

UAS_MachineLearning_3012300 24_JIHADAKBAR-2.pdf

by louionfu fugu

Submission date: 22-Jan-2026 03:32AM (UTC+0900)

Submission ID: 2856600796

File name: UAS_MachineLearning_301230024_JIHADAKBAR-2.pdf (566.5K)

Word count: 3460

Character count: 22229

Perbandingan Metode K-Nearest Neighbor dan Random Forest dalam Klasifikasi Status Stunting Anak Berdasarkan Data Antropometri

Jihad Akbar¹

Prodi Teknik Informatika – Fakultas Teknologi Informasi
Universitas Bale Bandung, Jl. R.A.A Wiranata Kusumah No.7, Baleendah,
Kec. Baleendah, Kabupaten Bandung, Jawa Barat 40375
jihadar165@gmail.com rektorat@unibba.ac.id

Abstrak

Stunting merupakan masalah gizi jangka panjang pada anak yang dapat menghambat pertumbuhan fisik serta perkembangan kognitif dalam jangka panjang. Penentuan status stunting biasanya dilakukan melalui pengukuran antropometri secara manual, namun metode ini berpotensi menimbulkan ketidakkonsistenan pada hasil klasifikasi. Untuk meningkatkan ketepatan penentuan status stunting, penelitian ini mengusulkan pendekatan berbasis machine learning dengan membandingkan dua algoritma, yaitu **K-Nearest Neighbor (KNN)** dan **Random Forest**. Data yang digunakan berupa data antropometri anak dan dibagi menjadi empat kategori status gizi, yaitu **Normal**, **Tall**, **Stunted**, dan **Severely Stunted**. Proses pengolahan data dilakukan menggunakan metode **supervised learning** dengan evaluasi performa melalui metrik **accuracy, precision, recall, dan F1-score**, serta analisis tambahan menggunakan **confusion matrix** dan **ROC Curve**. Berdasarkan hasil pengujian, algoritma **Random Forest** memberikan performa lebih tinggi dibandingkan KNN. Model **Random Forest** kemudian diterapkan dalam aplikasi berbasis **Streamlit** untuk membantu klasifikasi status stunting pada anak.

Kata kunci: stunting, machine learning, klasifikasi, K-Nearest Neighbor, Random Forest

Abstract

Stunting is a long-term nutritional problem in children that may negatively impact both physical growth and long-term cognitive development. In many cases, stunting classification is still performed using manual anthropometric measurements, which can lead to inconsistency during the decision-making process. To improve classification reliability, this study applies a machine learning approach by comparing **K-Nearest Neighbor (KNN)** and **Random Forest** methods. The dataset consists of children's anthropometric attributes and is divided into four categories, namely **Normal**, **Tall**, **Stunted**, and **Severely Stunted**. A supervised learning scheme is implemented and evaluated using **accuracy, precision, recall, and F1-score**, supported by **confusion matrix** and **ROC curve** analysis. The experimental results indicate that **Random Forest** achieves better overall performance than KNN. Therefore, the **Random Forest** model is integrated into a **Streamlit-based application** as a decision-support tool for stunting classification.

Keywords: stunting, machine learning, classification, K-Nearest Neighbor, Random Forest

PENDAHULUAN

Stunting merupakan kondisi kekurangan gizi yang terjadi dalam periode waktu panjang pada anak, terutama pada masa pertumbuhan. Permasalahan ini tidak hanya berdampak pada tinggi badan, tetapi juga dapat mempengaruhi kemampuan kognitif dan produktivitas anak ketika dewasa. Sejumlah penelitian menunjukkan bahwa indikator antropometri, khususnya tinggi badan terhadap umur, menjadi variabel utama dalam menentukan status stunting pada anak (Andriyani et al., 2023; Atik et al., 2025; Ndagijimana et al., 2023). Dengan kemajuan teknologi, penerapan pendekatan *machine learning* berbasis data antropometri semakin banyak dimanfaatkan untuk mendukung proses klasifikasi status stunting secara lebih objektif, konsisten, dan efisien dibandingkan metode manual (Lestari et al., 2024; Mohammed et al., 2022; Putri & Arminarahmah, 2024; Uddin et al., 2022).

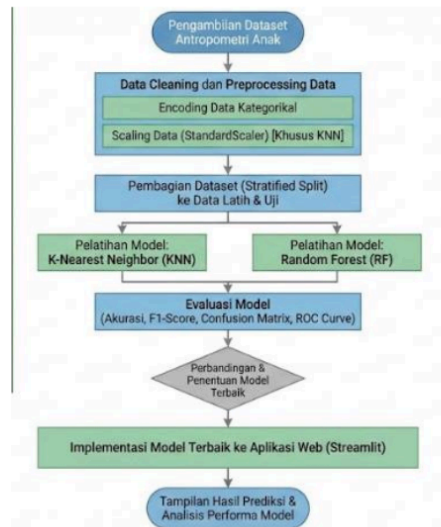
Meskipun berbagai metode telah dikembangkan, penentuan status stunting di lapangan umumnya masih dilakukan secara manual melalui perhitungan antropometri, yang berpotensi menimbulkan ketidakkonsistenan data dan kesalahan interpretasi hasil. Selain itu, variasi karakteristik dataset, distribusi kelas yang tidak seimbang, serta pemilihan algoritma yang kurang sesuai dapat memengaruhi kinerja model klasifikasi, khususnya pada kasus dengan banyak kelas (*multiclass*) dan data berskala besar (Fitri et al., n.d.; Rosta et al., 2023; Yunus et al., 2025). Beberapa penelitian juga mengungkapkan bahwa model dengan performa tinggi pada satu dataset belum tentu memberikan hasil optimal ketika diterapkan pada dataset yang berbeda, sehingga diperlukan proses evaluasi dan perbandingan algoritma secara lebih sistematis (Marleny & Zulfadhilah, 2023; Status et al., 2023).

Berbagai penelitian sebelumnya telah menerapkan sejumlah algoritma *machine learning* dalam klasifikasi status stunting, maupun stunting pada anak, antara lain *K-Nearest Neighbor*, *Naïve Bayes*, *Support Vector Machine*, *Random Forest*, hingga pendekatan *deep learning* (Bimawan et al., 2024; Lestari et al., 2024; Wahidin & Andika, 2024). Meski begitu, kebanyakan penelitian masih hanya menerapkan satu algoritma saja tanpa membandingkannya secara langsung dengan model lain, sehingga pemilihan algoritma terbaik untuk mengklasifikasikan status stunting menjadi sulit ditentukan. Selain itu, tidak semua penelitian

mengimplementasikan model yang telah dikembangkan ke dalam sistem aplikasi yang memiliki nilai praktis dalam penggunaannya (Fadilla et al., 2024; Vlachas et al., 2022). Berdasarkan celah penelitian tersebut, studi ini mengusulkan perbandingan langsung antara algoritma *K-Nearest Neighbor* sebagai model dasar dan *Random Forest* sebagai model pengembangan, dengan menerapkan skema klasifikasi empat kelas status stunting serta evaluasi performa yang komprehensif.

Sebagai solusi, penelitian ini membangun model klasifikasi status stunting anak berbasis *machine learning* menggunakan data antropometri dan membandingkan performa algoritma KNN dan Random Forest untuk menentukan model terbaik. Tujuan penelitian ini adalah memperoleh model klasifikasi yang akurat dan stabil serta mengimplementasikannya ke dalam aplikasi web berbasis Streamlit. Output yang dihasilkan meliputi model klasifikasi, hasil evaluasi perbandingan performa algoritma, serta aplikasi web interaktif. Outcome yang diharapkan adalah tersedianya sistem pendukung yang mampu membantu proses klasifikasi status stunting anak secara cepat, konsisten, dan berbasis data, sehingga dapat dimanfaatkan sebagai alat bantu pengambilan keputusan di bidang kesehatan anak (Aprilia et al., n.d.; Fauzan & Rosita, 2025; Hendy et al., 2025).

METODE PENELITIAN



Gambar 1. Flowchart Klasifikasi Status Stunting Anak

16

2.1 Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimen pada penerapan Machine Learning.. Jenis pembelajaran yang digunakan adalah *supervised learning*, karena data yang diolah telah memiliki label kelas yang merepresentasikan status stunting anak. Pendekatan ini banyak digunakan dalam penelitian klasifikasi di bidang kesehatan, sebab mampu mengenali pola dari data historis untuk kemudian melakukan prediksi terhadap data baru (Mohammed et al., 2022; Putri & Arminarahmah, 2024).

2.2 Dataset dan Variabel Penelitian

Dataset yang digunakan dalam penelitian ini merupakan data antropometri anak yang mencakup atribut umur, jenis kelamin, tinggi badan, dan berat badan, dengan label target berupa empat kategori status stunting, yaitu *Normal*, *Tall*, *Stunted*, dan *Severely Stunted*. Pemanfaatan data antropometri sebagai dasar dalam proses klasifikasi status stunting telah banyak diterapkan dalam berbagai penelitian,

karena dianggap mampu merepresentasikan kondisi pertumbuhan anak secara objektif dan terukur (Andriyani et al., 2023; Hendy et al., 2025).

2.3 Tahapan Preprocessing Data

Tahapan preprocessing dilakukan guna menjamin bahwa data yang dipakai benar-benar sesuai dengan kebutuhan analisis yang baik serta sesuai dengan karakteristik algoritma yang diterapkan dalam pemodelan klasifikasi. Proses ini meliputi beberapa langkah sebagai berikut:

1. Data Cleaning

Pada tahap ini, data dianalisis untuk mengidentifikasi dan menghapus data duplikat serta menangani keberadaan nilai kosong. Proses *data cleaning* bertujuan mengurangi *noise* dan meningkatkan kinerja model klasifikasi, sebagaimana direkomendasikan dalam berbagai penelitian klasifikasi data kesehatan (Rosta et al., 2023).

2. Encoding Data Kategorikal

Variabel jenis kelamin yang bersifat kategorikal dikonversi menjadi representasi angka agar algoritma Machine Learning dapat menggunakannya dalam proses komputasi. Tahap *encoding* ini merupakan praktik umum dalam pemodelan klasifikasi data kesehatan berbasis *machine learning* karena memungkinkan algoritma mengolah atribut non-numerik secara lebih efektif.

3. Scaling Data

Proses *scaling* fitur dilakukan menggunakan *StandardScaler* pada data yang akan digunakan oleh algoritma *K-Nearest Neighbor*, karena algoritma ini berbasis perhitungan jarak dan sensitif terhadap perbedaan skala antar fitur (Bimawan et al., 2024; Syam et al., 2021). Sebaliknya, algoritma *Random Forest* tidak memerlukan *scaling* karena bekerja menggunakan struktur pohon keputusan yang tidak dipengaruhi oleh perbedaan skala fitur.

4. Partisi Data

Dataset kemudian dibagi menjadi data latih dan data uji menggunakan metode *stratified split*, sehingga proporsi setiap kelas tetap terjaga seimbang pada masing-masing subset data. Pendekatan ini digunakan untuk meningkatkan keandalan evaluasi model, khususnya pada permasalahan klasifikasi multikelas dengan distribusi kelas yang perlu dijaga tetap seimbang (Uddin et al., 2022).

2.4 Parameter Model

1. Parameter K-Nearest Neighbor (KNN)

Pada penelitian ini, algoritma K-Nearest Neighbor menerapkan parameter jumlah tetangga terdekat (k) sebagai faktor utama dalam proses pengambilan keputusan klasifikasi. Nilai k ditentukan melalui serangkaian eksperimen awal dengan tujuan memperoleh performa terbaik pada data latih, sehingga model dapat menghasilkan prediksi yang lebih optimal. Metode perhitungan jarak yang diterapkan pada penelitian ini adalah **Euclidean distance**, yang secara luas diaplikasikan dalam klasifikasi data numerik, termasuk data antropometri, karena mampu merepresentasikan kedekatan antar titik data dalam ruang fitur secara intuitif. Pengaturan parameter tersebut sejalan dengan temuan berbagai penelitian terdahulu yang menyimpulkan bahwa pemilihan nilai k yang tepat berkontribusi signifikan terhadap peningkatan akurasi model klasifikasi.

2. Parameter K-Nearest Neighbor (KNN)

Algoritma Random Forest pada penelitian ini menggunakan parameter utama berupa **jumlah pohon** (**$n_estimators$**) dan **random state**. Parameter **$n_estimators$** menentukan jumlah pohon yang dibangun dalam model, sedangkan **random state** digunakan untuk menjaga konsistensi hasil eksperimen. Pemilihan parameter ini bertujuan untuk memperoleh model yang stabil dan mengurangi risiko overfitting, sebagaimana direkomendasikan dalam penelitian klasifikasi data kesehatan menggunakan Random Forest (Bimawan et al., 2024; Syam et al., 2021).

3. Algoritma Klasifikasi

1. K-Nearest Neighbor (KNN)

K-Nearest Neighbor digunakan sebagai model baseline karena kesederhanaannya dan kemampuannya dalam melakukan klasifikasi berbasis jarak. Algoritma ini menentukan kelas pada data baru dengan melihat jarak terdekat terhadap data latih yang memiliki kemiripan paling tinggi. KNN sering digunakan sebagai pembandingan awal dalam penelitian klasifikasi kesehatan dan status gizi anak.

2. Random Forest

Random Forest digunakan sebagai model lanjutan karena memiliki kemampuan untuk menangani pola

data yang bersifat non-linear, mengurangi risiko *overfitting*, serta menghasilkan performa yang lebih stabil pada tugas klasifikasi multikelas. Algoritma ini juga telah banyak diaplikasikan dalam klasifikasi data kesehatan dan secara umum dilaporkan mampu memberikan kinerja yang lebih unggul dibandingkan sejumlah algoritma konvensional lainnya (Bimawan et al., 2024; Fitri et al., n.d.; Jannah et al., 2024; Vlachas et al., 2022).

2.6 Evaluasi Model

Penilaian kinerja model dalam penelitian ini dilakukan dengan menggunakan sejumlah metrik evaluasi, yaitu **akurasi**, **precision**, **recall**, **F1-score**, **confusion matrix**, serta **ROC curve**. Penggunaan **F1-score** menjadi krusial pada kasus klasifikasi multikelas karena metrik ini menilai keseimbangan antara *precision* dan *recall* secara simultan, sehingga memberikan gambaran yang lebih representatif terhadap kinerja model pada distribusi kelas yang beragam. **Confusion matrix** dimanfaatkan untuk mengidentifikasi dan menganalisis pola kesalahan klasifikasi pada setiap kelas secara lebih rinci. Sementara itu, **kurva ROC** dimanfaatkan untuk mengevaluasi sejauh mana model mampu membedakan kelas secara keseluruhan dengan mengamati kompromi antara **true positive rate** dan **false positive rate** (Uddin et al., 2022).

2.7 Implementasi Sistem

Model terbaik yang diperoleh dari proses evaluasi kemudian diimplementasikan ke dalam aplikasi web berbasis **Streamlit**. Aplikasi ini memungkinkan pengguna untuk memasukkan data antropometri anak secara interaktif dan langsung memperoleh hasil klasifikasi status stunting secara real-time. Implementasi model ke dalam sistem aplikasi bertujuan untuk mendemonstrasikan penerapan *Machine Learning* secara menyeluruh (*end-to-end*) serta meningkatkan nilai praktis dari penelitian dalam mendukung pengambilan keputusan di bidang kesehatan anak (Fadilla et al., 2024; Mohammed et al., 2022).

HASIL DAN PEMBAHASAN

3.1 Perbandingan Performa Model

Penelitian ini mengevaluasi performa dua algoritma klasifikasi yaitu, **K-Nearest Neighbor (KNN)** dan **Random Forest**, untuk mengklasifikasikan status stunting anak berdasarkan data antropometri. Evaluasi dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score.

| Model | Accuracy | Precision | Recall | F1-Score |
|-------|----------|-----------|--------|----------|
| KNN | 96,67% | 96,67% | 96,67% | 96,67% |
| RF | 100% | 100% | 100% | 100% |

Tabel 1. Perbandingan Performa Model KNN dan Random Forest

Berdasarkan Tabel 1, algoritma **Random Forest** menghasilkan performa yang lebih baik dibandingkan **KNN** pada seluruh metrik evaluasi. KNN masih menghasilkan beberapa kesalahan klasifikasi, sedangkan Random Forest mampu mengklasifikasikan seluruh data uji dengan benar. Temuan ini konsisten dengan studi sebelumnya yang mengemukakan bahwa Random Forest memiliki performa yang stabil dan unggul pada data kesehatan dengan karakteristik non-linear dan multikelas (Bimawan et al., 2024; Syam et al., 2021; Vlachas et al., 2022).

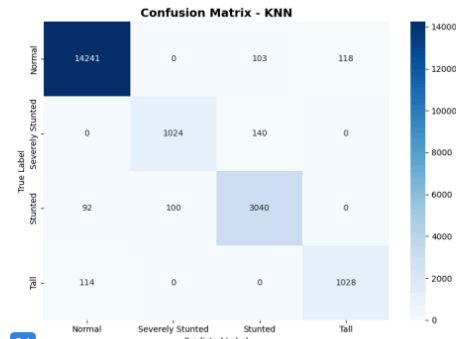
3.2 Analisis Confusion Matrix

Confusion matrix dimanfaatkan untuk menilai kinerja model dalam membedakan dan mengklasifikasikan setiap kelas status stunting, yaitu *Normal*, *Tall*, *Stunted*, dan *Severely Stunted*.

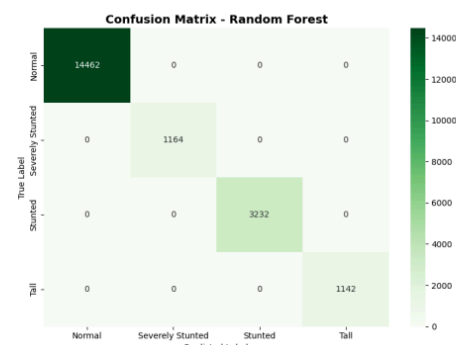
- KNN** masih menunjukkan kesalahan klasifikasi pada beberapa kelas, terutama pada kelas yang memiliki karakteristik antropometri yang saling berdekatan.
- Random Forest** mampu mengklasifikasikan seluruh kelas dengan benar tanpa kesalahan, yang ditunjukkan oleh nilai diagonal penuh pada confusion matrix.

Hasil ini mengindikasikan bahwa Random Forest lebih efektif dalam menangkap hubungan kompleks antar fitur dibandingkan KNN yang hanya mengandalkan perhitungan jarak. Temuan ini konsisten dengan penelitian sebelumnya yang menyebutkan bahwa algoritma berbasis ensemble lebih unggul dalam menangani data medis dan

kesehatan yang kompleks (Uddin et al., 2022).



Gambar 2. Confusion Matrix KNN



Gambar 3. Confusion Matrix Random Forest

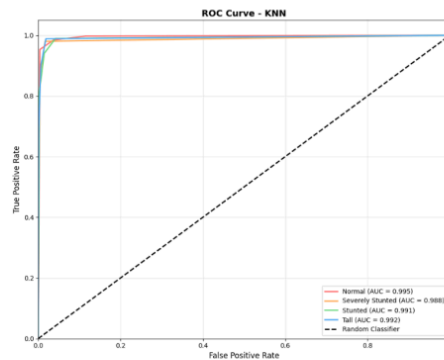
3.3 Analisis ROC Curve

Evaluasi lanjutan dilakukan menggunakan **ROC Curve multiclass** dengan pendekatan *one-vs-rest*. Hasil visualisasi menunjukkan bahwa:

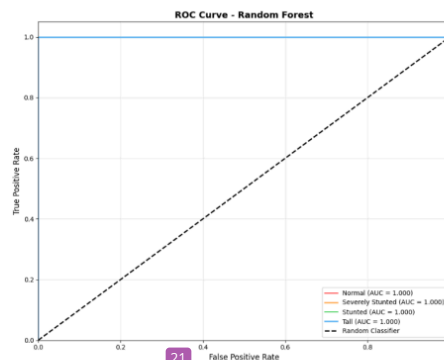
- Kurva ROC **Random Forest** berada sangat dekat dengan titik (0,1) untuk seluruh kelas, yang menunjukkan kemampuan klasifikasi yang sangat baik.
- Nilai **Area Under Curve (AUC) Random Forest** mencapai 1,00, sedangkan KNN memiliki nilai AUC yang lebih rendah.

Temuan ini mengonfirmasi bahwa **Random Forest** memiliki kemampuan diskriminasi antar kelas

yang jauh lebih unggul dibandingkan algoritma KNN. Hasil tersebut konsisten dengan berbagai penelitian internasional yang melaporkan performa superior *Random Forest* dalam tugas klasifikasi penyakit serta data kesehatan anak (Mohammed et al., 2022; Uddin et al., 2022).



Gambar 4. ROC Curve KNN



Gambar 5. ROC Curve Random Forest

3.4 Analisis Feature Importance

Analisis *feature importance* dilakukan dengan memanfaatkan algoritma *Random Forest* untuk mengidentifikasi variabel yang memiliki pengaruh terbesar dalam klasifikasi status stunting pada anak. Hasil analisis menunjukkan urutan pengaruh fitur sebagai berikut:

1. **Tinggi badan** ($\approx 56,3\%$)
2. **Umur** ($\approx 33,5\%$)
3. **Berat badan**

4. Jenis kelamin

Atribut tinggi badan dan umur terbukti menjadi faktor paling dominan dalam penentuan status stunting, yang selaras dengan prinsip antropometri di mana indikator tinggi badan terhadap umur menjadi parameter utama untuk mendeteksi kondisi stunting. Temuan ini konsisten dengan sejumlah penelitian yang menegaskan bahwa variabel pertumbuhan linier memberikan kontribusi terbesar dalam identifikasi stunting pada anak (Andriyani et al., 2023; Hendy et al., 2025).

Dominasi kedua fitur tersebut juga menjelaskan mengapa *Random Forest* mampu mencapai performa yang sangat tinggi, sebab algoritma ini efektif dalam menangkap dan memanfaatkan hubungan non-linear antar fitur-fitur krusial. Hasil analisis ini mendukung temuan penelitian sebelumnya yang menyoroti keunggulan *Random Forest* dalam menentukan *feature importance* pada dataset kesehatan (Syam et al., 2021).

3.5 Pembahasan Umum

Secara umum, hasil penelitian menunjukkan bahwa *Random Forest* memberikan performa yang lebih baik dibandingkan KNN dalam proses klasifikasi status stunting pada anak. Keunggulan tersebut terlihat jelas dari nilai akurasi, *F1-score*, *confusion matrix*, kurva ROC, serta analisis *feature importance* yang dilakukan. Penerapan model terbaik ke dalam aplikasi web berbasis *Streamlit* menunjukkan bahwa penelitian ini tidak hanya bernilai secara teoritis, tetapi juga dapat dimanfaatkan secara praktis sebagai sistem pendukung keputusan dalam proses klasifikasi status stunting anak (Bimawan et al., 2024; Fadilla et al., 2024).

KESIMPULAN DAN SARAN

4.1 Kesimpulan

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa algoritma *Random Forest* memiliki keunggulan yang signifikan dibandingkan KNN dalam mengklasifikasikan status stunting anak. Keunggulan tersebut terlihat jelas dari nilai akurasi, *F1-score*, *confusion matrix*, kurva ROC, serta analisis *feature importance* yang dilakukan. Penerapan model terbaik ke dalam aplikasi web berbasis *Streamlit* menunjukkan bahwa penelitian ini tidak hanya bermanfaat secara teoritis, tetapi juga dapat diterapkan secara praktis sebagai sistem pendukung keputusan dalam proses klasifikasi status stunting anak.

4.2 Saran

Berdasarkan temuan penelitian yang telah diperoleh, beberapa saran dapat diberikan untuk pengembangan penelitian selanjutnya, yaitu:

1. Penelitian berikutnya sebaiknya memanfaatkan dataset yang lebih besar dan berasal dari berbagai wilayah geografis guna meningkatkan kemampuan generalisasi model yang dihasilkan.
2. Penggunaan algoritma alternatif seperti XGBoost, Support Vector Machine, atau pendekatan Deep Learning dapat dipertimbangkan untuk mengidentifikasi potensi peningkatan performa dibandingkan dengan Random Forest.
3. Studi lanjutan disarankan untuk mengintegrasikan variabel tambahan seperti faktor lingkungan, kondisi sosial-ekonomi, serta riwayat kesehatan guna memperkaya analisis klasifikasi status stunting.
4. Pengembangan sistem aplikasi dapat ditingkatkan dengan menambahkan fitur visualisasi hasil analisis dan fungsi penyimpanan data untuk mendukung pemantauan status gizi anak secara berkelanjutan.

DAFTAR PUSTAKA

- Alenany, E., & El-baz, M. A. (2017). *Modelling a Hospital as a Queueing Network: Analysis for Improving Performance*. 11(5), 1181–1187.
- Andriyani, Y., Lydia, M. S., Efendi, S., Technology, I., & Utara, U. S. (2023). *Optimization of Support Vector Machine Algorithm Using Stunting Data Classification*. 11(1), 164–174.
- Aprilia, Y. N., Sani, D. A., Anggadimas, N. M., Studi, P., Informatika, T., Informasi, F. T., & Pasuruan, U. M. (n.d.). *Klasifikasi Status Penderita Gizi Stunting Pada Balita Menggunakan Metode Random Forest (Studi Kasus di Kelurahan Petamanan Kota Pasuruan)*. 143–154.
- Atik, M., Ekowati, S., Hidayat, N., & Karim, A. (2025). *Predictive Modeling for Underweight Detection in Toddlers Using Support Vector Machine , K-Nearest Neighbors , and Decision Tree C4 . 5 Algorithms*. 6(6), 5570–5589.
- Bahadori, M., Mohammadnejhad, S. M., & Ravangard, R. (2014). *Using Queueing Theory and Simulation Model to Optimize Hospital Pharmacy Performance*. 16(3).
<https://doi.org/10.5812/ircmj.16807>
- Berlianty, I., Astanti, Y. D., Soejanto, I., Industri, J. T., Pembangunan, U., & Veteran, N. (2019). *APPLICATION OF DISCRETE-EVENT SIMULATION IN HEALTH CARE : A PRELIMINARY STUDIES. V*, 501–506. <https://doi.org/10.28989/senatik.v5i0.380>
- Bimawan, Z. I., Astuti, T., & Arsi, P. (2024). *COMPARISON OF RANDOM FOREST , K-NEAREST NEIGHBOR , DECISION TREE , AND XGBOOST ALGORITHMS FOR DETECTING STUNTING IN TODDLERS KOMPARIASI ALGORITMA RANDOM FOREST , K-NEAREST NEIGHBOR ,. 5(6)*, 1599–1607.
- Fadilla, M. A., Sholahuddin, M. F., & Sutabri, T. (2024). *Pengembangan Sistem Klasifikasi Diagnosa Medis Menggunakan Progressive Web Application Terintegrasi Machine Learning*. 5(12), 5488–5503.
- Fauzan, R., & Rosita, A. (2025). *Comparison of KNN and Naïve Bayes Classification Algorithms for Predicting Stunting in Toddlers in Banjaran District*. 9(5), 271–2717.
- Fitri, R. R., Apriandari, W., Studi, P., Informatika, T., Sukabumi, U. M., Sukabumi, K., & Barat, J. (n.d.). *PENGUNAAN RANDOM FOREST DALAM SISTEM*. 13(3).
- Hendy, A., Ibrahim, R. K., Mohammed, S., Abdelaliem, F., & Zaher, A. (2025). *Supervised machine learning for classification and prediction of stunting among under-five Egyptian children*.
- Jannah, M., Hasan, M. A., & Fajar, M. Al. (2024). *PERBANDINGAN METODE NAÏVE BAYES DAN K-NEAREST NEIGHBOR DALAM MENGLASIFIKASI STATUS PERTUMBUHAN ANAK STUNTING (STUDI KASUS : POSYANDU CEMARA)*. 14(1), 250–255.
- Lestari, W. S., Saragih, Y. M., Technology, I., & Mikroskil, U. (2024). *MULTICLASS CLASSIFICATION FOR STUNTING PREDICTION USING*. 10(2), 386–393.
<https://doi.org/10.33480/jitk.v10i2.5636>
- Marleny, F. D., & Zulfadhilah, M. (2023). *Prediction of linear model on stunting prevalence with machine learning approach*. 12(1), 483–492.
<https://doi.org/10.11591/eei.v12i1.4028>
- Mohammed, N. I., Jarde, A., Mackenzie, G., Alessandro, U. D., & Jeffries, D. (2022). *Deploying Machine Learning Models Using Progressive Web Applications : Implementation Using a Neural Network Prediction Model for Pneumonia Related Child Mortality in The Gambia*. 9(February), 1–6.
<https://doi.org/10.3389/fpubh.2021.772620>
- Ndagijimana, S., Kabano, I. H., Masabo, E., & Ntaganda, J. M. (2023). *Prediction of Stunting Among Under-5 Children in Rwanda Using Machine Learning Techniques*. 41–49.

- Putri, I. P., & Arminarahmah, N. (2024). *Comparative Analysis of Machine Learning Algorithms for Predicting Child Stunting Analisis Perbandingan Algoritma Machine Learning untuk Prediksi Stunting pada Anak*. 4(January), 257–265.
- Rosta, E., Sebayang, B., Chrisnanto, Y. H., Jenderal, U., Yani, A., Terusan, J., Sudirman, J., Selatan, C., & Barat, J. (2023). *Klasifikasi Data Kesehatan Mental di Industri Teknologi Menggunakan Algoritma Random Forest*. 1(3), 237–253.
- Shella, S., Wara, M., Nasrudin, M., Adziima, A. F., & Pratama, A. R. (2025). *Optimasi Sistem Antrian Pada Medical Center Its Dengan Simulasi Discrete Event Dan Response Surface Methodology*. 99(99), 1–8.
- Status, K., Anak, G., & Kecamatan, D. I. (2023). *Perbandingan metode lvq dan backpropagation untuk klasifikasi status gizi anak di kecamatan sangkub*. 12, 314–321.
<https://doi.org/10.14710/j.gauss.12.3.314-321>
- Syam, R., Djatmiko, W., & Identification, H. D. (2021). *Prediction of Heart Diseases using Random Forest Prediction of Heart Diseases using Random Forest*.
<https://doi.org/10.1088/1742-6596/1817/1/012009>
- Uddin, S., Haque, I., Lu, H., Moni, M. A., & Gide, E. (2022). Comparative performance analysis of K - nearest neighbour (KNN) algorithm and its different variants for disease prediction. *Scientific Reports*, 1–11. <https://doi.org/10.1038/s41598-022-10358-x>
- Vlachas, C., Damianos, L., Gousetis, N., Mouratidis, I., Kelepouris, D., Kollias, K., Asimopoulos, N., & Fragulis, G. F. (2022). *Random forest classification algorithm for medical industry data*. 03008, 2–7.
- Wahidin, A. J., & Andika, T. H. (2024). *Deteksi Dini Stunting Pada Anak Berdasarkan Indikator Antropometri dengan Menggunakan Algoritma Machine Learning*. 378–387.
<https://doi.org/10.33364/algoritma/v.21-2.2122>
- Yunus, M., Biddinika, M. K., Fadlil, A., & Info, A. (2025). *Classification of Stunting in Children Using the C4 . 5 Algorithm*. 8(1), 99–106.
<https://doi.org/10.15575/join.v8i1.1062>

ORIGINALITY REPORT

| | | | |
|------------------|------------------|--------------|----------------|
| 20% | 14% | 10% | 4% |
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|----|--|----|
| 1 | ejournal.unibba.ac.id Internet Source | 4% |
| 2 | Submitted to FAKULTAS ILMU KOMPUTER Student Paper | 1% |
| 3 | Aryadeep Roychoudhury. "Epigenetic Regulations in Plants - Mechanisms, Plant Development and Stress Responses", CRC Press, 2026 Publication | 1% |
| 4 | jurnal.itg.ac.id Internet Source | 1% |
| 5 | journal.untar.ac.id Internet Source | 1% |
| 6 | www.ijaeb.org Internet Source | 1% |
| 7 | www.stmik-budidarma.ac.id Internet Source | 1% |
| 8 | Herlambang Irawan. "Credit Card Fraud Detection with Logistic Regression, Random Forest, and Gradient Boosting", Jurnal Informatika Terpadu, 2025 Publication | 1% |
| 9 | ijirt.org Internet Source | 1% |
| 10 | A. Reza Baehaqa Jamroni Jamroni, Wahyu Hadikristanto, Muhamad Fatchan. "Analisis Faktor dan Prediksi Atrisi untuk Optimalisasi | 1% |

Retensi Karyawan Menggunakan Machine Learning", bit-Tech, 2025

Publication

| | | |
|----|---|------|
| 11 | Submitted to Universitas Islam Riau Student Paper | 1 % |
| 12 | repository.ipb.ac.id Internet Source | 1 % |
| 13 | Submitted to Universitas Malikussaleh Student Paper | <1 % |
| 14 | Submitted to Universitas Muhammadiyah Purwokerto Student Paper | <1 % |
| 15 | Submitted to Universitas Brawijaya Student Paper | <1 % |
| 16 | repository.radenfatah.ac.id Internet Source | <1 % |
| 17 | Submitted to Sriwijaya University Student Paper | <1 % |
| 18 | Submitted to Universitas Pertahanan RI Student Paper | <1 % |
| 19 | journal.irpi.or.id Internet Source | <1 % |
| 20 | Karunia Viandra Yoanisa Yoanisa, Dhiny Easter Yanti, Lolita Sary, Christin Angelina F. "PENGARUH MEDIA EDUKASI VIDEO ANTISIPASI STUNTING (VAS) TERHADAP PENINGKATAN PENGETAHUAN IBU DALAM PENCEGAHAN STUNTING DI WILAYAH KERJA PUSKESMAS PAGELARAN TAHUN 2025", PREPOTIF : JURNAL KESEHATAN MASYARAKAT, 2025 Publication | <1 % |

21 Muhammad Ilham Manzis Nur Jamal. "IMPLEMENTASI MACHINE LEARNING BERBASIS FITUR WARNA RGB DAN HSV UNTUK KLASIFIKASI KUALITAS AIR", Jurnal Informatika dan Teknik Elektro Terapan, 2026
Publication

22 link.springer.com
Internet Source

23 Abhishek Agarwal, Amit Rana, Karan Gupta, Neeta Verma. "A Comparative Study and enhancement of classification techniques using Principal Component Analysis for credit card dataset", 2020 International Conference on Intelligent Engineering and Management (ICIEM), 2020
Publication

24 Adji Surya Pamungkas, Nuri Cahyono. "Analisis Sentimen Review ChatGPT di Play Store menggunakan Support Vector Machine dan K-Nearest Neighbor", Edumatic: Jurnal Pendidikan Informatika, 2024
Publication

25 Erviana Erviana, Widiani Widiani, Muhammad Taufik Page Taufik, Risna Damayanti. "Hubungan Pola Asuh Orang Tua Dalam Pemberian Makan Dengan Kejadian Stunting Pada Anak Usia 24-59 Bulan", Citra Delima Scientific journal of Citra Internasional Institute, 2024
Publication

26 Rafael Nicholas Tanaja, Johnny, Musyaffa Ayman Rafif, Alexander Agung Santoso Gunawan. "Fake News Detection using Machine Learning: Integrating FakeBERT Classification, Style Analysis, and Credibility

-
- | | | |
|-----------|---|----------------|
| 27 | Sekar Aulia Sista, Dadan Tonny Kusnandar, Neva Satyahadewi. "A ORDINAL LOGISTIC REGRESSION BAGGING FOR MODELING AND CLASSIFICATION OF THE NUTRITIONAL STATUS OF TODDLERS IN SOUTHEAST PONTIANAK SUB-DISTRICT", VARIANCE: Journal of Statistics and Its Applications, 2024 <small>Publication</small> | <1 % |
|-----------|---|----------------|
-
- | | | |
|-----------|--|----------------|
| 28 | Yudhi Fajar Saputra, Mahmoud Ahmad Al-Khasawneh, Milkhatun Milkhatun, Ni Wayan Wiwin Asthiningsih, Sitti Rahmah. "ANALISIS KOMPARATIF KLASIFIKASI MACHINE LEARNING UNTUK MEMPREDIKSI STUNTING PADA ANAK USIA DI BAWAH LIMA TAHUN", Syntax : Journal of Software Engineering, Computer Science and Information Technology, 2025 <small>Publication</small> | <1 % |
|-----------|--|----------------|
-
- | | | |
|-----------|--|----------------|
| 29 | ejurnal.methodist.ac.id <small>Internet Source</small> | <1 % |
|-----------|--|----------------|
-
- | | | |
|-----------|--|----------------|
| 30 | eprints.uad.ac.id <small>Internet Source</small> | <1 % |
|-----------|--|----------------|
-
- | | | |
|-----------|--|----------------|
| 31 | hal.science <small>Internet Source</small> | <1 % |
|-----------|--|----------------|
-
- | | | |
|-----------|--|----------------|
| 32 | jpti.journals.id <small>Internet Source</small> | <1 % |
|-----------|--|----------------|
-
- | | | |
|-----------|--|----------------|
| 33 | repository.usd.ac.id <small>Internet Source</small> | <1 % |
|-----------|--|----------------|
-
- | | | |
|-----------|--|----------------|
| 34 | www.mdpi.com <small>Internet Source</small> | <1 % |
|-----------|--|----------------|
-
- | | | |
|-----------|--|--|
| 35 | lib.ibs.ac.id | |
|-----------|--|--|

| | | | |
|----------------------|-----|-----------------|-----|
| Exclude quotes | Off | Exclude matches | Off |
| Exclude bibliography | On | | |