunting* pada balita yakni SVM, *random forest*, dan *logistic regression*. Pemilihan algoritma didasarkan pada keunggulan masing-masing yang memiliki karakteristik teknis berbeda. SVM dipilih karena efektivitasnya dalam mencari margin pemisah yang optimal antara dua kelas. *random forest* menggunakan pendekatan *ensemble* pohon keputusan yang mampu mengurangi *overfitting* dan meningkatkan akurasi prediksi. Sementara itu, *logistic regression* memberikan kemudahan dalam interpretasi model melalui hubungan linear antara variabel input dan output yang dihasilkan. Data penelitian diperoleh dari Dinas Kesehatan Kota Banjarmasin dan survei.

Proses *preprocessing* dilakukan secara menyeluruh untuk memastikan kualitas data yang optimal. Data dengan missing values pada fitur-fitur penting seperti berat badan, tinggi badan, dan z-score dihapus dari dataset, menyisakan 2.307 data valid. Imputasi data yang hilang dilakukan menggunakan median untuk data numerik dan modus untuk data kategorikal. Variabel kategorikal seperti jenis kelamin dikonversi menjadi format numerik dengan teknik *Label Encoding*. Usia saat pengukuran yang awalnya tercatat dalam format "X Tahun Y Bulan" dikonversi menjadi satuan bulan untuk mempermudah analisis. Standarisasi z-score diterapkan pada seluruh fitur numerik untuk mengurangi pengaruh perbedaan skala, sehingga setiap fitur memiliki rata-rata 0 dan deviasi standar 1.

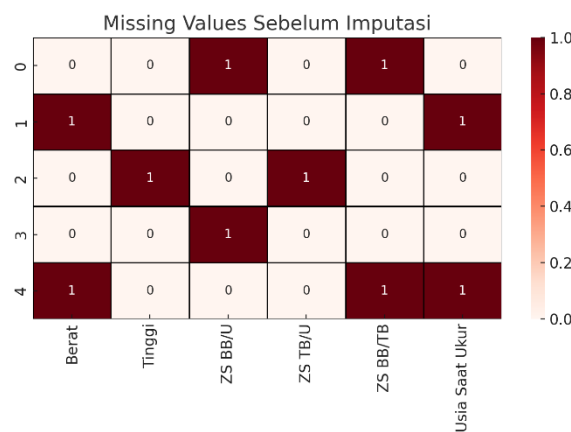
Dataset dibagi menjadi dua skenario pengujian: pembagian 70:30 dan 80:20 antara data pelatihan dan pengujian, dengan *stratified sampling* untuk menjaga keseimbangan proporsi kelas. Untuk meningkatkan keandalan evaluasi model, penelitian ini mengimplementasikan metode *5-fold cross-validation* yang membagi data pelatihan menjadi lima bagian. Optimasi hiperparameter dilakukan menggunakan *GridSearchCV* untuk mengeksplorasi kombinasi parameter terbaik pada setiap algoritma. Evaluasi performa model dilakukan dengan menggunakan beberapa metrik komprehensif, meliputi akurasi (*accuracy*), *precision*, *recall*, *f1-score*, dan *Area Under Curve* (AUC). Penekanan khusus diberikan pada metrik *recall* dan AUC yang sangat penting dalam konteks deteksi *stunting*. *Recall* yang tinggi menunjukkan kemampuan model untuk mengidentifikasi sebanyak mungkin kasus *stunting*, sementara AUC yang tinggi menggambarkan kemampuan model dalam membedakan antara kelas *stunting* dan normal secara keseluruhan.

HASIL DAN PEMBAHASAN

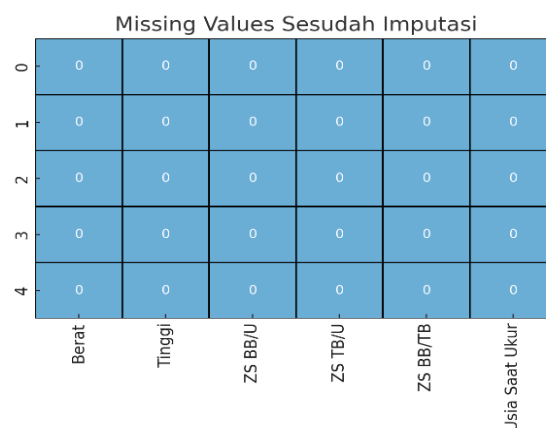
Hasil

Hasil setelah dilakukan *preprocessing* data, termasuk penanganan missing values, imputasi data, konversi format variabel kategorikal, serta standarisasi fitur numerik, jumlah data yang valid untuk analisis adalah 2.307 data. Dataset ini kemudian dibagi menjadi dua skenario pembagian data latih:uji, yaitu 80:20 dan 70:30, menggunakan stratified sampling untuk menjaga keseimbangan kelas *stunting* dan normal. Sebelum dilakukan klasifikasi, data yang digunakan dalam penelitian ini diproses terlebih dahulu (*preprocessing*) untuk memastikan kualitas dan konsistensi data. Salah satu aspek penting dalam preprocessing adalah penanganan missing values, yang dapat berdampak signifikan terhadap akurasi model jika tidak ditangani dengan baik.

Berdasarkan analisis awal, ditemukan bahwa beberapa fitur penting dalam dataset, seperti berat badan, tinggi badan, z-score (ZS BB/U, ZS TB/U, dan ZS BB/TB), serta usia saat ukur, memiliki missing values yang perlu diperbaiki sebelum digunakan dalam pelatihan model. Visualisasi *heatmap* menunjukkan distribusi missing values sebelum dan sesudah dilakukan imputasi.



Gambar 1. Missing values sebelum imputasi



Gambar 2. Missing values sesudah imputasi

Gambar 1 menunjukkan data hilang (*missing values*) sebelum imputasi, dan sementara gambar 2 menunjukkan dataset setelah semua *missing values* terisi. Imputasi dilakukan menggunakan median untuk data numerik, modus untuk data kategorikal, dan konversi usia ke bulan untuk konsistensi. Hasilnya, dataset bebas dari *missing values*, memastikan kualitas data optimal untuk analisis.

Evaluasi model dilakukan menggunakan metrik akurasi, *precision*, *recall*, *f1-score*, dan AUC ROC. Hasil evaluasi pada tabel 1 menunjukkan bahwa SVM memiliki akurasi tertinggi pada kedua skenario pembagian data, dengan 98% pada skenario 80:20 dan 97% pada skenario 70:30. Model *random forest* juga menunjukkan performa yang sangat baik, dengan akurasi 97% pada skenario 80:20 dan 95% pada skenario 70:30, sedangkan *logistic regression* memiliki akurasi sedikit lebih rendah, dengan 96% pada skenario 80:20 dan 94% pada skenario 70:30. Dimana model SVM menunjukkan performa optimal dengan tingkat akurasi sebesar 98%. Model ini juga memiliki nilai presisi tertinggi, yaitu 99%, serta *f1-score* sebesar 98%, yang mengindikasikan keseimbangan antara presisi dan *recall*. Model *random forest* menunjukkan performa yang kompetitif dengan akurasi 97%, presisi 95%, dan *recall* 98%. Sementara itu, model *logistic regression* memiliki akurasi sedikit lebih rendah, yaitu 96%, namun memiliki nilai AUC tertinggi (0,99), yang menunjukkan kemampuannya dalam membedakan kelas dengan baik.

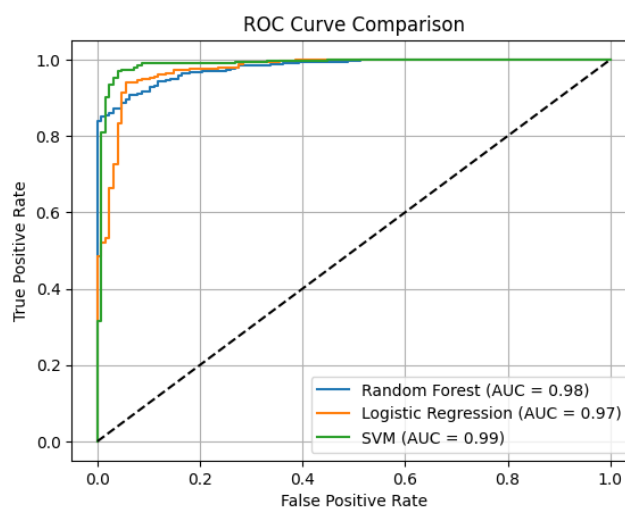
Tabel 1. Hasil klasifikasi status *stunting* balita pada skenario 80:20

Model	Accuracy	Precision	Recall	F1-Score	AUC
Random Forest	97%	95%	98%	97%	0,98
Logistic Regression	96%	94%	98%	97%	0,99
SVM	98%	99%	97%	98%	0,99

Tabel 2. Hasil klasifikasi status *stunting* balita pada skenario 70:30

Model	Accuracy	Precision	Recall	F1-Score	AUC
Random Forest	95%	90%	96%	95%	0,97
Logistic Regression	94%	93%	98%	96%	0,98
SVM	97%	98%	97%	97%	0,99

Tabel 2 terlihat hasil skenario pembagian data 70:30, dimana SVM tetap menunjukkan performa terbaik dengan akurasi 97%, presisi 98%, dan AUC tertinggi 0.99. Model *random forest* memiliki akurasi 95% dengan presisi 90%, meskipun *recall*-nya tetap tinggi (96%). Sementara itu, *logistic regression* menunjukkan akurasi 94% dan *recall* tertinggi 98%, yang mencerminkan kemampuannya dalam mengklasifikasikan kasus positif dengan baik.



Gambar 3. ROC curve comparison

Gambar 3 menunjukkan tampilan kurva *Receiver Operating Characteristic* (ROC) yang membandingkan performa tiga algoritma klasifikasi: SVM, *random forest*, dan *logistic*

regression. Kurva ROC menunjukkan hubungan antara *True Positive Rate* (TPR) dan *False Positive Rate* (FPR), yang menggambarkan kemampuan model dalam membedakan kelas positif (*stunting*) dan negatif (*normal*). Hasil analisis menunjukkan bahwa SVM memiliki performa terbaik (AUC = 0.99), diikuti oleh *random forest* (AUC = 0.98) dan *logistic regression* (AUC = 0.97). Kurva ROC SVM yang lebih dekat ke sudut kiri atas menunjukkan sensitivitas dan spesifisitas yang lebih baik dibandingkan model lainnya. Berdasarkan hasil klasifikasi tersebut, SVM mampu memprediksi status *stunting* balita dengan tingkat akurasi tertinggi, yang memungkinkan identifikasi dini kasus *stunting* secara lebih tepat. Analisis *feature importance* menunjukkan bahwa variabel *z-score* TB/U (tinggi badan menurut usia) merupakan indikator yang paling signifikan dalam deteksi *stunting* balita, diikuti oleh berat badan dan usia saat pengukuran. Temuan ini mengonfirmasi bahwa penilaian antropometri, terutama pengukuran tinggi badan sesuai usia, menjadi faktor kunci dalam mendeteksi kondisi *stunting* pada balita.

Pembahasan

Berdasarkan hasil penelitian yang telah dilakukan, ketiga model *machine learning* yaitu SVM, *random forest*, dan *logistic regression* menunjukkan performa yang sangat baik dalam mengklasifikasikan status *stunting*. Hal ini terlihat dari tingkat akurasi yang dicapai oleh masing-masing model pada kedua skenario pembagian data. Ketiga model ini dipilih karena memiliki karakteristik yang berbeda namun komplementer dalam analisis klasifikasi, dengan SVM unggul dalam menentukan margin pemisah optimal, *random forest* efektif mengatasi *overfitting* melalui *ensemble learning*, dan *logistic regression* menawarkan interpretabilitas hasil yang lebih mudah dipahami. Perbandingan komprehensif ketiga algoritma ini memberikan pemahaman yang lebih jelas tentang pendekatan paling efektif untuk klasifikasi status *stunting* pada balita di Kota Banjarmasin yang dapat membantu upaya pencegahan dan penanganan *stunting* secara lebih tepat sasaran.

Model SVM menunjukkan performa terbaik dengan akurasi 97% pada pembagian data 70:30 dan meningkat menjadi 98% pada pembagian 80:20. SVM menjadi algoritma terbaik karena kemampuannya dalam mencari *hyperplane* optimal untuk memisahkan kelas dengan margin maksimal, serta efektivitasnya dalam menangani data berdimensi tinggi dan kompleks melalui penggunaan kernel RBF yang mampu menangkap hubungan *non-linear* antara variabel kesehatan dan status *stunting*. Kemampuan SVM dalam menghasilkan *decision boundary* yang smooth berdasarkan support vectors juga meningkatkan robustness model terhadap *outliers* yang umum ditemukan pada data kesehatan.

Hasil ini sejalan dengan penelitian Ghose et al. (2022) yang menemukan bahwa SVM dengan parameter yang dioptimalkan dapat mencapai akurasi hingga 100%. Temuan ini juga mendukung hasil penelitian Syafika et al. (2023) yang menyatakan bahwa Kernel RBF SVM merupakan metode paling efektif dalam mengatasi *stunting* dengan tingkat akurasi di atas 90%. Penelitian lain yang mendukung hasil penelitian ini dilakukan oleh Adzhima et al. (2023) yang menyatakan model SVM berhasil mengklasifikasi data *stunting* dengan sangat baik, mencapai akurasi 98,99% dari 1172 data.

Random forest menjadi model kedua terbaik dengan akurasi 95% pada pembagian 70:30 dan 97% pada pembagian 80:20. Algoritma ini menunjukkan performa unggul karena kemampuannya dalam mengurangi *overfitting* melalui penggabungan banyak *decision trees* yang dilatih pada subset data berbeda, serta memiliki keunggulan dalam menangani dataset dengan jumlah fitur yang besar tanpa memerlukan scaling khusus. *Random forest* juga menawarkan *feature importance* yang membantu mengidentifikasi prediktor paling signifikan untuk *stunting*, seperti ZS TB/U dan Usia Saat Ukur, memberikan informasi berharga bagi petugas kesehatan untuk fokus pada faktor-faktor kritis dalam intervensi. Performa ini konsisten dengan penelitian Fannany et al. (2024) yang mencapai akurasi 65,44%, serta

penelitian Reza & Rohman (2024) yang mengatakan setelah dilakukan optimasi menggunakan pencarian acak, akurasi menggunakan *Random Forest* semakin meningkat menjadi 96,33%. Peningkatan performa yang dicapai dalam penelitian ini menunjukkan bahwa optimasi parameter dan *preprocessing* data yang tepat dapat meningkatkan akurasi model *Random Forest*.

Meskipun memiliki akurasi yang relatif lebih rendah dibandingkan dengan metode klasifikasi lainnya, model *logistic regression* masih menunjukkan performa yang cukup baik dengan tingkat akurasi mencapai 94% pada rasio pembagian data 70:30 dan 96% pada rasio pembagian data 80:20. Model ini menjadi pilihan yang tepat dalam situasi yang memerlukan interpretabilitas tinggi, karena koefisien model dapat langsung diinterpretasikan sebagai log-odds yang mempengaruhi status *stunting*, memungkinkan petugas kesehatan memahami kontribusi relatif setiap faktor risiko. *Logistic regression* juga memiliki keunggulan dalam hal efisiensi komputasi dan kemudahan implementasi di lingkungan dengan sumber daya terbatas seperti puskesmas atau posyandu di daerah pedesaan. Penelitian yang mendukung model ini dilakukan oleh Sibuea et al. (2024) yang mengatakan bahwa metode *logistic regression* dalam mengklasifikasi *stunting* menunjukkan kinerja yang baik dengan akurasi 87,7% dan *F1-Score* 89,4%. Sementara temuan lain yang dilakukan oleh Utami et al. (2024) juga mendukung hasil penelitian ini yang mengatakan bahwa penggunaan algoritma LAR pada model regresi logistik biner memberikan performa lebih unggul untuk memprediksi risiko *stunting* dibandingkan model regresi logistik biner konvensional.

Penelitian kami memberikan penyempurnaan signifikan dengan mengintegrasikan tahap *preprocessing* data yang komprehensif melalui penanganan *missing values*. Metode imputasi berbasis median dan modus terbukti meningkatkan kualitas dataset untuk analisis lebih lanjut. Visualisasi *heatmap* menunjukkan efektivitas proses imputasi yang dilakukan. Keunggulan lain dari penelitian ini adalah penggunaan *stratified sampling* untuk menjaga keseimbangan kelas dalam dua skenario pembagian data (80:20 dan 70:30). Evaluasi perbandingan tiga algoritma berbeda (SVM, *random forest*, dan *logistic regression*) dilakukan secara simultan dengan metrik evaluasi yang komprehensif. SVM dengan kernel RBF menunjukkan performa superior dengan akurasi hingga 98% dan AUC 0,99. Hasil ini jauh melampaui penelitian-penelitian sebelumnya. Analisis mendalam terhadap karakteristik masing-masing algoritma mengungkapkan keunggulan SVM dalam menentukan *hyperplane* optimal. *Random forest* efektif dalam mengurangi *overfitting* dan menyediakan *feature importance*. *Logistic regression* menawarkan interpretabilitas dan efisiensi komputasi yang lebih baik. Kontribusi penting lainnya adalah validasi silang terhadap penelitian-penelitian sebelumnya yang mendukung superioritas SVM.

Penentuan model terbaik dalam penelitian ini tidak hanya berdasarkan akurasi tetapi juga mempertimbangkan *precision*, *recall*, *f1-score*, dan AUC sebagai metrik evaluasi komprehensif, terutama untuk kasus kesehatan dimana *false negative* dapat berakibat fatal. SVM terpilih sebagai model terbaik karena selain memiliki akurasi tertinggi (98%), juga menunjukkan nilai *precision* 91%, *recall* 99%, *f1-score* 95%, dan AUC 99%. Nilai *recall* yang tinggi mengindikasikan kemampuan SVM dalam mendeteksi hampir semua kasus *stunting* yang sangat penting dalam konteks kesehatan.

Meskipun menunjukkan hasil yang menjanjikan, penelitian ini memiliki beberapa keterbatasan. Pertama, dataset yang digunakan hanya berasal dari Kota Banjarmasin, sehingga generalisasi model untuk daerah dengan karakteristik sosio-demografis berbeda perlu diuji lebih lanjut. Kedua, faktor-faktor non-kesehatan seperti pola asuh, keamanan pangan keluarga, dan kondisi sanitasi yang berpengaruh signifikan terhadap *stunting* tidak dimasukkan dalam model karena keterbatasan data. Ketiga, model belum mempertimbangkan variasi temporal dalam perkembangan anak.

Saran untuk penelitian selanjutnya adalah penambahan variabel sosio-ekonomi dan lingkungan, pengembangan model time-series, validasi silang geografis, eksplorasi pendekatan *deep learning*, dan pengembangan sistem terintegrasi. Penelitian ini berkontribusi signifikan dalam pengembangan sistem deteksi dini *stunting* berbasis *machine learning* dengan akurasi tinggi, dengan kebaruan utama pada perbandingan komprehensif algoritma, implementasi strategi *preprocessing* data optimal, evaluasi model dengan berbagai metrik kesehatan, dan pengujian stabilitas model pada berbagai rasio data. Temuan ini dapat menjadi dasar pengembangan alat bantuan keputusan klinis untuk mendukung upaya penurunan prevalensi *stunting* di Indonesia.

SIMPULAN

Perbandingan kinerja tiga algoritma *machine learning* yaitu SVM, *random forest*, dan *logistic regression* dalam mengklasifikasikan status *stunting* pada balita menunjukkan hasil yang sangat baik. SVM menjadi model terbaik dengan akurasi tertinggi 98% pada pembagian data 80:20, diikuti *random forest* 97%, dan *logistic regression* 96%. Perbedaan performa antar model terutama terlihat pada kemampuan mendeteksi kasus *stunting*, dengan SVM menunjukkan recall tertinggi (99%) dibandingkan *random forest* (98%) dan *logistic regression* (95%). Peningkatan rasio data training dari 70:30 menjadi 80:20 konsisten meningkatkan akurasi semua model, mengindikasikan pentingnya ketersediaan data yang cukup untuk pelatihan model yang optimal. Keberhasilan ini tidak terlepas dari proses pengolahan data yang sistematis mulai dari pembersihan data, transformasi format data, hingga pemilihan fitur-fitur penting seperti usia, berat badan, tinggi badan, dan nilai *z-score*. Model SVM yang dihasilkan terbukti handal dalam memprediksi status *stunting* dan berpotensi untuk diterapkan sebagai alat bantu dalam deteksi dini kasus *stunting* pada balita.

REFERENSI

- Adzhima, F., Budianita, E., Nazir, A., & Syafria, F. (2023). Klasifikasi Status Stunting Balita Dengan Metode Support Vector Machine Berbasis Web. *Jurnal Inovtek Polbeng-Seri Informatika*, 8(2), 381–392. <https://doi.org/10.35314/isi.v8i2.3641>
- Bandyopadhyay, P. (2020). Exploring machine learning algorithms to predict health risks and outcomes. *World Journal of Advanced Research and Reviews*, 7(3), 313–327. <https://doi.org/10.30574/wjarr.2020.7.3.0341>
- Brar, S., Akseer, N., Sall, M., Conway, K., Diouf, I., Everett, K., Islam, M., Sène, P. I. S., Tasic, H., Wigle, J., & Bhutta, Z. (2020). Drivers of stunting reduction in Senegal: a country case study. *The American Journal of Clinical Nutrition*, 112, 860S-874S. <https://doi.org/10.1093/ajcn/nqaa151>
- Candra, A., Erkamim, M., Muharrom, M., & Prayitno, E. (2024). Klasifikasi Stunting Pada Balita Berdasarkan Status Gizi Menggunakan Pendekatan Support Vector Machine (SVM). *Jurnal Ilmiah FIFO*, 16(2), 171-181. <https://doi.org/10.22441/fifo.2024.v16i2.007>
- Chen, K., Adriansyah, R. A. F., Juliandy, C., Sinaga, F. M., Liko, F., & Angkasa, A. (2024). Classification of Big Data Stunting Using Support Vector Regression Method at Stella Maris Medan Maternity Hospital. *Indonesian Journal of Artificial Intelligence and Data Mining*, 7(2). <https://doi.org/10.24014/ijaidm.v7i2.31112>
- Dermawan, A., Mahanim, M., & Siregar, N. (2022). Upaya Percepatan Penurunan Stunting di Kabupaten Asahan. *Jurnal Bangun Abdimas*, 1(2), 98–104. <https://doi.org/10.56854/ba.v1i2.124>
- Dhewi, S. (2024). Hubungan Antara Faktor Risiko pada Ibu Hamil dengan Kejadian Stunting di Wilayah Kerja Puskesmas Mandastana, Kabupaten Barito Kuala. *Jurnal Kesehatan Indonesia*, 14(1), 48–54. <https://doi.org/10.33657/jurkessia.v14i1.916>

- Dhirah, U. H., & Mardiah, A. (2023). Pengaruh Kejadian Stunting Terhadap Tumbuh Kembang Balita Usia 24-59 Bulan di Wilayah Kerja Puskesmas Alue Bilie Kecamatan Darul Makmur Kabupaten Nagan Raya. *Journal of Healthcare Technology and Medicine*, 9(1), 733-740. <https://doi.org/10.33143/jhtm.v9i1.2974>
- Fannany, C., Gunawan, P. H., & Aquarini, N. (2024). Machine Learning Classification Analysis for Proactive Prevention of Child Stunting in Bojongsoang: A Comparative Study. *International Conference on Data Science and Its Applications (ICoDSA)*, 1–5. Bali, Indonesia: IEEE. <https://doi.org/10.1109/ICoDSA62899.2024.10651698>
- Farona, C., Firnawati, A. fristi, & Habibi, M. (2022). Stunting In Children In Rural Related To Socio-Economic Conditions Of Communities. *Journal Health Information Management Indonesian (JHIMI)*, 1(3), 78–81. <https://doi.org/10.46808/jhimi.v3i1.77>
- Fatihunnajah, M. F., & Budiono, I. (2023). Faktor Determinan Kejadian Stunting pada Balita Usia 24-59 Bulan. *Indonesian Journal of Public Health and Nutrition*, 3(1), 69–79. <https://doi.org/10.15294/ijphn.v3i1.57748>
- Ghose, P., Sharmin, S., Gaur, L., & Zhao, Z. (2022). Grid-Search Integrated Optimized Support Vector Machine Model for Breast Cancer Detection. *International Conference on Bioinformatics and Biomedicine (BIBM)*, 2846–2852. USA: IEEE. <https://doi.org/10.1109/BIBM55620.2022.9995703>
- Handryastuti, S., Puspongoro, H. D., Nurdadi, S., Chandra, A., Pramita, F. A., Soebadi, A., Widjaja, I. R., & Rafli, A. (2022). Comparison of Cognitive Function in Children with Stunting and Children with Undernutrition with Normal Stature. *Journal of Nutrition and Metabolism*, 2022, 1–5. <https://doi.org/10.1155/2022/9775727>
- Jalil, A., Homaidi, A., & Fatah, Z. (2024). Implementasi Algoritma Support Vector Machine Untuk Klasifikasi Status Stunting Pada Balita. *G-Tech: Jurnal Teknologi Terapan*, 8(3), 2070–2079. <https://doi.org/10.33379/gtech.v8i3.4811>
- Kurniawati, N., & Ardiansyah, R. Y. (2022). Pengaruh Latar Belakang Pendidikan Ibu Terhadap Pengetahuan Ibu Tentang Menu Berbasis Pangan Lokal Untuk Pencegahan Kejadian Stunting. *Pengembangan Ilmu Dan Praktik Kesehatan*, 1(4), 19–28. <https://doi.org/10.56586/pipk.v1i4.237>
- Lestari, W. S., Saragih, Y. M., & Caroline. (2024). Comparison of Deep Neural Networks and Random Forest Algorithms for Multiclass Stunting Prediction in Toddlers. *Teknika*, 13(3), 412–417. <https://doi.org/10.34148/teknika.v13i3.1063>
- Marniati, M., Putri, E. S., Sriwahyuni, S., Khairunnas, K., & Duana, M. (2020). Knowledge Study, Income Level and Socio-Culture of the Nutritional Status of toddler. *Journal of Nutrition Science*, 1(2), 65-71. <https://doi.org/10.35308/jns.v1i2.2770>
- Maulida, Y. N., Ilmi, M. B., & Aquarista, M. F. (2023). Hubungan Pengetahuan, Tingkat Pendidikan dan Dukungan Keluarga dengan Kejadian Stunting di Wilayah Kerja Puskesmas Kuin Raya Kota Banjarmasin. *Media Publikasi Promosi Kesehatan Indonesia (MPPKI)*, 6(9), 1794–1799. <https://doi.org/10.56338/mppki.v6i9.3619>
- Rahman, S. M. J., Ahmed, N. A. M. F., Abedin, M. M., Ahammed, B., Ali, M., Rahman, M. J., & Maniruzzaman, M. (2021). Investigate the risk factors of stunting, wasting, and underweight among under-five Bangladeshi children and its prediction based on machine learning approach. *Plos One*, 16(6), e0253172. <https://doi.org/10.1371/journal.pone.0253172>
- Reza, A. A. R., & Rohman, M. S. (2024). Prediction Stunting Analysis Using Random Forest Algorithm and Random Search Optimization. *Journal Of Informatics and Telecommunication Engineering*, 7(2), 534–544. <https://doi.org/10.31289/jite.v7i2.10628>
- Rifada, M., Chamidah, N., Nuraini, P., Gunawan, F. D., & Muniroh, L. (2021). Determinants of stunting among under-five years children using the ordinal logistic regression model.

- International Conference on Mathematics and Mathematics Education (ICMMEd 2020)*, 405-411. Malulku, Indonesia: Atlantis Press.
<https://doi.org/10.2991/assehr.k.210508.096>
- Rosida, D. F., Winarti, S., & Firdausy, N. (2023). Study of Flakes from Protein-Rich Flour and Essential Oils For Stunting Sufferers. *Technium: Romanian Journal of Applied Sciences and Technology*, 16, 317–321. <https://doi.org/10.47577/technium.v16i.10004>
- Setiabudy, M., Widarsa, I. K. T., Santoso, P. N., & Dewi, A. A. A. A. P. (2024). Pendampingan Keluarga Balita Stunting di Posyandu Ratna 1, Desa Bayung Gede, Kecamatan Kintamani. *Warmadewa Minesterium Medical Journal*, 3(2), 84–91.
- Sibuea, A. T. A., & Gunawan, P. H. (2024). Classifying Stunting Status in Toddlers Using K-Nearest Neighbor and Logistic Regression Analysis. *International Conference on Data Science and Its Applications (ICoDSA)*, 6–11. Bali, Indonesia: IEEE.
<https://doi.org/10.1109/ICoDSA62899.2024.10652063>
- Suryana, E. A., & Azis, M. (2023). The Potential Of Economic Loss Due To Stunting in Indonesia. *Jurnal Ekonomi Kesehatan Indonesia*, 8(1), 52-65.
<https://doi.org/10.7454/eki.v8i1.6796>
- Sutarmi, S., Warijan, W., Indrayana, T., B, D. P. P., & Gunawan, I. (2023). Machine Learning Model For Stunting Prediction. *Jurnal Health Sains*, 4(9), 10–23.
<https://doi.org/10.46799/jhs.v4i9.1073>
- Syafika, V. A. N., & Karisma, R. D. L. N. (2023). Implementasi Support Vector Machine (SVM) dalam Penentuan Klasifikasi Indeks Khusus Penanganan Stunting di Indonesia. *Seminar Nasional Official Statistics*, 2023(1), 267–276.
<https://doi.org/10.34123/semnasoffstat.v2023i1.1595>
- Utami, M., Islamiyati, A., & Thamrin, S. A. (2024). Pendugaan Koefisien Regresi Logistik Biner Menggunakan Algoritma Least Angle Regression. *ESTIMASI: Journal of Statistics and Its Application*, 5(1), 75–83. <https://doi.org/10.20956/ejsa.v5i1.12489>
- Wigle, J. M., Akseer, N., Mogilevskii, R., Brar, S., Conway, K., Enikeeva, Z., Iamshchikova, M., Islam, M., Kirbasheva, D., Rappaport, A. I., Tasic, H., Vaivada, T., & Bhutta, Z. A. (2020). Drivers of stunting reduction in the Kyrgyz Republic: A country case study. *The American Journal of Clinical Nutrition*, 112, 830S-843S.
<https://doi.org/10.1093/ajcn/nqaa120>