

Performansi Algoritma C4.5 untuk Prediksi Gizi pada Balita

Ahmad Nur Sahal^{1*}, Siska Narulita²
Institut Teknologi dan Bisnis Semarang

*Email: sahalnur2@gmail.com

Abstrak

Pemenuhan gizi yang baik merupakan komponen penting dari kesehatan yang optimal. Namun berbagai gangguan gizi dan gizi buruk disebabkan oleh buruknya kualitas makanan atau jumlah makanan yang tidak mencukupi kebutuhan tubuh. Terutama kesehatan pada gizi balita. Anak pada usia ini sangat rentan mengalami masalah gizi, karena asupan gizi yang cukup mutlak diperlukan dalam masa pertumbuhan dan perkembangannya. Tumbuh kembang anak menjadi salah satu hal yang penting bagi para orang tua dalam memperhatikan masalah gizi pada anak. Pola makan juga berpengaruh pada keadaan gizi balita. Gizi merupakan asupan terpenting untuk tumbuh kembang anak, karena gizi yang baik membantu anak dalam berkembang secara normal. Oleh karena itu perlu dilakukan pengolahan nilai gizi pada balita untuk memprediksi akurasi terlepas dari apakah bayi tersebut bernilai gizi buruk, gizi baik, atau gizi lebih. Dari hasil pengujian dan evaluasi performansi algoritma *data mining* C4.5 menghasilkan nilai akurasi sebesar 100% pada penggunaan perbandingan *data training* dan *data testing* 90%:10%, 80%:20%, dan 70%:30%.

Kata kunci: *Data Mining; Prediksi; C4.5; Gizi Balita*

Performance of the C4.5 Algorithm for Nutrition Prediction in Toddlers

Abstract

Fulfillment of good nutrition is an important component of optimal health. However, various nutritional disorders and malnutrition are caused by poor food quality or an amount of food that is not sufficient for the body's needs. especially in toddler nutrition. Children at this age are very vulnerable to experiencing nutritional problems because adequate nutritional intake is absolutely necessary during their growth and development. Child growth and development are important for parents to pay attention to nutritional problems in children. Diet also influences the nutritional status of toddlers. Nutrition is the most important intake for children's growth and development because good nutrition helps children develop normally. Therefore, it is necessary to process the nutritional value of toddlers to predict accuracy, regardless of whether the baby is undernourished, well-nourished, or overnourished. From the results of testing and evaluating the performance of the C4.5 data mining algorithm, it produces an accuracy value of 100% when using training data and testing data comparisons of 90%:10%, 80%:20%, and 70%:30%.

Keywords: *Data Mining; Prediction; C4.5; Toddler Nutrition*

Pendahuluan

Sehat merupakan suatu kondisi atau keadaan sejahtera secara fisik, mental, maupun sosial yang utuh dan bukan hanya sekedar bebas dari suatu penyakit atau

Performansi Algoritma C4.5 untuk Prediksi Gizi pada Balita

kelemahan/cacat (Allensworth, 2010). Kesehatan juga dapat dikatakan sebagai bentuk keseimbangan antara individu (sebagai inang), patogen (bakteri, virus, racun, dan lain sebagainya), serta lingkungan. Oleh karena itu, interaksi tidak hanya terjadi antara individu dan patogen, tetapi juga dengan lingkungan, yang menciptakan suatu kondisi untuk berkembang biak (Allensworth, 2010). Permasalahan kesehatan seringkali diabaikan atau dipandang sebelah mata oleh masyarakat, terutama kesehatan pada balita. Kesehatan pada anak tidak hanya dipengaruhi oleh asupan makanan, tetapi juga oleh kebersihan lingkungan, perilaku orang tua, dan kondisi ibu (Bappenas, 2018). Anak pada usia balita ini sangat rentan mengalami masalah gizi, karena asupan gizi yang cukup, mutlak diperlukan dalam masa pertumbuhan dan perkembangan balita (Marsita, 2018).

Pertumbuhan dan perkembangan anak menjadi salah satu hal yang sangat penting bagi para orang tua dalam memperhatikan masalah gizi pada anak. Pola makan juga mempengaruhi keadaan gizi balita. Gizi merupakan asupan terpenting untuk tumbuh kembang anak, karena gizi yang baik membantu anak untuk dapat berkembang secara normal. Berdasarkan Riset Kesehatan Dasar (Riskesdas) tahun 2018 menyatakan 17,7% balita mengalami masalah gizi dengan rincian 3,9% mengalami gizi buruk, dan 13,8% mengalami gizi kurang (Kementerian Kesehatan RI, 2013). Sangat penting untuk mengetahui kondisi balita sejak dini, agar penanggulangan gizi buruk pada balita dapat teratasi. Untuk membantu memprediksi karakteristik gizi balita (Narulita et al., 2023), dapat memanfaatkan metode *data mining*.

Data mining itu sendiri merupakan proses penggalian data dari sebuah kumpulan data untuk mendapatkan pola hubungan antar data yang nantinya dapat digunakan sebagai informasi untuk pemecahan masalah tertentu (Riani et al., 2019). Klasifikasi merupakan salah satu teknik dalam metode *data mining*. Teknik klasifikasi digunakan untuk melakukan prediksi, pengelompokan atau pengklasifikasian suatu kelas berdasarkan label-label tertentu (Wanto et al., 2020). Untuk melakukan prediksi nilai gizi pada balita atau anak-anak, peneliti menggunakan metode *data mining* prediksi algoritma C4.5 atau *decision tree* dengan berdasarkan atribut jenis kelamin, umur, berat badan, dan tinggi. Algoritma C4.5 merupakan algoritma klasifikasi pohon keputusan yang paling banyak digunakan karena algoritma ini memiliki keunggulan yang besar dibandingkan dengan algoritma lainnya. Kelebihan dari algoritma C4.5 adalah algoritma ini mudah untuk diinterpretasikan dan dapat membuat pohon keputusan dengan tingkat akurasi yang dapat diterima, sehingga dapat menangani atribut diskrit secara efisien, serta dapat menangani atribut diskrit dan numerik (Han et al., 2012).

Beberapa penelitian terdahulu yang menggunakan algoritma C4.5 dalam memprediksi atau mengklasifikasikan status gizi balita, antara lain penelitian yang dilakukan oleh Islam et al., (2022) yang dalam penelitiannya menyebutkan bahwa dengan algoritma C4.5 dapat diklasifikasikan status gizi balita ke dalam enam kategori berdasarkan indeks BB/TB, yaitu gizi normal, gizi kurang, buruk, lebih, dan obesitas. Model yang terbentuk memudahkan petugas Posyandu dalam menentukan status gizi balita (Islam et al., 2022). Penelitian lainnya dilakukan oleh Zami et al., (2022), hasil penelitian yang telah dilakukan ditarik kesimpulan bahwa akurasi yang dihasilkan oleh algoritma *decision tree* dalam memprediksi gizi balita sebesar 98,86% (Zami et al., 2022). Hutasoit et al., (2018) dalam penelitiannya menggunakan algoritma C4.5 untuk mengklasifikasikan status gizi balita. Dari penelitiannya tersebut, implementasi algoritma C4.5 memberikan

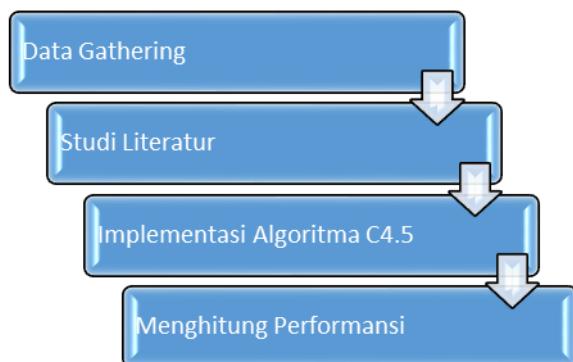
Performansi Algoritma C4.5 untuk Prediksi Gizi pada Balita

akurasi yang baik (Hutasoit et al., 2018). Algoritma C4.5 juga digunakan dalam pengklasifikasian status gizi balita pada Posyandu Desa Dames Damai Kabupaten Lombok Timur oleh Mahfuz et al., (2022). Dalam penelitiannya diperoleh hasil bahwa algoritma C4.5 mempunyai kecepatan yang tinggi dalam mengklasifikasikan status gizi balita dan hasil akurasi yang diperoleh juga sangat baik, yaitu 97,02% (Mahfuz et al., 2022).

Berdasarkan latar belakang tersebut di atas, maka dalam penelitian ini digunakan metode *data mining* dengan algoritma C4.5 untuk memprediksi status gizi balita, sehingga penanganan gizi buruk pada balita dapat tepat sasaran dan cepat untuk diatasi oleh Pemerintah.

Metode Penelitian

Penelitian ini dilakukan dengan beberapa tahapan. Adapun tahapan dari penelitian yang dilakukan ditunjukkan pada gambar 1 berikut ini:



Gambar 1. Tahapan Penelitian

Pada tahap awal penelitian dilakukan *data gathering* atau pengumpulan data. Jenis data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari *website Kaggle* (<https://www.kaggle.com/datasets>). *Dataset* yang digunakan adalah *dataset baby nutrition classification*. Tahapan kedua adalah studi literatur, dimana peneliti mencari bahan-bahan dari sumber penelitian sebelumnya melalui metode pengumpulan data, serta membaca literatur yang digunakan sebagai referensi penelitian yang dilakukan. Tahap ketiga adalah implementasi algoritma C4.5 pada *dataset*. Adapun tahapan dari algoritma C4.5, yaitu pemilihan atribut sebagai akar sesuai dengan nilai *gain* tertinggi dari atribut-atribut yang digunakan. Rumus untuk menghitung *gain* sebagai berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (1)$$

Selanjutnya menghitung nilai entropi melalui persamaan:

$$Entropy(S) = \sum_{i=1}^n -p_i \times \log_2 p_i \quad (2)$$

Performansi Algoritma C4.5 untuk Prediksi Gizi pada Balita

Setelah menghitung entropi, langkah berikutnya adalah membuat cabang untuk setiap nilai di dalam akar tersebut, dilanjutkan dengan membagi atribut dalam cabang. Ulangi proses untuk setiap atribut sampai semua atribut pada cabang mempunyai *class* yang sama.

Untuk proses analisis data, peneliti menggunakan alat bantu (*tools*) *RapidMiner*. Tahap keempat atau terakhir dilakukan evaluasi terhadap performansi algoritma C4.5 yang digunakan pada penelitian ini. Pengukuran performansi menggunakan *confusion matrix*.

Hasil dan Pembahasan

Dataset yang digunakan dalam penelitian ini terdiri atas 200 *record data*. Berikut ini sampel dari *dataset* yang digunakan:

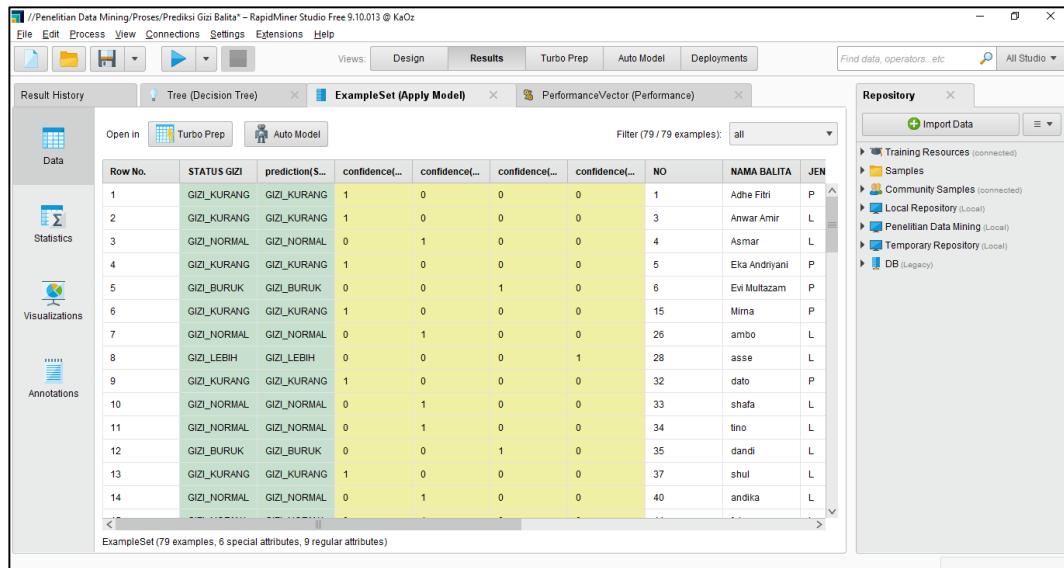
Tabel 1. Sampel *Dataset Baby Nutrition Classification*

No.	Nama Balita	Jenis Kelamin	Umur	Berat (Kg)	Tinggi (cm)	Tinggi (m)	Tinggi (m ²)	BMI	Status Gizi
1	Adhe Fitri	P	24	5,8	65	0,65	0,42	13,73	Gizi_Kurang
2	Andi Hariati	P	24	5,5	59	0,59	0,35	15,8	Gizi_Normal
3	Anwar Amir	L	28	6,7	71,5	0,715	0,51	13,11	Gizi_Kurang
4	Asmar	L	30	8,1	72,5	0,725	0,53	15,41	Gizi_Normal
5	Eka Andriyani	P	28	6,9	73	0,73	0,53	12,95	Gizi_Kurang
6	Evi Multazam	P	25	5,8	77	0,77	0,59	9,78	Gizi_Buruk
7	Firdaus Zubair	L	23	4,7	52,5	0,525	0,28	17,05	Gizi_Normal
8	Ilham Wahyudi	L	28	6,8	54	0,54	0,29	23,32	Gizi_Normal
9	Irmayaeni	P	21	3,5	55	0,55	0,3	11,57	Gizi_Kurang
10	Lieri Aprilyanti	P	27	5,8	53	0,53	0,28	20,65	Gizi_Normal
...
191	Gasali	P	23	7,3	61	0,61	0,37	19,62	Gizi_Normal
192	Amanda	L	33	5,6	62	0,62	0,38	14,57	Gizi_Kurang
193	Dita	L	25	5,5	62	0,62	0,38	14,31	Gizi_Kurang
194	Danis	P	34	7,8	63	0,63	0,4	19,65	Gizi_Normal
195	Inyong	L	35	9,8	63	0,63	0,4	24,69	Gizi_Normal
196	Danss	P	36	6,6	64	0,64	0,41	16,11	Gizi_Normal
197	Sama	L	38	9	76	0,76	0,58	15,58	Gizi_Normal
198	Mita	P	42	12	75	0,75	0,56	21,33	Gizi_Normal
199	Unta	L	44	11	74	0,74	0,55	20,09	Gizi_Normal
200	Fahmi	L	47	10	73	0,73	0,53	18,77	Gizi_Normal

Sumber: <https://www.kaggle.com/datasets>, 2022

Dataset dalam bentuk *file Excel* tersebut diimport ke *tools RapidMiner* untuk dilakukan proses pembagian antara *data training* dan *data testing* menggunakan operator *split data*. Pada penelitian ini, pembagian *data training* dan *data testing* dilakukan beberapa kali, yaitu 90%:10%, 80%:20%, 70%:30%, dan 60%:40%.

Performansi Algoritma C4.5 untuk Prediksi Gizi pada Balita

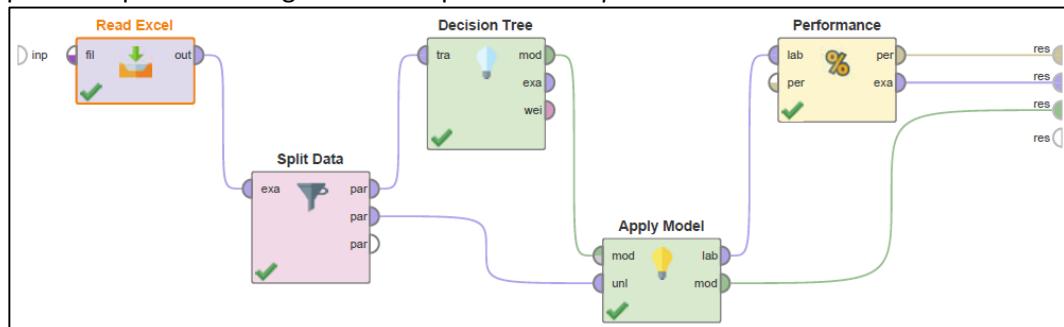


The screenshot shows the RapidMiner Studio interface. On the left, there are three panels: 'Statistics' (with a table of 14 rows), 'Visualizations' (empty), and 'Annotations' (empty). The main area displays a table titled 'ExampleSet (Apply Model)' with 79 rows. The columns include 'Row No.', 'STATUS GIZI', 'prediction(...)', 'confidence(...)', 'confidence(...)', 'confidence(...)', 'confidence(...)', 'NO', 'NAMA BALITA', and 'JEN'. The 'Repository' panel on the right shows 'Training Resources (connected)' with 'Community Samples (connected)', 'Local Repository (Local)', 'Penelitian Data Mining (Local)', 'Temporary Repository (Local)', and 'DB (Legacy)'.

Gambar 1. Dataset yang Diimport pada RapidMiner

Sumber: Data Sekunder Diolah, 2023

Setelah dilakukan pembagian *dataset* menjadi *data training* dan *data testing*, implementasi algoritma C4.5 dilakukan. Untuk mengukur performansi algoritma yang digunakan dalam penelitian ini, yaitu C4.5 digunakan metode *confusion matrix*. Dimana pada tabel *confusion matrix* yang dihasilkan akan terlihat nilai akurasi dari algoritma C4.5 yang diimplementasikan pada *dataset baby nutrition classification*. Berikut ini merupakan proses implementasi algoritma C4.5 pada *tools RapidMiner*.



Gambar 2. Implementasi Algoritma C4.5 pada RapidMiner

Sumber: Data Sekunder Diolah, 2023

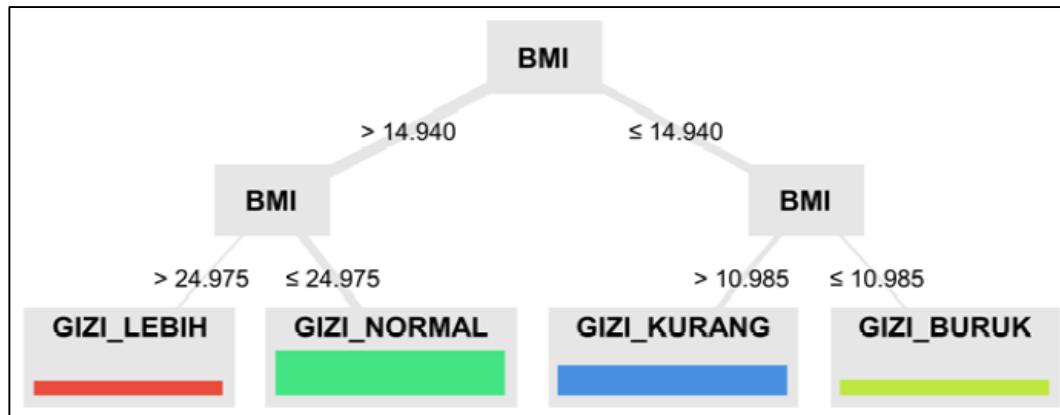
Dari beberapa kali eksperimen dengan mengubah persentase perbandingan *data training* dan *data testing*, diperoleh nilai akurasi sebagai berikut:

Tabel 2. Nilai Akurasi *Confusion Matrix*

Data Training : Data Testing	90% : 10%	80% : 20%	70% : 30%	60% : 40%
Nilai Akurasi	100%	100%	100%	97,47%

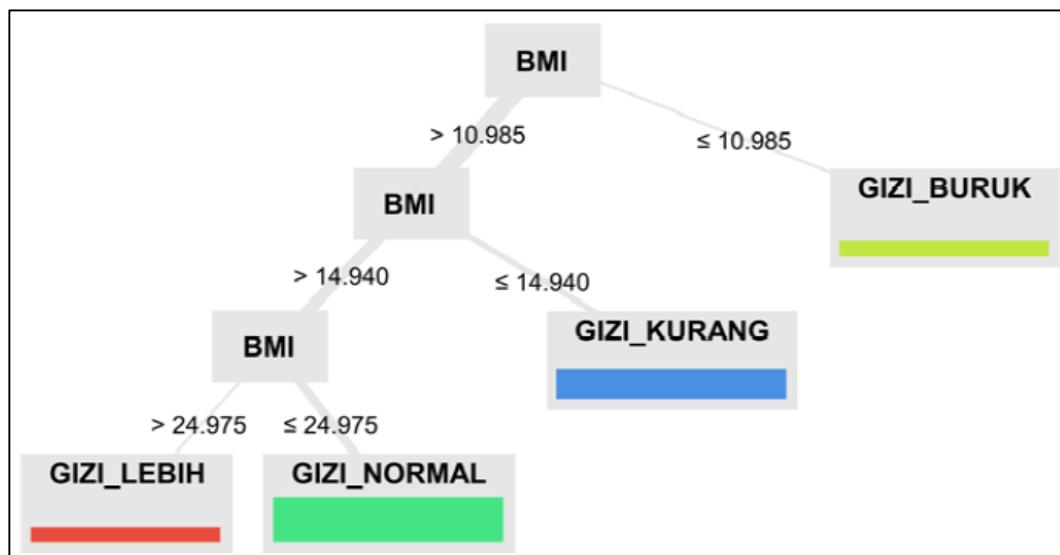
Performansi Algoritma C4.5 untuk Prediksi Gizi pada Balita

Model pohon keputusan atau *decision tree* yang dihasilkan dari masing-masing eksperimen ditunjukkan pada gambar-gambar berikut ini:



Gambar 3. Pohon Keputusan Hasil Eksperimen dengan Pembagian *Data Training* dan *Data Testing* 90%:10%

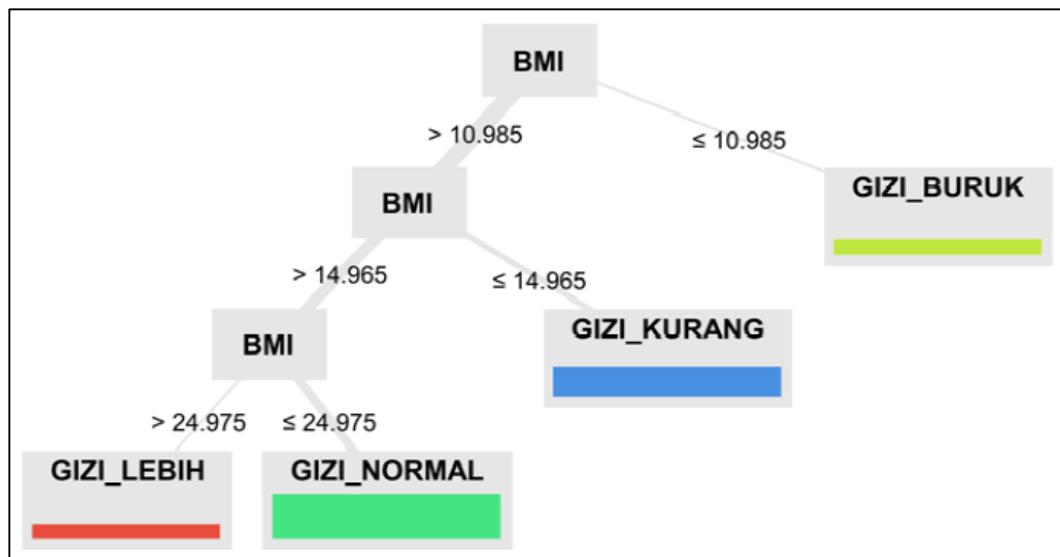
Sumber: Data Sekunder Diolah, 2023



Gambar 4. Pohon Keputusan Hasil Eksperimen dengan Pembagian *Data Training* dan *Data Testing* 80%:20%

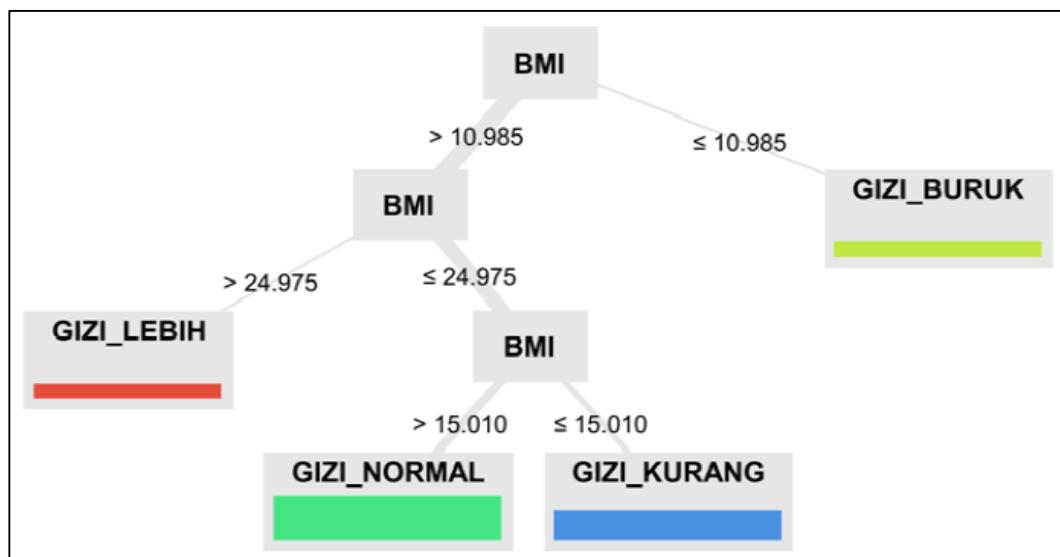
Sumber: Data Sekunder Diolah, 2023

Performansi Algoritma C4.5 untuk Prediksi Gizi pada Balita



Gambar 5. Pohon Keputusan Hasil Eksperimen dengan Pembagian *Data Training* dan *Data Testing* 70%:30%

Sumber: Data Sekunder Diolah, 2023



Gambar 6. Pohon Keputusan Hasil Eksperimen dengan Pembagian *Data Training* dan *Data Testing* 60%:40%

Sumber: Data Sekunder Diolah, 2023

Simpulan

Dari hasil penelitian yang dilakukan, nilai akurasi yang diperoleh dari beberapa eksperimen pembagian *data training* dan *data testing* dapat dikatakan sangat bagus, yaitu dengan rata-rata 99,37%. Untuk penelitian selanjutnya dapat dicoba dengan

Performansi Algoritma C4.5 untuk Prediksi Gizi pada Balita

menggunakan algoritma klasifikasi lainnya, seperti *Naive Bayes*, *K-Nearest Neighbor*, *ID3*, *CART*, *Random Forest*, dan lain sebagainya.

Daftar Pustaka

- Allensworth, D. (2010). *Health Promotion Programs from Theory to Practice*. Jossey-Bass.
- Bappenas. (2018). *Pedoman Pelaksanaan Intervensi Penurunan Stunting Terintegrasi di Kabupaten/Kota*. Kementerian Perencanaan Pembangunan Nasional.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining, Concepts and Techniques* (Third Edit). Morgan Kaufmann Publishers.
- Hutasoit, A. S., Tarigan, P., & Siagian, E. R. (2018). Implementasi Data Mining Klasifikasi Status Gizi Balita pada Posyandu Medan Timur dengan Menggunakan Metode C4.5. *Jurnal Pelita Informatika Budi Darma*, 7(2), 120–125.
- Islam, H. I., Mulyadien, M. K., & Enri, U. (2022). Penerapan Algoritma C4.5 dalam Klasifikasi Status Gizi Balita. *Jurnal Ilmiah Wahana Pendidikan*, 8(10), 116–125.
- Kementerian Kesehatan RI. (2013). *Hasil Riset Kesehatan Dasar*.
- Mahfuz, Nur, A. M., & Samsu, L. M. (2022). Penerapan Algoritma C4.5 dalam Mengklasifikasi Status Gizi Balita pada Posyandu Desa Dames Damai Kabupaten Lombok Timur. *Infotek: Jurnal Informatika Dan Teknologi*, 5(1), 72–81.
- Marsita, M. (2018). *Implementasi Algoritma Decision Tree C4.5 untuk Mengidentifikasi Gizi Balita Berdasarkan Indeks Antropometri (Studi Kasus Posyandu Seruni)*. UIN Sunan Gunung Djati.
- Narulita, S., Prihati, Oktaga, A. T., & Widayantoro, A. E. (2023). Performansi Algoritma Clustering K-Means untuk Penentuan Status Malnutrisi pada Balita. *Jurnal Informasi, Sains, Dan Teknologi*, 6(1), 188–202. <https://isaintek.polinef.ac.id/index.php/isaintek/article/view/128>
- Riani, A., Susianto, Y., & Rahman, N. (2019). Implementasi Data Mining untuk Memprediksi Penyakit Jantung Menggunakan Metode Naive Bayes. *JINITA: Journal of Innovation Information Technologi and Application*, 1(1), 25–34.
- Wanto, A., Siregar, M. N. H., Windarto, A. P., Hartama, D., Ginantra, N. L. W. S. R., Napitupulu, D., Negara, E. S., Lubis, M. R., Dewi, S. V., & Prianto, C. (2020). *Data Mining: Algoritma dan Implementasi*. Yayasan Kita Menulis.
- Zami, A. Z., Nurdianwan, O., & Dwilestari, G. (2022). Klasifikasi Kondisi Gizi Bayi Bawah Lima Tahun pada Posyandu Melati dengan Menggunakan Algoritma Decision Tree. *JSON: Jurnal Sistem Komputer Dan Informatika*, 3(3), 305–310.