

Rancangan Aplikasi Algoritma C4.5 pada Stunting Balita Menggunakan Bahasa Phyton

¹Sigit Yugi Wargiyo, ²Susliansyah, ³Heny Sumarno, ⁴Hendro Priyono, ⁵Linda Maulida

^{1,2,3,4,5}Universitas Bina Sarana Informatika

Jakarta, Indonesia

¹sigit.syw@bsi.ac.id, ²susliansyah.slx@bsi.ac.id, ³heny.hnm@bsi.ac.id, ⁴hendro.hop@bsi.ad

⁵linda.lma@bsi.ac.id

*Penulis Korespondensi

Diajukan : 02/12/2024

Diterima : 17/01/2025

Dipublikasi : 19/01/2025

ABSTRAK

Stunting pada balita merupakan salah satu masalah kesehatan serius di Indonesia, yang memengaruhi pertumbuhan fisik dan kognitif anak. Dalam upaya memahami dan memprediksi faktor-faktor risiko yang berkaitan dengan *stunting* pada balita, digunakan teknologi *data mining*. Penelitian ini bertujuan mengembangkan aplikasi berbasis algoritma C4.5 untuk memprediksi status gizi balita, menggunakan bahasa pemrograman *Python* dan aplikasi *Orange*. Data yang berasal dari dataset "*Stunting Toddler Detection*" di Kaggle, dengan fokus pada variabel umur, tinggi badan, dan status gizi. Data tersebut digunakan sebagai bahan analisis, dengan tahapan *preprocessing*, integrasi data, hingga penerapan algoritma C4.5. Metode penelitian melibatkan pengolahan data menggunakan *Python* untuk analisis awal, sementara *Orange* dimanfaatkan untuk membangun pohon keputusan dan evaluasi model. Hasil pengujian menunjukkan algoritma C4.5 menghasilkan akurasi sebesar 36% di *Orange* dan 40% di *Python*, dengan faktor utama yang memengaruhi status gizi balita adalah tinggi badan. Aplikasi yang dikembangkan juga dilengkapi antarmuka visual untuk mempermudah tenaga kesehatan dan pemangku kebijakan dalam menganalisis risiko *stunting*.

Kata Kunci: Bahasa Phyton, C4.5, Stunting Balita

I. PENDAHULUAN

Stunting pada balita merupakan salah satu masalah kesehatan serius yang dihadapi di banyak negara, termasuk Indonesia. Kementerian Kesehatan mengumumkan hasil Survei Status Gizi di Indonesia (SSGI) pada Rapat Kerja Nasional BKKBN, dimana prevalensi *stunting* di Indonesia turun dari 24,4% di tahun 2021 menjadi 21,6% di 2022 (Amanda Amalia & Lestari, 2023). Kondisi *stunting* ditandai dengan tinggi badan yang berada di bawah standar untuk anak seusianya dan merupakan indikator kekurangan gizi kronis yang berkepanjangan. *Stunting* tidak hanya berdampak pada kondisi fisik anak, tetapi juga memengaruhi perkembangan kognitif dan produktivitas anak di masa depan.

Dalam upaya memahami dan memprediksi faktor-faktor risiko yang berkaitan dengan *stunting* pada balita, pemanfaatan teknologi *data mining* menjadi semakin penting. *Data mining* merupakan penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data dalam jumlah besar yang diharapkan dapat mengatasi kondisi tersebut. *Data mining* memungkinkan penggalian pola dan hubungan tersembunyi dalam data yang kompleks, sehingga dapat membantu dalam pengambilan keputusan berbasis data (Sari et al., 2021).

Data mining sendiri memiliki beberapa Teknik, salah satunya yaitu klasifikasi. Klasifikasi merupakan salah satu proses pada *data mining* yang bertujuan untuk menemukan pola yang berharga dari data yang berukuran relatif besar hingga sangat besar. Algoritma yang digunakan dalam penelitian ini yaitu algoritma C4.5 dengan Teknik pohon keputusan (*decision tree*), yang dikenal efektif dalam klasifikasi dan pengambilan keputusan berbasis data. Algoritma ini mampu menghasilkan pohon keputusan yang mudah dipahami serta memberikan informasi penting mengenai variabel-variabel yang berpengaruh dalam klasifikasi data (Kurniah et al., 2022).

Penelitian ini bertujuan untuk mengembangkan aplikasi prediksi *stunting* pada balita dengan menerapkan algoritma C4.5 melalui bahasa pemrograman *Python* dan aplikasi *Orange*. *Python* dipilih karena memiliki banyak pustaka yang mendukung proses *data mining* dan *machine learning*, sementara *Orange* merupakan perangkat lunak visualisasi *data mining* yang memungkinkan analisis data dengan mudah. Dengan mengombinasikan keduanya, diharapkan penelitian ini mampu menghasilkan sistem yang akurat dan efisien dalam mengidentifikasi faktor risiko *stunting* pada balita serta memberikan wawasan yang bermanfaat bagi tenaga kesehatan dan pemangku kebijakan dalam upaya penanggulangan *stunting*.

II. STUDI LITERATUR

1. *Data Mining*

Data mining adalah suatu proses ekstraksi pengetahuan atau informasi yang berharga dari suatu dataset yang besar dan kompleks. Proses *data mining* melibatkan penggunaan berbagai teknik statistik, matematis, dan kecerdasan buatan untuk menganalisis data dengan cara yang sistematis dan otomatis. Hasil dari *data mining* dapat digunakan untuk mendukung pengambilan keputusan, mengidentifikasi tren pasar, meningkatkan efisiensi operasional, atau merumuskan strategi bisnis (Rahayu et al., 2024).

Data mining bukan hanya tentang aplikasi algoritma dan teknik analisis, tetapi juga melibatkan pemahaman konteks bisnis dan interpretasi hasil untuk pengambilan keputusan yang efektif. Tujuan utama dari *data mining* adalah mengidentifikasi hubungan atau pola yang dapat membantu dalam pengambilan keputusan dan memahami fenomena yang terkandung dalam data (Masruriyah et al., 2024).

2. *Algoritma C4.5*

Algoritma C4.5 adalah algoritma klasifikasi data dengan teknik pohon keputusan yang memiliki kelebihan-kelebihan. Kelebihan ini misalnya dapat mengolah data numerik (kontinyu) dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat diantara algoritma-algoritma yang lain (Latifah et al., 2018). Algoritma C4.5 atau pohon keputusan merupakan metode yang umum digunakan untuk melakukan klasifikasi pada data mining. Metode ini populer karena mampu melakukan klasifikasi sekaligus menunjukkan hubungan antar atribut ((Riandari & Simangunsong, 2019)).

Tahapan dalam algoritma C4.5 yang pertama adalah menghitung nilai *entropy* total dari setiap label, kemudian menghitung nilai *entropy* dari masing-masing atribut. Setelah mendapatkan nilai *entropy* dari masing-masing atribut dapat diketahui nilai *gain*. Nilai *gain* dari setiap atribut inilah yang menentukan node dalam pohon Keputusan (Islam et al., 2022).

$$\text{Entropy}(S) = \sum_{i=1}^n - p_i * \log_2$$

Keterangan:



S = Himpunan kasus

n = Jumlah partisi S

p_i = Proporsi dari S_i dan S

$$\text{Gain}(S,A) = \text{Entropy} - \sum_{i=1}^n \frac{S_i}{S} * \text{entropy}(S_i)$$

Keterangan:

S = Himpunan kasus A = Atribut

n = Jumlah partisi atribut A

$|S_i|$ = Jumlah kasus pada partisi ke- i

$|S|$ = Jumlah kasus dalam S

3. Python

Python adalah bahasa pemrograman tingkat tinggi, berjalan dengan sistem *interpreted*, dan bisa dipakai untuk berbagai jenis tujuan. sebuah bahasa pemrograman disebut sebagai bahasa pemrograman tingkat tinggi karena perintah atau kode program yang dipakai sudah mirip dengan bahasa manusia. *Python* menggunakan metode pemrosesan *interpreted*, kode program akan diproses baris per baris langsung dari kode program (tidak butuh proses compile) (Sembiring, 2021).

Python merupakan bahasa pemrograman tingkat tinggi yang dibuat oleh Guido Van Rossum dan dirilis pada tahun 1991. *Python* juga merupakan bahasa yang sangat populer belakangan ini. Selain itu *python* juga merupakan bahasa pemrograman yang multi fungsi contohnya *Python* dapat digunakan untuk *Machine Learning* dan *Deep Learning*. *Python* dipilih sebagai penelitian karena *Python* memiliki penulisan sintaksis yang mudah selain itu *Python* juga memiliki *library* yang lengkap dan memiliki dukungan komunitas yang kuat karena *Python* bersifat *open source*. Untuk menuliskan *source code Python* dapat menggunakan IDE seperti vs code, sublime text, PyCharm atau anda juga dapat menggunakan IDE online seperti Jupyter notebook dan google colab (Riziq sirfatullah Alfarizi et al., 2023).

4. Orange

Aplikasi *Orange* merupakan perangkat lunak yang dapat dijadikan suatu alat untuk dapat mengukur pencemaran udara yang bersifat *opensource*. *Orange* menyediakan pemrograman visual yang berbasis pada komponen-komponen untuk penambangan data, analisis data, pembelajaran mesin dan visualisasi data. Aplikasi *Orange* sudah dikembangkan sejak tahun 1996 di Ljubljana University dan Jožef Stefan Institute (Sitorus et al., 2022).

Perangkat lunak ini merupakan sebuah visual programming berbasis komponen yang digunakan untuk visualisasi data, *machine learning*, dan analisis data tanpa perlu membuat kode program. Komponen pada perangkat lunak *Orange* disebut dengan *widget*. Komponen *widget* ini memiliki fungsi dari visualisasi data sederhana, pemilihan subset data, dan preprocessing. Untuk evaluasi secara empiris, *widget* juga dapat digunakan untuk evaluasi hasil algoritma pembelajaran prediktif (Yunardi & Dina, 2022).

III. METODE

Penelitian ini menggunakan data sekunder yang diperoleh dari sumber-sumber terpercaya seperti data kesehatan nasional, survei status gizi, dan data demografis terkait stunting pada balita. Data sekunder dipilih karena memungkinkan analisis pola pada sampel yang lebih besar dan memiliki cakupan yang lebih luas dibandingkan dengan pengumpulan data primer. Berikut tahapan pengumpulan dan pengolahan data secara rinci:

1. Identifikasi dan Pemilihan Sumber Data

Data sekunder yang digunakan pada penelitian ini merupakan dataset “Stunting Toddler (Balita) Detection” yang diperoleh dari Kaggle. Data ini berupa kumpulan data berdasarkan rumus z-score penentuan stunting menurut WHO (World Health Organization), yang berfokus pada deteksi stunting pada balita (bayi dibawah lima tahun). Dataset ini merinci informasi mengenai umur, jenis kelamin, tinggi badan, dan status gizi balita.

2. Preprocessing & Cleansing

- Secara umum, preprocessing adalah tahapan awal dalam pengolahan data di mana data mentah (raw data) dipersiapkan dan diubah menjadi format yang lebih sesuai untuk dianalisis. Hal ini memiliki pengaruh terhadap hasil dari proses data mining. Kualitas data yang digunakan untuk proses data mining, berbanding lurus dengan hasil yang akan didapat (Hakim, 2021).
- Cleansing atau data cleaning adalah bagian dari preprocessing yang berfokus pada identifikasi dan perbaikan kesalahan, menghapus duplikasi, dan menyelaraskan format dalam data, sehingga dapat memanfaatkan data tersebut secara efektif untuk pengambilan keputusan. Tujuannya adalah untuk memastikan data akurat, konsisten, dan valid (Hanafi, 2023).

3. Integrasi Data ke Dalam Python dan Orange

- Data yang telah diproses selanjutnya diimpor ke dalam Python menggunakan pustaka seperti Pandas untuk analisis lebih lanjut dan manipulasi data jika diperlukan.
- Setelah itu, data diimpor ke dalam aplikasi Orange, perangkat lunak data mining berbasis visualisasi yang mendukung algoritma C4.5, sehingga memudahkan dalam membangun pohon keputusan dan menganalisis hubungan antar variabel yang mempengaruhi stunting pada balita.

4. Penggunaan Algoritma C4.5 untuk Analisis Data

Dengan menggunakan Orange, data akan diolah dengan algoritma C4.5 untuk menghasilkan pohon keputusan yang memetakan hubungan antara variabel independen (faktor-faktor risiko) dan variabel dependen (status stunting). Model C4.5 kemudian diuji dan divalidasi untuk memastikan keakuratan dalam memprediksi faktor risiko stunting.

5. Analisis Hasil dan Interpretasi

Data yang telah diproses oleh algoritma C4.5 kemudian dianalisis untuk mengidentifikasi faktor-faktor utama yang memiliki pengaruh signifikan terhadap stunting pada balita. Hasil analisis dari Python dan Orange digunakan untuk membuat laporan serta interpretasi yang dapat dimanfaatkan sebagai rekomendasi bagi kebijakan kesehatan dalam upaya penanggulangan stunting. Beberapa ukuran yang digunakan dalam stunting balita adalah akurasi, presisi, recall, specificity dan F1-score.

IV. HASIL DAN PEMBAHASAN

Pengujian terhadap analisa sangat penting dilakukan untuk menentukan dan memastikan apakah hasil analisa tersebut sesuai dengan Keputusan yang diharapkan (Amalia et al., 2023). Pada tahapan ini, proses data mining menerapkan algoritma C4.5 dengan menggunakan Bahasa pemrograman Python dengan tools Google Colab, lalu dilakukan pengujian model menggunakan aplikasi Orange.

1. Penerapan Algoritma C4.5 dengan Bahasa Pemrograman Python

a. Pengolahan data menggunakan Python.

Pada tahap pengolahan data ini menggunakan libraries dari Python untuk data science seperti pandas, numpy, seaborn, matplotlib, mlxtend, dan scikit-learn yang berfungsi sebagai pencetak grafik atau visualisasi berdasarkan inputan data dengan menggunakan algoritma C4.5 yang termasuk dalam bagian dari supervised learning.

1) Menampilkan Data Stunting Balita

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 4 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   Umur (bulan)           50 non-null    int64
1   Jenis Kelamin          50 non-null    object
2   Tinggi Badan (cm)      50 non-null    float64
3   Status Gizi            50 non-null    object
dtypes: float64(1), int64(1), object(2)
```

Gambar 1. Informasi dasar dataset Informasi dataset sederhana dari 50 data balita yang diambil.

	Umur (bulan)	Jenis Kelamin	Tinggi Badan (cm)	Status Gizi
0	13	laki-laki	79.8	normal
1	27	perempuan	76.0	severely stunted
2	10	perempuan	66.3	stunted
3	16	laki-laki	88.6	tinggi
4	20	perempuan	76.9	normal
5	43	perempuan	84.4	severely stunted
6	39	laki-laki	112.2	tinggi
7	21	laki-laki	87.7	normal
8	40	perempuan	110.4	tinggi
9	40	perempuan	109.4	normal

Gambar 2. 10 data teratas dari dataset

Dataset ini memberikan gambaran awal yang dapat dianalisis lebih lanjut untuk memahami hubungan antara umur, jenis kelamin, tinggi badan (cm), dan status gizi.

Status Gizi	count
normal	15
tinggi	13
severely stunted	12
stunted	10

Gambar 3. Total kolom status gizi

Kolom dari status gizi di dataset terbagi menjadi 4: normal, tinggi, stunted, severely stunted.

2) Preprocessing & Cleansing

```
[15] # menghapus/menghilangkan spasi di awal dan di akhir
data['Jenis Kelamin'] = data['Jenis Kelamin'].str.strip()
data['Status Gizi'] = data['Status Gizi'].str.strip()

[16] # menghapus semua baris yang memiliki nilai NULL pada kolom 'Umur (bulan)'
data.dropna(axis=0, subset=['Umur (bulan)'], inplace=True)
```

Gambar 4. Preprocessing dan cleansing

Membersihkan dan membuang data yang kotor (tidak diperlukan).

3) Data Transformation

Pada penelitian ini, Data transformation digunakan untuk mentransformasi dataset yang berisi informasi tentang *stunting* balita untuk fokus pada hubungan antara usia, status gizi, dan tinggi badan dengan jenis kelamin (membandingkan distribusi

Status Gizi	normal	severely stunted	stunted	tinggi
Umur (bulan)				
7	0.0	0.0	0.0	1.0
8	1.0	1.0	0.0	0.0
10	0.0	0.0	0.0	1.0
13	1.0	0.0	0.0	1.0
16	0.0	0.0	0.0	1.0

pengukuran tinggi badan di berbagai kategori status gizi untuk anak laki-laki dan Perempuan dari berbagai usia).

Gambar 5. Kategori jenis kelamin laki-laki
 Status Gizi jenis kelamin laki-laki yang dikelompokkan berdasarkan Umur (bulan).

Status Gizi	normal	severely stunted	stunted	tinggi
Umur (bulan)				
1	0.0	1.0	0.0	1.0
3	1.0	0.0	0.0	0.0
6	0.0	0.0	0.0	1.0
10	0.0	0.0	1.0	0.0
18	0.0	1.0	0.0	0.0

Gambar 6. Kategori jenis kelamin Perempuan
 Status Gizi jenis kelamin perempuan yang dikelompokkan berdasarkan Umur (bulan).

4) Evaluasi Model

Evaluasi model adalah merupakan penting dalam analisis data terstruktur maupun tidak terstruktur. Evaluasi model membantu kita mengetahui seberapa baik model kita dalam memberikan hasil yang tepat. Untuk mengevaluasi model, kita dapat menggunakan beberapa metode atau ukuran yang bisa memberikan gambaran objektif tentang kinerja model. Beberapa ukuran yang sering digunakan adalah akurasi, presisi, *recall*, *specificity*, dan *F1-score*. Pada evaluasi model terdapat *confusion matrix*, yang merupakan alat untuk menggambarkan kinerja model klasifikasi pada data uji yang sudah diketahui hasil sebenarnya.

Accuracy: 0.40

	precision	recall	f1-score	support
0	0.40	0.67	0.50	3
1	0.50	0.33	0.40	3
2	0.00	0.00	0.00	1
3	1.00	0.33	0.50	3
accuracy			0.40	10
macro avg	0.47	0.33	0.35	10
weighted avg	0.57	0.40	0.42	10

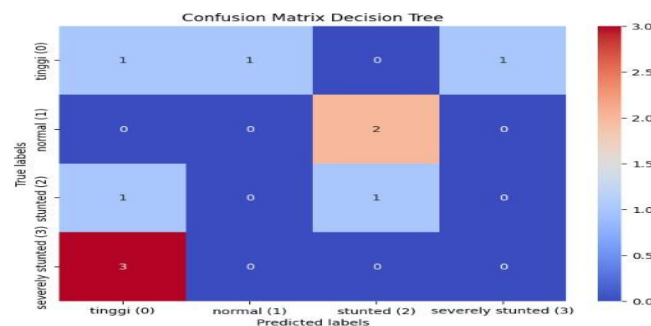
```

[[2 0 1 0]
 [1 1 1 0]
 [0 1 0 0]
 [2 0 0 1]]

```

Gambar 7. Hasil Akurasi, Presisi, dan Recall

Matrix precision, recall, f1-score, dan support yang ditampilkan dalam hasil evaluasi klasifikasi.

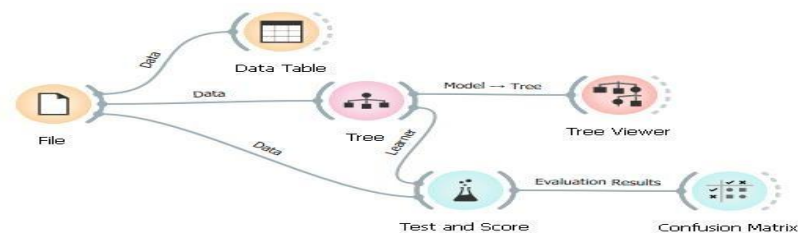


Gambar 8. Confusion matrix

Digunakan untuk mengevaluasi kinerja model klasifikasi dalam *machine learning*.

2. Penerapan Algoritma C4.5 dengan Tool Orange

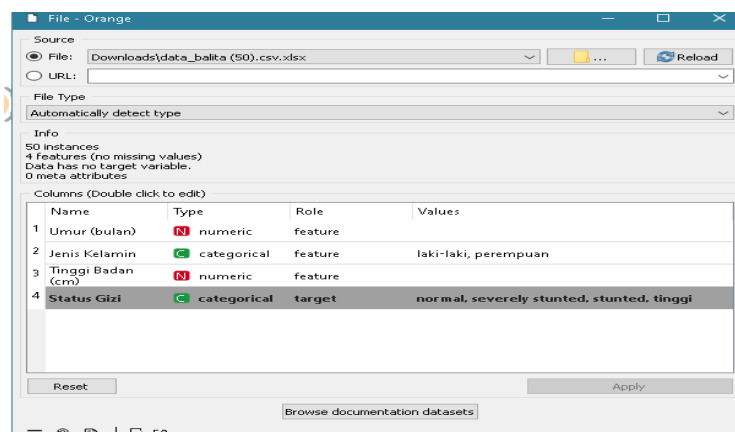
a. Struktur Workflow



Gambar 9. Struktur Workflow

Rangkaian langkah yang digunakan untuk menganalisis dan memproses data menggunakan antarmuka grafis berbasis drag-and-drop.

b. Dataset Stunting Balita



Gambar 10. Dataset *Stunting* Balita

Dataset yang menjadi *target variable* yang menggambarkan status gizi balita.

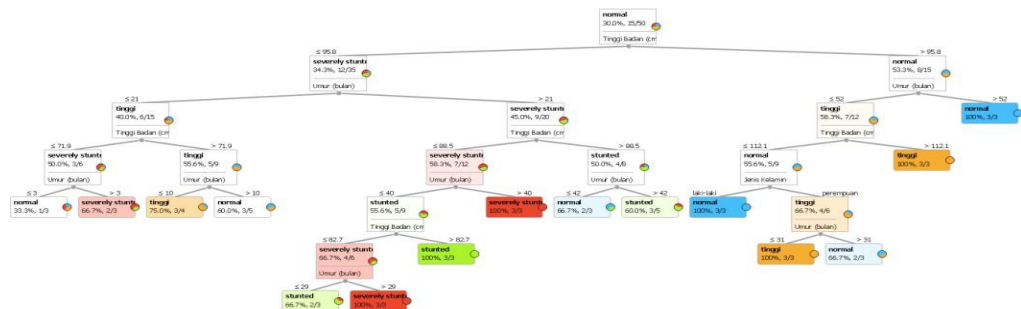
c. Output Data Tabel

	Status gizi	Umur (bulan)	Jenis Kelamin	Tinggi Badan (cm)
1	severely stunted	13	laki-laki	79.0
2	severely stunted	27	perempuan	76
3	stunted	10	perempuan	66.3
4	tinggi	16	laki-laki	86.6
5	normal	20	perempuan	76.9
6	severely stunted	43	perempuan	64.4
7	tinggi	39	laki-laki	112.2
8	normal	21	laki-laki	87.7
9	tinggi	40	perempuan	110.4
10	normal	40	perempuan	100.4
11	normal	37	laki-laki	102.9
12	stunted	37	laki-laki	89.9
13	stunted	36	perempuan	87.9
14	severely stunted	31	laki-laki	80.7
15	severely stunted	31	laki-laki	81
16	stunted	28	laki-laki	87.2
17	normal	29	laki-laki	89.8
18	tinggi	52	perempuan	120.5
19	severely stunted	52	perempuan	88.5
20	normal	56	perempuan	109.2
21	normal	3	perempuan	66.1

Gambar 11. Output Data Table

Output seluruh data balita yang ada pada dataset (50 data).

d. Output Decision Tree



Gambar 16. Output Decision Tree

Berdasarkan gambar hasil dari perhitungan pohon keputusan didapatkan sebanyak 28 rules yang mempengaruhi status gizi pada balita. Hasil dari 8 atribut yang diuji menunjukkan bahwa atribut yang paling mempengaruhi balita dengan status gizi *severely stunted* adalah tinggi badan, dan diikuti oleh Umur dengan pola variabel yang mempengaruhi adalah pada tinggi badan ≤ 95 cm, dan umur > 21 bulan.

3. Desain Interface Aplikasi Algoritma C4.5

Desain *Interface* merupakan penggambaran tampilan dari aplikasi prediksi *stunting* pada balita yang akan dibangun.

a. Form sign up dan login



Gambar 17. Form *Sign Up* dan *Login*

b. Form *Login* setelah meng-klik *button Login* di Form *login*



Gambar 18. Form *Login* setelah meng-klik *button Login* di Form *login* (bisa mengakses atau masuk dengan berbagai pilihan : Google, Facebook dan Nomor Telepon (WhatsApp)).

c. Tampilan untuk memilih dan mengetahui porsi dan standar Gizi



Gambar 19. Tampilan untuk memilih dan mengetahui porsi dan standar Gizi; Baik, Normal dan Buruk.

- d. Tampilan Decision Tree Untuk Memprediksi Status Gizi Balita dengan Jenis Kelamin Perempuan



Gambar 20. Tampilan *Decision Tree* untuk memprediksi status gizi balita dengan jenis kelamin perempuan.

- e. Tampilan Input Data Diri dan Karakteristik Balita

Gambar 21. Input data diri dan karakteristik Balita (*Toddler*) untuk pengecekan, pengukuran serta prediksi gizi.

- f. Tampilan For Hasil



Gambar 22. Form Hasil (*Output*) setelah menginput data diri dan karakteristik Balita (*Toddler*) ; apakah gizinya terpenuhi atau belum (*Congrats!* Muncul ketika gizi sudah baik dan normal).

V. KESIMPULAN

Penelitian ini telah menghasilkan klasifikasi status gizi pada balita menggunakan bahasa pemrograman *Python* dengan tools *Google Colab* dan aplikasi *Orange*. Dengan menggunakan *Python*, hasil akurasi pada penelitian ini menunjukkan angka 40%. Sedangkan pada pengujian model algoritma C4.5 menggunakan aplikasi *Orange*, mendapatkan hasil akurasi sebesar 36. Dari pohon keputusan yang telah terbentuk, variabel tertinggi dalam memprediksi status gizi *severely stunted* adalah variabel tinggi badan ≤ 95 cm, artinya tinggi badan sangat berpengaruh terhadap status gizi pada balita. Sedangkan variabel yang tidak terlalu berpengaruh adalah variabel jenis kelamin. Untuk pengembangan penelitian yang akan datang, diharapkan menambahkan data responden dan kriteria lebih banyak lagi dari data yang digunakan dalam penelitian tersebut, agar hasil yang diperoleh lebih bisa dijadikan patokan yang lebih akurat dan dapat dibandingkan dengan algoritma lain seperti algoritma k-means, algoritma apriori dan algoritma naïve bayes.

VI. REFERENSI

- Amalia, F., Aditya Purnama, M., & Studi Teknologi Rekayasa Perangkat Lunak Politeknik Kelapa Sawit Citra Widya Edukasi -Bekasi Email Penulis Korespondensi, P. (2023). Klasifikasi Peserta Didik Menggunakan Algoritma C4.5. *Jurnal Citra Widya Edukasi*, 15(3).
- Amanda Amalia, R., & Lestari, S. (2023). Penerapan Algoritma C.45 Pada Klasifikasi Status Gizi Balita di Posyandu Desa Sukalilah Cibatu Kabupaten Garut Jawa Barat. *Jurnal Sains Dan Teknologi*, 5(1), 177–182. <https://doi.org/10.55338/saintek.v5i1.1375>
- Hakim, B. (2021). Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning. *JBASE - Journal of Business and Audit Information Systems*, 4(2). <https://doi.org/10.30813/jbase.v4i2.3000>
- Hanafi, H. (2023). *Data Cleaning dalam Big Data : Review*. <https://www.researchgate.net/publication/376758940>
- Islam, H. I., Khandava Mulyadien, M., & Enri, U. (2022). Penerapan Algoritma C4.5 dalam Klasifikasi Status Gizi Balita. *Jurnal Ilmiah Wahana Pendidikan*, 8(10), 116–125. <https://doi.org/10.5281/zenodo.6791722>
- Kurniah, R., Yunika Surya Putra, D., & Diana, E. (2022). Penerapan Data Mining Decission Tree Algoritma C4.5 Untuk Mengetahui Tingkat Kepuasan. *Jurnal Informatika Dan Teknologi*, 5(2), 316. <https://doi.org/10.29408/jit.v5i2.5910>
- Latifah, K., Wibowo, S., & Qotrun Nada, N. (2018). Analisis Dan Penerapan Algorithma C45 Dalam Data Mining Untuk Menunjang Strategi Promosi Prodi Informatika Upgris. *JURNAL TEKNIK INFORMATIKA*, 11(2), 109–120. <https://doi.org/10.15408/jti.v11i2.6706>
- Masruriyah, A. F. N., Sukmawati, C. E., & Dermawan, B. A. (2024). *Memahami Data Mining Dengan Python: Implementasi Praktis*. Eureka Media Aksara.
- Prastyadi Wibawa Rahayu, S. Kom. , M. K., Suryani, SKom. , M. T., Achmad Ridwan, SKom. , M. K., I Gede Iwan Sudipa, S. Kom. , M. C., Arie Surachman, M. K., I Gede Mahendra Darmawiguna, S. Kom. , M. S., Ir. Muh. Nurtanzis Sutoyo, S. Kom. , M. Cs. , I., Drs. Isnandar Slamet, M. Sc. , Ph. D., Sitti Harlina, SE. , M. K., & I Made Dendi Maysanjaya, S. Pd. , M. E. (2024). *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia. <https://www.researchgate.net/publication/377415198>
- Riandari, F., & Simangunsong, A. (2019). *Penerapan Algoritma C4.5 Untuk Mengukur Tingkat Kepuasan Mahasiswa*. CV. Rudang Mayang.

-
- Riziq sirfatullah Alfarizi, M., Zidan Al-farish, M., Taufiqurrahman, M., Ardiansah, G., & Elgar, M. (2023). Penggunaan Python Sebagai Bahasa Pemrograman Untuk Machine Learning Dan Deep Learning. *Karimah Tauhid*, 2(1).
- Sari, M., Perdana Windarto, A., & Okprana, H. (2021). Penerapan Data Mining Klasifikasi C4.5 Pada Penerima Beasiswa di SMK Swasta Anak Bangsa. *BEES: Bulletin of Electrical and Electronics Engineering*, 1(3), 115–121.
- Sembiring, F. (2021). *Buku Ajar Dasar Pemrograman (Python)*. Nusaputra Press.
- Sitorus, M., Fitron, D., & Agung Segara Wisesa, C. (2022). Implementasi Algoritma K-Means Menggunakan Aplikasi Orange dalam Clustering Pencemaran Udara di DKI Jakarta Tahun 2021. *Journal of Informatics and Advanced Computing (JIAC)*, 3.
- Yunardi, R. T., & Dina, N. Z. (2022). *Data Mining dan Machine Learning dengan Orange: Tutorial dan Aplikasinya*. Airlangga University Press.