

KLASIFIKASI STATUS GIZI BALITA MENGGUNAKAN METODE DECISION TREE DAN K-FOLD CROSS VALIDATION

Sabrina Aisyah Putri¹, Rachmat Adi Purnama^{2*}

^{1,2} Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika
Jl.Kramat Raya No.98, Koya Jakarta Pusat, Daerah Khusus Ibukota Jakarta, Indonesia

e-mail: [1sabrina0806@gmail.com](mailto:sabrina0806@gmail.com), [2rachmat.rap@bsi.ac.id](mailto:rachmat.rap@bsi.ac.id)

Submission: dd-mm-yyyy	Revision: dd-mm-yyyy	Acceptance: dd-mm-yyyy	Available Online: dd-mm-yyyy
---------------------------	-------------------------	---------------------------	---------------------------------

Abstrak

Status gizi balita merupakan indikator penting bagi kesehatan masyarakat yang berpengaruh pada kualitas sumber daya manusia. Namun, pemantauan status gizi balita secara manual memiliki keterbatasan akurasi dan efisiensi. Dalam upaya mengatasi tantangan ini, metode berbasis machine learning seperti algoritma decision tree menawarkan solusi untuk meningkatkan kecepatan dan akurasi penilaian. Penelitian ini bertujuan untuk mengembangkan model klasifikasi status gizi balita menggunakan algoritma decision tree yang sederhana, mudah dipahami, dan diinterpretasikan. Keakuratan model diuji menggunakan Teknik validasi K-fold Cross Validation, yang memastikan evaluasi performa model dilakukan secara menyeluruh dengan membagi data menjadi beberapa subset untuk pelatihan dan pengujian. Hasil penelitian menunjukkan bahwa algoritma decision tree dengan K-fold Cross Validation mampu menghasilkan model klasifikasi yang akurat dan dapat diandalkan. Pendekatan ini diharapkan menjadi alternatif dalam pemantauan status gizi balita untuk mendukung peningkatan Kesehatan masyarakat Indonesia.

Kata Kunci : Status gizi balita, Decision Tree, K-fold Cross Validation, Klasifikasi, Machine Learning

Abstract

The nutritional status of children under five is an important indicator of public health which influences the quality of human resources. However, monitoring the nutritional status of toddlers manually has limited accuracy and efficiency. In an effort to overcome this challenge, machine learning-based methods such as decision tree algorithms offer solutions to increase the speed and accuracy of assessments. This research aims to develop a classification model for the nutritional status of toddlers using a decision tree algorithm that is simple, easy to understand and interpret. The accuracy of the model was tested using the K-fold Cross Validation validation technique, which ensures that the model performance evaluation is carried out thoroughly by dividing the data into several subsets for training and testing. The research results show that the decision tree algorithm with K-fold Cross Validation is able to produce an accurate and reliable classification model. This approach is expected to be an alternative in monitoring the nutritional status of toddlers to support improving the health of Indonesian society.

Keywords : Status gizi balita, Decision Tree, K-fold Cross Validation, Klasifikasi, Machine Learning

1. Pendahuluan

Status Gizi balita merupakan factor krusial yang berpengaruh terhadap tumbuh kembang anak. Menurut (Handayani & Charis Fauzan, 2024), penilaian status gizi balita dapat dilakukan melalui metode antropometri yang mencakup berat badan menurut usia (BB/U),

tinggi badan menurut usia (TB/U), dan indeks massa tubuh (IMT/U). Selain itu, penelitian dari (Latifah, 2018) menyebutkan bahwa klasifikasi status gizi sangat penting dalam mengidentifikasi anak-anak yang berisiko mengalami masalah Kesehatan di masa depan.

Konsep dasar klasifikasi sering digunakan dalam berbagai bidang termasuk ilmu komputer dan analisis data. Algoritma Decision Tree adalah salah satu metode yang efektif dalam mengklasifikasi karena mudah dipahami dan dapat menghasilkan aturan klasifikasi karena mudah dipahami dan dapat menghasilkan aturan klasifikasi yang jelas (Bahri & Lubis, 2020). Penelitian oleh (Romzi & Kurniawan, 2020) menunjukkan bahwa Decision Tree memiliki keunggulan dalam klasifikasi data Kesehatan dibandingkan dengan metode lainnya karena strukturnya yang hierarkis dan mudah diinterpretasikan.

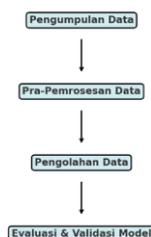
Salah satu Teknik yang digunakan untuk meningkatkan akurasi model adalah K-Fold Cross Validation. Teknik ini memastikan bahwa model diuji dengan data yang berbeda setiap iterasi sehingga meningkatkan keakuratan klasifikasi (Adhi Putrayana et al., n.d.). K-Fold Cross Validation dapat membantu mencegah overfitting dalam model klasifikasi status gizi balita.

Selain itu, perangkat lunak seperti Python dan Google COLab sering digunakan dalam implementasi algoritma machine learning, termasuk Decision Tree (Gelar Guntara, 2023). Python menyediakan Pustaka seperti Scikit-Learn yang mendukung pengolahan dan validasi data. Dengan meningkatnya penggunaan dataset dari platform seperti Kaggle, metode machine learning telah banyak diterapkan dalam analisis Kesehatan masyarakat (Azis Rahmat et al., 2023).

Dengan memanfaatkan algoritma Decision Tree dan K-Fold Cross Validation, penelitian ini bertujuan untuk mengembangkan model klasifikasi status gizi balita yang akurat dan dapat diandalkan. Model ini diharapkan dapat membantu tenaga Kesehatan dalam memantau status gizi balita dan memberikan rekomendasi intervensi yang tepat.

2. Metode Penelitian

Metode penelitian ini menggunakan serangkaian proses dan Langkah-langkah untuk mengklasifikasi status gizi balita. Metode ini didukung menggunakan metode Decision Tree dan K-Fold Cross Validation untuk mencapai hasil yang paling tepat dan efisien.



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Dataset yang digunakan dalam penelitian ini diperoleh dari Kaggle, yang mencakup atribut seperti usia, jenis kelamin, tinggi badan, dan berat badan balita. Total Dataset jumlah 120.999 entri.

Tabel 1 Dataset Gizi Balita

Umur (bulan)	Jenis Kelamin	Tinggi Badan (cm)	Status Gizi
0	Laki-laki	44.5919732943438	<i>stunted</i>
0	Laki-laki	56.7052033668847	tinggi
0	Laki-laki	46.8633575967919	normal
0	Laki-laki	47.5080256315438	normal
0	Laki-laki	42.7434938911793	<i>severely stunted</i>
0	Laki-laki	44.2577186391463	<i>stunted</i>
0	Laki-laki	59.5725227366786	tinggi
0	Laki-laki	42.7017961514634	<i>severely stunted</i>
0	Laki-laki	45.2517789938352	<i>stunted</i>
0	Laki-laki	46.8633575967919	<i>stunted</i>

2.2 Pra Pemrosesan Data

Sebelum data dianalisis menggunakan metode decision tree, langkah-langkah pra pemrosesan data dilakukan untuk memastikan kualitas dataset dan mempersiapkannya akagar siap digunakan dalam analisis. Berikut langkah-langkah pra pemrosesan yang dilakukan:

- Memeriksa tipe data pada setiap kolom untuk memastikan struktur data sudah benar dan tidak terdapat kesalahan dalam jenis data.
- Mengevaluasi dan menghitung keberadaan nilai yang hilang disetiap kolom dataset untuk memastikan bahwa tidak ada data yang hilang yang dapat memengaruhi hasil analisis.
- Menghapus kolom yang tidak relevan atau tidak diperlukan dalam proses analisis.

2.3 Pengolahan data

Dalam penelitian ini pengolahan data menggunakan Google Colaboratory (Google Colab) untuk mengelola dataset terkait dan melakukan analisis prediktif. Tujuannya adalah untuk memperoleh wawasan yang berharga serta mendukung pengambilan keputusan yang lebih efektif dalam upaya mengatasi permasalahan gizi pada anak balita

2.4 Evaluasi dan Validasi Model

Beberapa metrik yang umum digunakan dalam mengevaluasi performa model meliputi

akurasi, presisi, recall, dan F1-score. Akurasi merupakan rasio antara jumlah data yang diklasifikasikan dengan benar terhadap keseluruhan data yang digunakan.

Akurasi :

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

Presisi : menggambarkan proporsi data yang diklasifikasikan sebagai positif dengan benar dibandingkan dengan total data yang diklasifikasikan sebagai positif.

Rumus Presisi :

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

Recall : Merupakan rasio antara jumlah data yang benar di klasifikasikan sebagai positif dibandingkan dengan total data yang seharusnya positif.

Rumus Recall :

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

F1-score : Menggabungkan presisi dan recall untuk memberikan ukuran kinerja model yang lebih seimbang. Semakin tinggi skor F1, semakin baik kinerja model klasifikasi.

Rumus F1-score :

$$= 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \times 100$$

Menurut (Arther Sandag, 2020), pengujian dilakukan untuk menentukan apakah suatu objek diklasifikasikan dengan benar atau salah.

Tabel 2. Confusion Matrix

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

1. TP (True Positive): Jumlah data yang benar diklasifikasikan sebagai positif.
2. TN (True Negative): Jumlah data yang benar diklasifikasikan sebagai negatif.
3. FN (False Negative): Jumlah data yang salah diklasifikasikan sebagai negatif.
4. FP (False Positive): Jumlah data yang salah diklasifikasikan sebagai positif.

Confusion Matrix merupakan metode yang digunakan untuk mengevaluasi kinerja model klasifikasi dengan membandingkan hasil

prediksi model terhadap label yang sebenarnya. Confusion Matrix memberikan gambaran yang jelas tentang performa model dalam mengkategorikan data secara benar maupun salah.

3. Hasil dan Pembahasan

Model Decision Tree mencapai akurasi 99,92%, menunjukkan efektivitasnya dalam mengklasifikasikan status gizi balita.

```

Accuracy: 0.9992975206611571
====CONFUSION MATRIX====
[[13530  0  2  4]
 [  0 3979  3  0]
 [  2  5 2675  0]
 [  1  0  0 3999]]
precision  recall  f1-score  support

normal      1.00    1.00    1.00   13536
severely stunted  1.00    1.00    1.00    3982
stunted     1.00    1.00    1.00    2682
tinggi      1.00    1.00    1.00    4000

accuracy    1.00    1.00    1.00   24200
macro avg   1.00    1.00    1.00   24200
weighted avg 1.00    1.00    1.00   24200
    
```

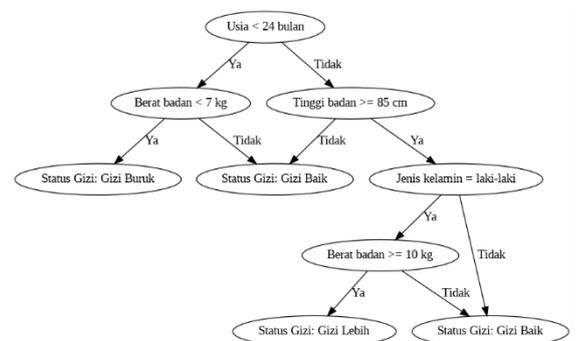
Gambar 2 hasil akurasi

Dari total 120,999 data dengan kata lain hanya sekitar 0,08% data uji yang di prediksi salah, yang merupakan jumlah kesalahannya yang sangat kecil menjadi bukti bahwa algoritma Decision Tree mampu mempelajari pola data sangat baik dan menghasilkan model yang sangat kuat untuk melakukan klasifikasi.

3.1 Hasil Pemodelan

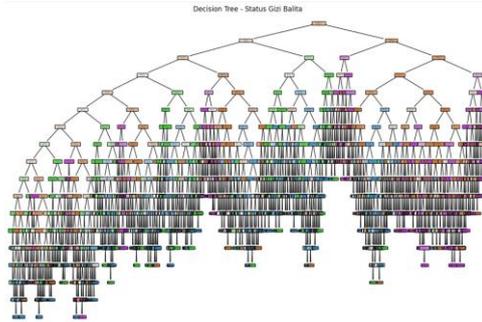
Setelah dilakukan pemodelan menggunakan algoritma Decision Tree, struktur pohon yang dihasilkan menunjukkan hubungan antara atribut dengan status gizi balita

- Jika usia < 24 bulan dan berat badan < 7 kg, maka status gizi = Gizi Buruk.
- Jika usia >= 24 bulan dan tinggi badan >= 85 cm, maka status gizi = Gizi Baik.
- Jika jenis kelamin = laki-laki dan berat badan >= 10 kg, maka status gizi = Gizi Lebih.



Gambar 3 pohon keputusan.

3.2 Visualisasi Decision Tree



Gambar 4. Visualisasi Pohon Keputusan

Setiap pohon dalam model Decision Tree memecahkan data berdasarkan fitur-fitur tertentu. Pohon keputusan berkerja dengan cara membagi dataset ke dalam cabang-cabang berdasarkan keputusan yang diambil dari atribut-atribut.

3.3 Hasil Pengujian

Dalam pengujian model ini, digunakan metode K-Fold Cross Validation dengan nilai K=10. Metode ini membagi dataset menjadi 10 bagian (fold) yang berukuran kurang lebih sama.

```
[42] print("Cross-validation scores for each fold:")  
print(scores)  
  
Cross-validation scores for each fold:  
[0.99900826 0.99942149 0.99933884 0.99975207 0.99909091 0.99909091  
0.99917355 0.99917355 0.99917355 0.99909083]
```

Gambar 5 hasil pengujian k-fold cross validation

Karena ada 10 skor yang diberikan, masing-masing merepresentasikan nilai evaluasi model pada satu lipatan. Skor yang ditampilkan sangat tinggi, mendekati angka 1 (contoh 0.99900826, 0.99942149, dan seterusnya). Setiap fold melibatkan pelatihan model pada K-1 bagian data dan menguji performannya pada fold yang tersisa.

```
# Tampilkan skor untuk setiap lipatan  
for i, score in enumerate(scores):  
    print(f"Nilai K-Fold Pengujian ke-{i+1}: {score:.2f}")  
  
# Cetak rata-rata akurasi  
print("Rata-rata Akurasi:", np.mean(scores))  
  
Nilai K-Fold Pengujian ke-1: 1.00  
Nilai K-Fold Pengujian ke-2: 1.00  
Nilai K-Fold Pengujian ke-3: 1.00  
Nilai K-Fold Pengujian ke-4: 1.00  
Nilai K-Fold Pengujian ke-5: 1.00  
Nilai K-Fold Pengujian ke-6: 1.00  
Nilai K-Fold Pengujian ke-7: 1.00  
Nilai K-Fold Pengujian ke-8: 1.00  
Nilai K-Fold Pengujian ke-9: 1.00  
Nilai K-Fold Pengujian ke-10: 1.00  
Rata-rata Akurasi: 0.9992313974449087
```

Gambar 6 Hasil rata-rata akurasi

Setelah semua skor ditampilkan, dihitung rata-rata akurasi dari keseluruhan lipatan yaitu, 0.999231. Nilai rata-rata ini menunjukkan bahwa

model memiliki performa yang mendekati sempurna (99,92%).

4. Kesimpulan

Penelitian ini berhasil mengembangkan model klasifikasi status gizi balita berbasis Decision Tree dengan tujuan meningkatkan efisiensi dan akurasi dalam pemantauan status gizi anak-anak. Hasil evaluasi menunjukkan bahwa model memiliki tingkat akurasi yang lumayan tinggi mencapai 99,92% menunjukkan model sangat mampu mengidentifikasi status gizi balita dengan tingkat kesalahan yang rendah. Dengan menggunakan pendekatan berbasis data ini, tenaga Kesehatan dapat memperoleh informasi yang lebih akurat dan real-time mengenai kondisi balita.

Dengan ini juga menunjukkan bahwa algoritma Decision Tree dapat menjadi alat yang efektif dalam mendukung upaya pemantauan gizi masyarakat, terutama dalam mengidentifikasi anak-anak yang beresiko mengalami masalah gizi seperti stunting, wasting, dan obesitas.

Referensi

Adhi Putrayana, R., Ramadhan, R., Saputra,

R. I., & Sahid, R. A. (n.d.). *Literatur*

Review: Klasifikasi Penyakit Menular

Seksual (PMS) Menggunakan Naïve

Bayes dan Metode Machine Learning

Terkait.

<https://journal.mediapublikasi.id/index.php/oktal>

[x.php/oktal](https://journal.mediapublikasi.id/index.php/oktal)

Arther Sandag, G. (2020). *Prediksi Rating*

Aplikasi App Store Menggunakan

Algoritma Random Forest Application

Rating Prediction on App Store using

Random Forest Algorithm. Cogito

- Smart Journal* /, 6(2). <https://doi.org/10.47233/jteksis.v5i1.7>
<https://www.kaggle.com/> [50](#)
- Azis Rahmat, W., Madinah Ladjamuddin, S., & Teruna Awaludin, D. (2023). PERBANDINGAN ALGORITMA DECISION TREE, RANDOM FOREST DAN NAIVE BAYES PADA PREDIKSI PENILAIAN KEPUASAN PENUMPANG MASKAPAI PESAWAT MENGGUNAKAN DATASET KAGGLE. *Jurnal Rekayasa Informasi*, 12(2). www.kaggle.com ,
- Bahri, S., & Lubis, A. (2020). METODE KLASIFIKASI DECISION TREE UNTUK MEMPREDIKSI JUARA ENGLISH PREMIER LEAGUE. 2(1).
- Gelar Guntara, R. (2023). Pemanfaatan Google Colab Untuk Aplikasi Pendeteksian Masker Wajah Menggunakan Algoritma Deep Learning YOLOv7. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 5(1), 55–60.
- Handayani, P., & Charis Fauzan, A. (2024). KLIK: Kajian Ilmiah Informatika dan Komputer Machine Learning Klasifikasi Status Gizi Balita Menggunakan Algoritma Random Forest. *Media Online*), 4(6), 3064–3072.
<https://doi.org/10.30865/klik.v4i6.190>
[9](#)
- Latifah, K. (2018). ANALISIS DAN PENERAPAN ALGORITMA C45 DALAM DATA MINING UNTUK MENUNJANG STRATEGI PROMOSI PRODI INFORMATIKA UPGRIS. *JURNAL TEKNIK INFORMATIKA*, 11(2), 109–120.
<https://doi.org/10.15408/jti.v11i2.670>
[6](#)
- Romzi, M., & Kurniawan, B. (2020). PEMBELAJARAN PEMROGRAMAN PYTHON DENGAN PENDEKATAN LOGIKA ALGORITMA (Issue 2).