

Deteksi *Stunting* Balita Menggunakan Algoritma ID3

^{1*}Yoga Firnanda Wicaksono, ²Ardi Sanjaya, ³Ahmad Bagus Setiawan

^{1,2,3} Teknik Informatika, Universitas Nusantara PGRI Kediri

E-mail: ¹firnanday89@gmail.com, ²dersky@gmail.com, ³ahmadbagus@unpkediri.ac.id

Penulis Korespondens : Yoga Firnanda Wicaksono

Abstrak— *Stunting* merupakan masalah gizi kronis pada balita yang dapat memengaruhi pertumbuhan fisik, perkembangan kognitif, dan kualitas sumber daya manusia di masa depan. Proses deteksi *stunting* di Posyandu Kamboja Desa Kwaron masih dilakukan secara manual sehingga rentan terhadap kesalahan pencatatan dan keterlambatan analisis data. Penelitian ini bertujuan untuk membangun sistem deteksi *stunting* balita otomatis menggunakan algoritma *Iterative Dichotomiser 3* (ID3) berbasis *Decision Tree*. Dataset yang digunakan berjumlah 223 data balita dengan atribut jenis kelamin, umur, tinggi badan, berat badan, lingkar kepala, dan Lingkar Lengan Atas (LiLa). Tahapan penelitian meliputi *preprocessing* data, pembagian data dengan rasio 80:20, proses pelatihan model, serta evaluasi menggunakan *confusion matrix*. Hasil pengujian menunjukkan bahwa model mampu mengklasifikasikan status *stunting* ke dalam kategori normal, pendek, sangat pendek, dan tinggi dengan tingkat akurasi sebesar 73,33%. Sistem yang dibangun diharapkan dapat membantu kader posyandu dalam melakukan deteksi *stunting* secara lebih cepat, terstruktur, dan akurat.

Kata Kunci— *Decision Tree*, ID3, klasifikasi, *machine learning*, *stunting*.

Abstract— *Stunting* is a chronic nutritional problem in toddlers that can affect physical growth, cognitive development, and the quality of human resources in the future. The process of *stunting* detection at the Posyandu Kamboja Kwaron Village is still done manually, making it vulnerable to recording errors and delays in data analysis. This study aims to build an automatic toddler *stunting* detection system using the *Iterative Dichotomiser 3* (ID3) algorithm based on *Decision Trees*. The dataset used was 223 toddlers with attributes of gender, age, height, weight, head circumference, and Mid-Upper Arm Circumference (LILA). The research stages included data *preprocessing*, data division with a ratio of 80:20, model training, and evaluation using a *confusion matrix*. The test results showed that the model was able to classify *stunting* status into normal, short, very short, and tall categories with an accuracy level of 73.33%. The system developed is expected to assist Posyandu cadres in detecting *stunting* more quickly, structured, and accurately..

Keywords— *Decision Tree*, ID3, classification, *machine learning*, *stunting*.

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Stunting merupakan kondisi gangguan pertumbuhan fisik pada anak akibat malnutrisi kronis, yang ditandai dengan ketidaksesuaian antara tinggi badan dan usia anak[1]. Kondisi ini berdampak buruk tidak hanya pada postur tubuh yang tidak optimal, tetapi juga pada perkembangan kognitif, kerentanan terhadap penyakit infeksi, serta penurunan produktivitas di

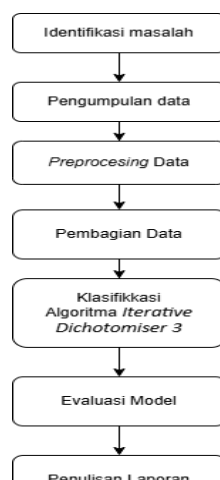
masa depan. Menyadari besarnya ancaman tersebut bagi kualitas sumber daya manusia, Pemerintah Indonesia menjadikan percepatan penurunan angka *stunting* sebagai agenda prioritas nasional yang menuntut keterlibatan aktif dari berbagai sektor pelayanan kesehatan.

Sebagai garda terdepan di tingkat masyarakat desa, posyandu memegang peran krusial dalam deteksi dini *stunting*. Namun, pemantauan *stunting* di Posyandu Kamboja, Desa Kwaron, Kecamatan Papar, Kabupaten Kediri, saat ini masih bertumpu pada mekanisme pencatatan manual menggunakan buku Kesehatan Ibu dan Anak (KIA). Data *antropometri* seperti jenis kelamin, umur, tinggi badan, berat badan, lingkar kepala, dan lila dicatat secara konvensional oleh kader posyandu. Keterbatasan sistem manual ini menyebabkan proses identifikasi berjalan lambat, rentan terhadap kesalahan manusia (*human error*), dan menyulitkan kader dalam melakukan analisis cepat terhadap tren pertumbuhan anak karena harus menelusuri grafik data historis secara manual.

Untuk mengatasi kendala tersebut, integrasi teknologi informasi berbasis *Machine Learning* menawarkan solusi yang efektif, objektif, dan akurat. Beberapa penelitian terdahulu telah membuktikan keandalan algoritma klasifikasi dalam mendeteksi status gizi balita. [2] membuktikan bahwa algoritma *Decision Tree* yang didukung optimasi *preprocessing* dan *GridSearch tuning* mampu memprediksi status gizi balita dengan tingkat akurasi sebesar 89,75%. Sejalan dengan hal tersebut, studi oleh [3] menegaskan bahwa penerapan *Decision Tree* yang dipadukan dengan teknik *preprocessing* komprehensif menghasilkan model prediksi yang sangat stabil, dengan capaian akurasi mencapai 95,73% berdasarkan evaluasi *Confusion Matrix*.

Berpijak pada tingginya tingkat keandalan dan kemampuan interpretasi aturan dari metode tersebut, penelitian ini bertujuan untuk mengimplementasikan metode *Decision Tree* pada sistem klasifikasi *stunting* di Posyandu Kamboja. Model ini dilatih menggunakan parameter pengukuran balita setempat yang telah divalidasi oleh tenaga kesehatan (bidan), guna mengklasifikasi *stunting* ke dalam empat kategori yaitu: normal, pendek, sangat pendek, dan tinggi. Transisi dari pendataan konvensional menuju sistem klasifikasi diharapkan dapat membantu kader posyandu serta pemerintah desa dalam mengambil keputusan kesehatan yang cepat dan presisi, sekaligus mendukung upaya penekanan prevalensi *stunting* secara nasional.

II. METODE



Gambar 1. Tahapan penelitian

Diagram alir ini memperlihatkan langkah-langkah penelitian yang akan dilakukan. Tahapan yang dilakukan yaitu untuk mengklasifikasi *stunting* pada balita dengan menerapkan algoritma *Iterative Dichotomiser 3 (ID3)* terlihat pada Gambar 1.

A. Pengumpulan Data

Pada tahap pengumpulan data, langkah awal yang dilakukan adalah menghimpun data sekunder yang diperoleh langsung dari bidan. Tahap ini dilanjutkan dengan mengimpor file dataset berisi 223 baris data historis pertumbuhan balita. Atribut data yang digunakan dalam penelitian ini meliputi jenis kelamin, umur (bulan), tinggi badan (cm), berat badan (kg), lingkar kepala (cm), lingkar lengan atas atau LiLa (cm), serta klasifikasi status *stunting* pada balita.

B. Preprocessing Data

Pada tahap *preprocessing* data, dilakukan proses pengolahan data awal agar dataset yang digunakan menjadi lebih bersih, terstruktur, dan siap digunakan dalam proses klasifikasi menggunakan algoritma *Iterative Dichotomiser3 (ID3)*. Tahapan *preprocessing* dilakukan untuk meningkatkan kualitas data sehingga hasil klasifikasi yang diperoleh menjadi lebih

	Jenis Kelamin	Umur (bulan)	Tinggi Badan (cm)	Berat Badan (kg)	Lingkar Kepala (cm)	LiLa (cm)	Stunting
0	Laki-laki	15	81.7	13.2	47.2	14.2	Normal
1	Perempuan	9	71.0	11.9	44.1	13.6	Normal
2	Perempuan	17	80.2	9.7	46.8	13.0	Normal
3	Perempuan	20	73.8	14.0	46.5	14.2	Pendek
4	Perempuan	7	74.8	8.3	44.2	13.1	Tinggi
...	...	...	...	...	...	...	...
218	Laki-laki	12	79.2	11.6	45.7	14.2	Normal
219	Laki-laki	4	60.9	3.9	40.6	11.2	Normal
220	Laki-laki	2	64.4	3.0	38.8	11.2	Normal
221	Laki-laki	23	80.3	14.3	47.8	14.9	Pendek
222	Perempuan	8	73.2	12.1	44.1	13.4	Normal

223 rows x 7 columns

Gambar 2. Tampilan Dataset

akurat. Adapun tahapan *preprocessing* data meliputi:

a. Duplicate Data

Pada tahap ini dilakukan pengecekan dan penghapusan data yang memiliki nilai sama atau duplikat. Proses ini bertujuan agar tidak terjadi pengulangan data yang dapat mempengaruhi hasil klasifikasi [4].

b. Missing Values

Tahap ini dilakukan untuk mendeteksi data yang kosong atau tidak lengkap pada setiap atribut. Jika ditemukan data yang hilang, maka dilakukan penanganan agar dataset tetap dapat digunakan dengan baik pada proses pelatihan model [5].

c. Encoding Data

Pada tahap encoding, data kategorikal seperti jenis kelamin dan status *stunting* diubah ke dalam bentuk numerik agar dapat diproses oleh algoritma *Iterative Dichotomiser3 (ID3)*. Proses ini dilakukan menggunakan teknik label encoding sehingga setiap kategori memiliki representasi angka tertentu [6].

C. Pembagian Data

Pembagian data (*data splitting*) merupakan tahapan krusial dalam proses pemodelan yang bertujuan untuk memisahkan *dataset* menjadi beberapa partisi dengan fungsi yang berbeda. Dua komponen utama dalam tahapan ini adalah data latih (*training data*) dan data uji (*testing data*). Data latih digunakan sebagai dasar untuk membangun dan melatih model, sedangkan data uji berfungsi untuk mengevaluasi kinerja model yang telah dibentuk. Proporsi pembagian

data yang tepat sangat penting agar model dapat mempelajari pola klasifikasi secara optimal dan menghasilkan prediksi yang akurat[7].

D. Klasifikasi Algoritma *Iterative Dichotomiser 3* (ID3)

Untuk membuat model prediksi dalam klasifikasi *stunting* pada balita, algoritma ID3 (*Iterative Dichotomiser 3*) merupakan salah satu metode penting yang sering digunakan dalam *machine learning*. Algoritma ini bekerja dengan membentuk pohon keputusan (*Decision Tree*) melalui proses pemilihan atribut terbaik berdasarkan nilai *information gain*. Proses klasifikasi dilakukan dengan menghitung nilai *entropy* dan gain pada setiap atribut untuk menentukan pemecahan data terbaik sehingga terbentuk node-node dalam pohon keputusan.[8]

E. Evaluasi

Proses evaluasi model dilakukan untuk menentukan kombinasi model yang paling efisien dengan melakukan evaluasi melalui metrik seperti akurasi, presisi, recall, dan skor F1. Dalam *machine learning*, Untuk menilai performa suatu model, teknik evaluasi yang umum digunakan adalah *confusion matrix*. Pada *machine learning*, *confusion matrix* juga digunakan untuk mendeteksi permasalahan dalam klasifikasi, terutama ketika output mencakup dua kelas atau lebih.[9]

III. HASIL DAN PEMBAHASAN

A. *Preprocessing* Data

Pada tahap ini, data yang telah terkumpul perlu melalui proses *preprocessing* sebelum digunakan untuk analisis lebih lanjut. Langkah ini bertujuan untuk mengolah data dengan cara mendeteksi dan menangani strukturnya dataset melalui beberapa tahapan[10]. Proses *preprocessing* yang meliputi *duplicate* data, *missing values*, dan *encoding* data.

- Duplicated* data Data *duplicated* merupakan data dengan nilai identik yang muncul lebih dari satu kali di dalam sebuah dataset. Adapun tahap identifikasi duplikasi data merupakan proses pemeriksaan untuk mendeteksi ada tidaknya indikasi data ganda tersebut.

```
jumlah_duplikat = df_balita.duplicated().sum()
print(f"Jumlah data duplikat: {jumlah_duplikat}")

Jumlah data duplikat: 0
```

Gambar 3. Hasil *duplicated* data

- Missing values*

Keadaan ketika terdapat nilai yang tidak lengkap atau hilang pada suatu variabel dalam dataset. Hal ini kerap menjadi masalah utama karena berpotensi memengaruhi hasil.

```
df_balita.isnull().sum() #check missing value

Jenis Kelamin      0
Umur (bulan)       0
Tinggi Badan (cm)  0
Berat Badan (kg)   0
Lingkar Kepala (cm) 0
Lila (cm)          0
Stunting           0
```

Gambar 4. Hasil *missing values*

c. *Encoding data*

Pada tahap *label encoding*, variabel kategorikal seperti 'Jenis Kelamin' ditransformasikan menjadi representasi numerik. Proses ini mengubah nilai kategorikal menjadi angka yang unik agar data dapat diproses.

	Jenis Kelamin	Umur (bulan)	Tinggi Badan (cm)	Berat Badan (kg)	Lingkar Kepala (cm)	LiLa (cm)	Stunting
0	0	15	81.7	13.2	47.2	14.2	Normal
1	1	9	71.0	11.9	44.1	13.6	Normal
2	1	17	80.2	9.7	46.8	13.0	Normal
3	1	20	73.8	14.0	46.5	14.2	Pendek
4	1	7	74.8	8.3	44.2	13.1	Tinggi
...	...	...	...	...	...	...	...
218	0	12	79.2	11.6	45.7	14.2	Normal
219	0	4	60.9	3.9	40.6	11.2	Normal
220	0	2	64.4	3.0	38.8	11.2	Normal
221	0	23	80.3	14.3	47.8	14.9	Pendek
222	1	8	73.2	12.1	44.1	13.4	Normal

223 rows x 7 columns

Gambar 5. Hasil *encoding*

Gambar 5 menunjukkan bahwa variabel pada kategori 'Jenis Kelamin' telah berhasil dikonversi menjadi data numerik. Transformasi ini diperlukan agar algoritma *Iterative Dichotomiser 3* (ID3) dapat memproses dan melakukan klasifikasi secara akurat. Pada representasi numerik tersebut, nilai 0 merepresentasikan laki-laki, sedangkan nilai 1 merepresentasikan perempuan.

B. Pembagian Data

Setelah tahap *preprocessing*, data dibagi menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*). Dalam penelitian ini, *dataset* dipisahkan dengan rasio 80:20, di mana 80% digunakan sebagai data latih dan 20% sebagai data uji. Pembagian ini bertujuan

```

clf = DecisionTreeClassifier(criterion='entropy', random_state=42, max_depth=12)
clf.fit(x_train, y_train)
print("=== Data Latih (Train) ===")
print(f"Jumlah data latih: {len(x_train)} baris")

print("\n=== Data Uji (Test) ===")
print(f"Jumlah data uji: {len(x_test)} baris")

✓ 0.0s
=== Data Latih (Train) ===
Jumlah data latih: 178 baris

=== Data Uji (Test) ===
Jumlah data uji: 45 baris
    
```

Gambar 6. *Split dataset*

untuk mengevaluasi sejauh mana model mampu mengklasifikasikan data secara akurat. Berdasarkan Gambar 6, *dataset* telah dibagi menjadi dua bagian utama dengan rasio 80:20, yaitu 80% untuk proses petihan (*training*) dan 20% untuk pengujian (*testing*) model. Pembagian ini menghasilkan 178 sampel sebagai data latih dan 45 sampel sebagai data uji.

C. Model Algoritma *Iterative Dichotomiser3* (ID3)

Pada tahap ini, algoritma *Iterative Dichotomiser3* (ID3) digunakan untuk mengklasifikasikan data, yang diawali dengan proses pelatihan model. Selama proses

tersebut, model akan belajar dari data latih (training data) untuk mengidentifikasi pola hubungan antara berbagai fitur dengan label target.

```

clf = DecisionTreeClassifier(criterion='entropy', random_state=42, max_depth=12)
clf.fit(x_train, y_train)

y_pred = clf.predict(x_test)

print("Akurasi Test :", metrics.accuracy_score(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test, y_pred))

Akurasi Test : 0.7333333333333333

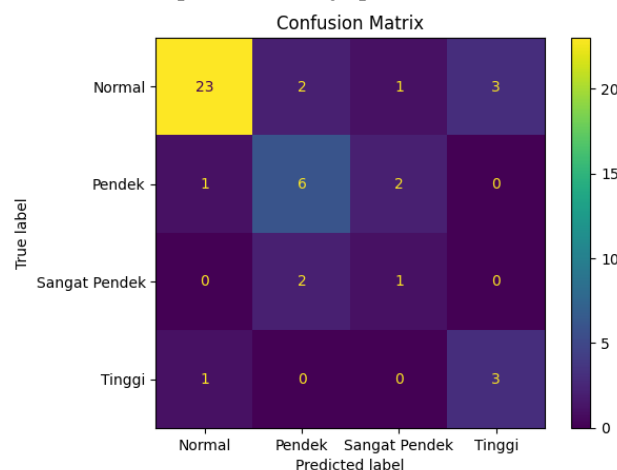
```

Gambar 7. Model klasifikasi

Berdasarkan Gambar 7, model Algoritma *Iterative Dichotomiser3* dikonfigurasi menggunakan parameter *max_depth* sebesar 12. Pengaturan batas kedalaman pohon ini bertujuan untuk mengoptimalkan kinerja model agar menghasilkan tingkat akurasi klasifikasi yang tinggi. Berdasarkan hasil pengujian, diperoleh tingkat akurasi model sebesar 0,7333 atau 73,33%. Hal ini menunjukkan bahwa model mampu memprediksi label kelas pada data uji dengan tepat sebesar 73,33%.

D. Evaluasi Model

Penelitian ini menggunakan confusion matrix sebagai metode untuk mengevaluasi kinerja model Algoritma *Iterative Dichotomiser3* (ID3). Pendekatan ini sangat efektif dalam menilai model klasifikasi karena memberikan gambaran detail mengenai jumlah prediksi yang tepat dan keliru pada masing-masing kategori. Selain itu, confusion matrix menjadi dasar dalam perhitungan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *f1-score*.



Gambar 8. Hasil *confusion matrix*

Gambar 8 memperlihatkan hasil evaluasi model menggunakan *confusion matrix*. Algoritma *iterative dichotomiser 3 (id3)* berhasil diterapkan untuk mengklasifikasikan *stunting*. Model yang dikembangkan mampu memprediksi hasil klasifikasi secara tepat pada 33 data dari total 45 data uji, sehingga diperoleh tingkat akurasi rata-rata sebesar 73,33%. Berdasarkan hasil tersebut, kinerja model yang dibangun masih memerlukan optimasi lebih lanjut untuk mencapai keefektifan yang maksimal.

Tabel 1. *hasil classification report*

Classification Report				
	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
Normal	0.92	0.79	0.85	29
Pendek	0.60	0.67	0.63	9
Sangat pendek	0.25	0.33	0.29	3
Tinggi	0.50	0.75	0.60	4

Berdasarkan Tabel 1, kinerja model terlihat bervariasi pada masing-masing kategori. Kategori 'Normal' menunjukkan performa paling baik dengan *precision* 0,92, *recall* 0,79, dan *f1-score* 0,85. Tingginya nilai metrik ini didukung oleh jumlah data uji (*support*) terbanyak. Sebaliknya, pada kategori 'Sangat Pendek', kinerja model masih sangat kurang optimal dengan *precision* 0,25, *recall* 0,33, dan *f1-score* 0,29. Hal ini mengindikasikan bahwa model gagal mendeteksi kategori ini dengan baik, tingginya angka *false positive* dan *false negative* kemungkinan disebabkan oleh minimnya jumlah sampel data (*support* = 3). Sementara itu, untuk kategori 'Pendek' dan 'Tinggi', model menunjukkan performa menengah dengan perolehan *f1-score* berturut-turut sebesar 0,63 dan 0,60.

IV. KESIMPULAN

Berdasarkan penelitian yang dilakukan, algoritma Iterative Dichotomiser 3 (ID3) berhasil diterapkan untuk mengklasifikasikan status stunting balita berdasarkan atribut jenis kelamin, umur, tinggi badan, berat badan, lingkar kepala, dan LiLa. Hasil pengujian menunjukkan model mampu mencapai akurasi sebesar 73,33% dalam mengklasifikasikan data ke dalam kategori normal, pendek, sangat pendek, dan tinggi.

Penerapan metode *Decision Tree* berbasis ID3 dapat membantu proses deteksi *stunting* menjadi lebih cepat dan terstruktur dibandingkan pencatatan manual di posyandu. Namun, performa model masih dapat ditingkatkan melalui penambahan jumlah data dan optimasi metode klasifikasi agar hasil prediksi menjadi lebih akurat.

DAFTAR PUSTAKA

- [1]. H. Rahman, M. Rahmah, and N. Saribulan, "Stunting Treatment Efforts in Indonesia," *Journal of Science of the Voice of Khatulistiwa (JIPSK)*, vol. VIII, no. 01, pp. 44–59, 2023.
- [2]. M. U. F. Millano, S. Kurniawan, and R. Gustriansyah, "KOMPUTA : Jurnal Ilmiah Komputer dan Informatika," *Penerapan Metode Decision Tree Dalam Klasifikasi Status Gizi Balita*, vol. 14, no. 2, pp. 2–10, 2025, doi: 10.34010/komputa.v14i2.
- [3]. L. D. H. Monasari and Nunik Pratiwi, "Klasifikasi Stunting Pada Balita Menggunakan Algoritma Decision Tree Pada Machine Learning," *Decode: Jurnal Pendidikan Teknologi Informasi*, vol. 5, no. 2, pp. 416–428, 2025, [Online]. Available: <https://journal.umkendari.ac.id/decode/article/view/1193>
- [4]. T. S. T. H. F. H. F. E. N. A. N. H. K. K. L. Y. Y. W. Rinci Kembang Hapsari, *DATA MINING*. Bandung, 2024. [Online]. Available: www.indiepress.id

- [5]. A. Widiyanti and I. Pratama, "PENANGANAN MISSING VALUES DAN PREDIKSI DATA TIMBUNAN SAMPAH BERBASIS MACHINE LEARNING," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 9, no. 2, pp. 242–251, Jul. 2024, doi: 10.36341/rabit.v9i2.4789.
- [6]. C. Herdian, A. Kamila, and I. G. Agung Musa Budidarma, "Studi Kasus Feature Engineering Untuk Data Teks: Perbandingan Label Encoding dan One-Hot Encoding Pada Metode Linear Regresi," *Technologia : Jurnal Ilmiah*, vol. 15, no. 1, p. 93, Jan. 2024, doi: 10.31602/tji.v15i1.13457.
- [7]. S. Marsya Finda and D. Wahyu Utomo, "Klasifikasi Stunting Balita menggunakan Metode Ensemble Learning dan Random Forest," *Jl. Imam Bonjol No*, vol. 15, no. 02, 2024, doi: 10.35970/infotekmesin.v15i2.2326.
- [8]. J. Tetra, M. Aliyah Nazanah, and M. I. Jambak, "KLIK: Kajian Ilmiah Informatika dan Komputer Pemanfaatan Algoritma Decision Tree ID3 Bagi Manajemen Bimbel Untuk Menentukan Faktor Kelulusan Pada Sekolah Kedinasan," *Media Online*, vol. 3, no. 6, pp. 915–924, 2023, doi: 10.30865/klik.v3i6.791.
- [9]. S. A. Hasibuan, V. Sihombing, and F. A. Nasution, "Analysis of Community Satisfaction Levels using the Neural Network Method in Data Mining," *Sinkron*, vol. 8, no. 3, pp. 1724–1735, Jul. 2023, doi: 10.33395/sinkron.v8i3.12634.
- [10]. A. A. Firdaus, A. Yudhana, and I. Riadi, "Public Opinion Analysis of Presidential Candidate Using Naïve Bayes Method," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, May 2023, doi: 10.22219/kinetik.v8i2.1686.