

Classification of Nutrition Problems for Indonesian Toddler With Decision Tree Algorithm C4.5

Nadhea Ovella Syaahsydy, Zamahsary Martha*, Nonong Amalita, dan Dina Fitria

Departemen Statistika, Universitas Negeri Padang, Kota Padang, Negara Indonesia

*Corresponding author: zamahsarymartha@fmipa.unp.ac.id

Submitted : 07 Agustus 2023

Revised : 11 Oktober 2023

Accepted : 16 November 2023

ABSTRACT

The health and economic sector presents several issues for the Indonesian government. Having excellent human resources is essential for Indonesia's development. The development of Indonesia is the key to improving the quality of life for its citizens, and a focus on this development can have a positive impact on the health and economy of the community. A healthy and educated generation is fundamental for the expected progress of this nation, as nutritional status is a significant factor affecting the quality of human resources. Nutritional problems can lead to serious consequences, such as abnormal physical growth, a decline in IQ quality, and even death. The objective of this research is to analyze the factors that influence the nutritional status of toddlers by classifying each variable using a decision tree. A decision tree is a flowchart resembling a branching tree structure. The C4.5 algorithm was utilized in this study. This algorithm can process both numeric and categorical data, handle missing attribute values, and generate easily interpretable rules. After conducting the analysis, it was found that the decision tree's results indicated that the attribute "Stunting < 20%" is a determining factor for acute chronic malnutrition issues in toddlers. There are 392 districts and cities in Indonesia where the prevalence of stunted toddler nutritional status is less than 20%. The model created using the C4.5 algorithm was evaluated using a confusion matrix, resulting in an accuracy of 99.8% and a kappa value close to 1. This indicates that the model is capable of accurately classifying toddler nutrition problems in Indonesia.

Keywords: Algorithm C4.5, Classification, Decision Tree, Nutrition Toddlers



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Indonesia menghadapi banyak masalah dibidang kesehatan dan ekonomi. Sebagai generasi penerus bangsa, kualitas Sumber Daya Manusia (SDM) merupakan syarat mutlak menuju pembangunan Indonesia. Pembangunan Indonesia menjadi poin utama agar meningkatnya kualitas hidup masyarakat, sehingga dengan adanya fokus dalam pembangunan Indonesia dapat memberikan dampak ke peningkatan kesehatan hingga ekonomi masyarakat. Menurut Pane dkk (2020) generasi yang sehat dan cerdas merupakan modal utama dalam melanjutkan pembangunan sesuai harapan bangsa, dimana salah satu faktor yang sangat berpengaruh pada kualitas SDM ialah status gizi. Gizi yang bermasalah dapat memberikan dampak yang serius, seperti pertumbuhan fisik yang tidak sesuai dengan usianya, penurunan kualitas IQ, hingga menyebabkan kematian.

Indonesia menduduki peringkat pertama jumlah penduduk kurang gizi di Asia Tenggara yakni 17,7 juta jiwa mengalami masalah gizi. Masalah gizi harus diperhatikan sejak dalam usia kandungan. Menurut Lasarudin dkk (2022) usia paling rentan mengalami masalah kesehatan dan gizi adalah usia bayi dan balita dimana usia bayi dan balita menjadi salah satu tolak ukur cerminan status gizi masyarakat. Upaya pemerintah dan masyarakat dalam mendeteksi status masalah gizi dengan hasil pemantauan status gizi balita di Posyandu dan instansi kesehatan resmi. Menurut hasil Studi Status Gizi Indonesia (SSGI) prevalensi status gizi balita sudah menurun di tahun 2021, namun balita yang menderita *underweight* memiliki prevalensi sebesar 17% pada 2021 dari 16,3% pada 2019, adapun balita yang menderita *stunting* memiliki prevalensi sebesar 24,4% pada 2021 dari 27,7% pada 2019, sedangkan balita yang menderita *wasting* memiliki prevalensi sebesar 7,1% pada 2021 dari 7,4% pada 2019 (Kemenkes, 2021).

Pendapat dari Jiawei Han dan Jian Pei (2023) bahwa klasifikasi data adalah proses dua langkah, yang terdiri dari langkah pembelajaran (dimana model klasifikasi dibangun) dan langkah klasifikasi (dimana model digunakan) untuk

mengklasifikasikan data baru. Banyak metode yang dapat digunakan untuk identifikasi status gizi pada balita, diantaranya metode Naïve Bayes, SVM, Random Forest, CART, dan Decision Tree. Salah satu metode klasifikasi yang digunakan adalah *decision tree* yang merupakan sebuah diagram alir mirip dengan struktur pohon, dimana setiap simpul menotasikan variabel yang diuji, cabangnya menunjukkan hasil dari variabel tes dan simpul daun mempresentasikan kelas-kelas tertentu (Hand dkk, 2001). Menurut Sifaunajah dkk (2022) Keistimewaan dari *decision tree* ini adalah visualisasi yang disajikan dalam bentuk pohon sehingga prosedur prediksinya dapat diamati dengan mudah. Algoritma yang digunakan pada penelitian ini adalah algoritma C4.5. Menurut Youn dkk (2006) Algoritma C4.5 dapat mengolah data numerik dan kategorik, serta dapat menangani nilai yang hilang, dan menghasilkan aturan-aturan yang mudah diinterpretasikan. Untuk menganalisis apa saja yang dapat mempengaruhi status gizi pada balita dapat dilakukan dengan cara mengklasifikasikan masing-masing kelas menggunakan pohon keputusan (*decision tree*).

Penelitian yang dilakukan oleh Ardiyanti (2018) dimana klasifikasi masalah gizi pada balita di Indonesia tepat diklasifikasikan menggunakan algoritma ID3 namun saran yang diberikan oleh Ardiyanti adalah menggunakan algoritma C4.5 untuk data yang bersifat numerik, maka dari itu pada analisis ini ingin mengetahui bagaimana klasifikasi masalah gizi pada balita di Indonesia menggunakan algoritma C4.5 untuk data numerik. Sehingga, hasil dari analisis ini dapat mengetahui variabel yang menjadi kriteria utama landasan seorang balita mengalami masalah gizi dan menjadi salah satu bahan pertimbangan bagi pemerintah dan masyarakat dalam penentuan masalah gizi pada balita.

II. METODE PENELITIAN

A. Sumber Data dan Teknik Analisis Data

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari buku saku yang diterbitkan oleh Kementerian Kesehatan Republik Indonesia mengenai SSGI Tahun 2021 yang tersebar di 514 Kabupaten/Kota. variabel target yang akan digunakan pada penelitian ini adalah Akut, Kronis dan Akut-Kronis dengan tiga variabel prediktor yakni *underweight*, *stunting*, dan *wasting*. Tabel 1 merupakan rincian dari variabel penelitian yang akan digunakan.

Tabel 1. Variabel Penelitian

Variabel	Istilah	Pengertian
<i>Y</i>	Akut (0)	Suatu wilayah mengalami masalah gizi akut
	Kronis (1)	apabila prevalensi balita <i>stunting</i> < 20% dan
	Akut-Kronis (2)	prevalensi balita <i>wasting</i> <5%
		Suatu wilayah mengalami masalah gizi kronis
		apabila prevalensi balita <i>stunting</i> 20% atau lebih
		dan prevalensi balita <i>wasting</i> <5%
		Suatu wilayah mengalami masalah gizi akut-
		kronis apabila prevalensi balita <i>stunting</i> 20% atau
		lebih dan prevalensi balita <i>wasting</i> 5% atau lebih
<i>X₁</i>	<i>Underweight</i>	Gabungan gizi buruk dan gizi kurang
<i>X₂</i>	<i>Stunting</i>	Gabungan sangat pendek dan pendek
<i>X₃</i>	<i>Wasting</i>	Gabungan sangat kurus dan kurus

langkah dalam membentuk *decision tree* menggunakan algoritma C4.5 sebagai berikut.

1. *Cleaning* data
Pada tahapan ini data yang tidak dibutuhkan saat melakukan perhitungan dan analisis dibersihkan. Selanjutnya data yang masih bersifat kategorik ditransformasikan atau diubah ke dalam bentuk factor.
2. Membagi data *training* 80% dan *testing* 20%
3. Menghitung nilai entropy

$$Entropy (D) = \sum_{i=1}^m -p_i \log_2 p_i$$

Dimana :

- m : Banyaknya kelas/kategori pada D
- p_i : Proporsi kasus D ke-i terhadap total himpunan kasus
- D : Himpunan kasus

4. Menghitung nilai gain info

$$Gain(A) = Entropy(D) - \sum_{j=1}^v \frac{D_j}{D} Entropy(D_j)$$

Dimana :

- v : Kelas/kategori atribut A
- D_j : Total atribut A dengan kelas ke-i
- D : Total himpunan kasus D
- $Entropy(D_j)$: *Entropy* untuk kelas ke-i

5. Menghitung nilai split info

$$Split\ Information_A(D) = - \sum_{j=1}^v \left| \frac{D_j}{D} \right| \times \log_2 \left| \frac{D_j}{D} \right|$$

Dimana :

- v : Total kelas atribut A dengan himpunan kasus D
- D_j : Total atribut A dengan kelas ke-i
- D : Total himpunan kasus D

6. Menghitung nilai gain ratio

$$Gain\ Ratio(A) = \frac{Gain\ Information(A)}{Split\ Information_A(D)}$$

7. Mengulang semua langkah 1 sampai 4 untuk setiap cabang hingga semua cabang memiliki daun keputusan
8. Membuat aturan berdasarkan keputusan decision tree

B. Evaluasi Ketepatan Hasil

Evaluasi ketepatan hasil klasifikasi dilakukan guna mengetahui seberapa besar sekumpulan dataset dapat dikategorikan dengan tepat pada suatu kelas. Metode yang digunakan untuk evaluasi ketepatan hasil klasifikasi adalah *confusion matrix* dengan memperkirakan hasil benar atau salah. *Confusion matrix* untuk klasifikasi *multi class* ditunjukkan pada Tabel 2.

Tabel 2. *Confusion Matrix*

n	Prediksi Akut (0)	Prediksi Kronis (1)	Prediksi Akut-Kronis (2)
Aktual Akut (0)	TP_0	FP_0	FP_0
Aktual Kronis (1)	FP_1	TP_1	FP_1
Aktual Akut-Kronis (2)	FP_2	FP_2	TP_2

Metode confusion matrix digunakan untuk menguji nilai *accuracy*, *error rate*, *true positive rate*, *false positive rate*, *specificity*, *precision*, *prevalence* dari algoritma C4.5. Untuk menghitung semua nilai-nilai kinerja berikut tabel rumus yang digunakan:

Tabel 3. Nilai Kinerja Klasifikasi

No	Nama	Rumus
1.	Accuracy Presentase klasifikasi benar melakukan prediksi	$\frac{TP_0 + TP_1 + TP_2}{n}$
3.	True Positive Rate Presentase data positif yang diprediksi positif	$\frac{TP}{TP + FN}$
4.	False Positive Rate Presentase data negatif diprediksi positif	$\frac{FP}{TN + FP}$

No	Nama	Rumus
5.	Specitiffity Presentase data negatif diprediksi negatif	$\frac{TN}{TN + FP}$
6.	Precision Presentase prediksi data sebagai positif yang benar	$\frac{TP}{FP + TP}$
7.	Prevalence Presentase jumlah instance positif pada data	$\frac{actual\ positive}{n}$

Terdapat istilah yang digunakan pada metode *confusion matrix* yaitu *True Positive* (TP) merupakan kasus dimana jumlah kelas diprediksi positif dan kelas aktualnya positif. *True Negative* (TN) merupakan kasus dimana jumlah kelas diprediksi negatif dan kelas aktualnya negatif. *False Positive* (FP) merupakan kasus dimana jumlah kelas diprediksi positif dan kelas aktualnya negatif. *False Negative* (FN) merupakan kasus dimana jumlah kelas diprediksi negatif dan kelas aktualnya positif.

III. HASIL DAN PEMBAHASAN

A. Cleaning Data

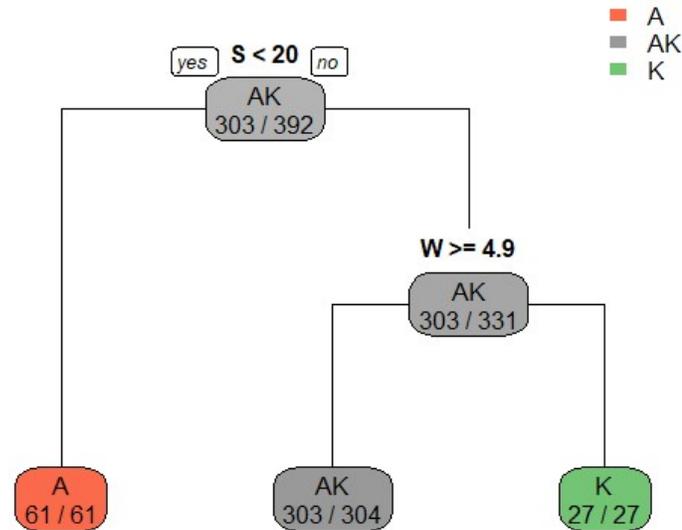
Cleaning data dilakukan untuk membuang data yang tidak diperlukan pada saat analisis. Pada data masalah gizi balita di Indonesia yang diterbitkan oleh SSGI kategori masalah gizi yang baik tidak digunakan pada analisis dan di hilangkan di dataset, adapun kabupaten/kota yang berkategori baik disajikan pada Tabel 4.

Tabel 4. Kabupaten/Kota Karakteristik Masalah Gizi Baik

Kabupaten/Kota	Status Gizi Balita			KMG
	U	S	W	
Kota Pematangsiantar	10,7	15,0	4,5	B
Kabupaten Siak	18,4	19,0	3,4	B
Kota Bogor	11,8	16,9	2,0	B
Kota Cimahi	11,2	19,9	3,8	B
Kabupaten Cilacap	9,7	17,9	3,9	B
Kabupaten Karanganyar	9,7	16,2	3,9	B
Kabupaten Purbalingga	10,4	16,8	3,3	B
Kabupaten Sragen	13,4	18,8	4,1	B
Kabupaten Wonogiri	10,9	14,0	4,2	B
Kabupaten Blitar	11,6	14,5	4,3	B
Kabupaten Kediri	11,4	18,0	4,9	B
Kabupaten Trenggalek	10,4	18,1	3,0	B
Kabupaten Tulungagung	11,8	13,1	4,3	B
Kota Blitar	9,6	12,9	4,4	B
Kota Jakarta Pusat	13,2	19,7	4,8	B
Kota Jakarta Selatan	11,8	15,7	4,8	B
Kabupaten Bangli	7,3	11,8	0,9	B
Kabupaten Buleleng	7,5	8,9	2,8	B
Kabupaten Gianyar	3,7	5,1	1,9	B
Kabupaten Jembrana	7,3	14,3	2,0	B
Kabupaten Tabanan	4,7	9,2	4,1	B
Kota Makassar	13,7	18,8	4,9	B
Kabupaten Minahasa	7,2	17,5	4,5	B
Kota Tomohon	6,0	18,3	2,4	B

B. Algoritma C4.5

Algoritma C4.5 menghitung dan mengklasifikasikan 490 data masalah gizi balita di Indonesia pada tahun 2021. Dimana data *training* sebanyak 392 data dan data *testing* sebanyak 98 data. Analisis algoritma C4.5 menggunakan software Rstudio. Dataset dihitung dengan cara memasukan data yang sudah di cleaning sebelumnya, setelah data dimasukkan ke Rstudio lalu algoritma C4.5 mulai mencari nilai sesuai dengan rumus pada persamaan 1 sampai persamaan 4. Nilai gain ratio tertinggi akan terpilih sebagai simpul akar, perhitungan terus diulang hingga semua simpul memiliki sebuah keputusan. Setelah mendapatkan keputusan menggunakan data *training* lalu algoritma langsung menguji keputusan tersebut pada data *testing* dan didapat hasil keputusan dan prediksi. Hasil dari analisis tersebut kemudian dievaluasi menggunakan metode *confusion matrix* seperti yang dijelaskan pada Tabel 1 dan Tabel 2. Hasil pohon keputusan menggunakan algoritma C4.5 seperti yang tertera pada Gambar 1.



Gambar 1. Hasil Pohon Keputusan

Hasil pohon keputusan menggunakan algoritma C4.5 menjelaskan bahwa variabel *Stunting* < 20% merupakan variabel yang menentukan masalah gizi balita Akut-Kronis. Hasil dari pohon keputusan terdiri dari satu *root node* sebagai kriteria utama, satu *internal node* dan tiga *leaf node* sebagai nilai kelas dari variabel yang terpilih. variabel yang berada pada *root node* merupakan variabel yang memiliki nilai gain ratio tertinggi yaitu variabel *Stunting*. variabel *Wasting* menempati *internal node* dan pada *leaf node* merupakan kategori masalah gizi yang terdiri dari Akut, Akut Kronis, dan Kronis.

Terdapat 392 kabupaten/kota di Indonesia dengan status gizi balita *Stunting* dengan kriteria prevalensi < 20%, dimana 303 kabupaten/kota tepat diklasifikasikan sebagai kategori Akut-Kronis, dan 61 kabupaten/kota tepat diklasifikasikan sebagai kategori masalah gizi Akut. Sisanya menjadi *internal node* dimana 331 kabupaten/kota dengan status gizi balita *Wasting* kriteria prevalensi >= 4.9%, dimana 303 kabupaten/kota tepat diklasifikasikan sebagai kategori Akut-Kronis dan 27 kabupaten/kota diklasifikasikan sebagai kategori Kronis.

C. Evaluasi Ketepatan Hasil Klasifikasi

Perhitungan ketepatan hasil klasifikasi menggunakan *confusion matrix*. Berikut adalah Tabel 5 yang merupakan hasil pengujian *confusion matrix* yang terdiri dari tiga kategori yaitu Akut, Akut-Kronis, dan Kronis.

Tabel 5. Hasil Perhitungan Confusion Matrix

n = 490	Prediksi Akut (0)	Prediksi Akut-Kronis (1)	Prediksi Kronis (2)
Aktual Akut (0)	77	0	0
Aktual Akut-Kronis (1)	0	378	0
Aktual Kronis (2)	0	1	34

Frekuensi aktual kategori Akut yang sesuai dengan prediksi kategori Akut berjumlah sebesar 77 kabupaten/kota, dan tidak ada frekuensi kategori aktual Akut yang tidak sesuai dengan prediksi Akut. Selanjutnya frekuensi aktual kategori Akut-Kronis yang sesuai dengan prediksi kategori Akut-Kronis berjumlah 378 kabupaten/kota dan tidak ada frekuensi aktual kategori Akut-Kronis yang tidak sesuai dengan prediksi kategori Akut-Kronis. Terakhir frekuensi aktual kategori Kronis yang sesuai dengan prediksi kategori Kronis berjumlah 34 kabupaten/kota, sedangkan ada frekuensi aktual kategori Kronis yang tidak sesuai dengan prediksi Kategori yakni Akut-Kronis yang berjumlah 1 kabupaten/kota. Adapun nilai evaluasi klasifikasi menggunakan *confusion matrix* yang dijelaskan pada Tabel 6.

Tabel 6. Nilai Evaluasi Klasifikasi Setiap Kategori

No	Nama	Akut	Akut-Kronis	Kronis
1.	Sensitivity	1.0000	0.9974	1.0000
2.	Specificity	1.0000	1.0000	0.9978
3.	Positive Prediksi Value	1.0000	1.0000	0.9714
4.	Negative Prediksi Value	1.0000	0.9911	1.0000
5.	Prevalence	0.1571	0.7735	0.0693
6.	Detection Rate	0.1571	0.7714	0.0693
7.	Detection Prevalence	0.1571	0.7714	0.0714
8.	Balanced Accuracy	1.0000	0.9987	0.9989

Nilai evaluasi klasifikasi setiap kategori menjelaskan bahwa model memiliki performa yang sangat baik dalam klasifikasi dimana nilai *sensitivity* untuk kategori Akut dan Kronis sebesar 100% artinya model mengklasifikasikan Akut dan Kronis dengan benar, sedangkan *sensitivity* untuk kelas Akut-Kronis sebesar 99.74% artinya sebagian besar model mengklasifikasikan Akut-Kronis dengan benar. Nilai *specificity* untuk kategori Akut dan Akut-Kronis sebesar 100% artinya model mengklasifikasikan negative dengan benar, dan untuk kategori Kronis sebagian besar model mengklasifikasikan negative dengan benar sebesar 99,78%. Nilai (*Positive Prediksi Value*) PPV dan (*Negative Prediksi Value*) NPV memiliki perbedaan pada masing-masing kategori, dimana PPV pada Akut dan Akut-Kronis menjelaskan model benar akurat ketika memprediksi kategori Akut dan Akut-Kronis sebesar 100%, dan PPV pada kategori Kronis sebesar 97.14%. nilai NPV kategori Akut dan Kronis memiliki model yang akurat ketika memprediksi negative sebesar 100%. Sedangkan NPV pada Kronis sebesar 100% sehingga model tepat mengklasifikasikan prediksi negative dengan benar untuk kategori Kronis.

Prevalence untuk kategori Akut sebesar 15.71%, untuk kategori Akut-Kronis sebesar 77.35%, dan untuk kategori Kronis sebesar 6.93% nilai prevalensi ini memberikan gambaran tentang sejauh mana suatu kondisi atau penyakit menyebar di antara populasi yang diuji. Nilai *balanced accuracy* pada kategori Akut 100% artinya akurasi model sangat tinggi dalam mengklasifikasikan kategori Akut dan kategori bukan Akut, disusul oleh Kronis sebesar 99.89% dan Akut Kronis sebesar 99.87% dimana nilai *balance akurasi* (akurasi seimbang) merupakan sebuah metrik yang mencoba mengukur keseimbangan antara sensitivitas dan spesifisitas. Adapun nilai secara keseluruhan yang dijelaskan pada Tabel 7.

Tabel 7. Nilai Evaluasi Klasifikasi

Overall Statistics	
Accuracy	: 0.998
95 % CI	: [0.9887, 0.9999]
No Information Rate	: 0.7735
P-value	: <22e-16
Kappa	: 0.9945

Tabel 7 menjelaskan bahwa model klasifikasi menggunakan algoritma C4.5 memiliki tingkat akurasi yang tinggi sebesar 99.8% dimana rentang nilai pada tingkat akurasi dengan kepercayaan 95% berada diantara 0.9887 sampai 0.9999 dengan No Information Rate sebesar 77.35 artinya tingkat akurasi yang akan dicapai model jika hanya memprediksi kategori mayoritas tanpa melakukan prediksi berdasarkan variabel. Nilai Kappa menjelaskan bahwa model memiliki tingkat kesepakatan tinggi dalam klasifikasi ditandai dengan nilai kappa mendekati 1.

IV. KESIMPULAN

Berdasarkan hasil dan pembahasan pada data SSGI mengenai masalah gizi balita di Indonesia tahun 2021 dengan menggunakan algoritma C4.5 dalam mengklasifikasikan data masalah gizi menyimpulkan bahwa variabel yang menjadi kriteria utama penentu masalah gizi balita Indonesia adalah *Stunting* kategori <20% artinya kriteria utama penentu suatu wilayah memiliki masalah gizi balita adalah *stunting* dengan prevalensi <20%. Model yang terbentuk dari algoritma C4.5 dievaluasi dan klasifikasi setiap kategori menjelaskan bahwa model memiliki performa yang sangat baik dalam klasifikasi dimana nilai *sensitivity* untuk kategori Akut dan Kronis sebesar 100% artinya model mengklasifikasikan Akut dan Kronis dengan benar, sedangkan *sensitivity* untuk kelas Akut-Kronis sebesar 99.74% artinya sebagian besar model mengklasifikasikan Akut-Kronis dengan benar. Nilai akurasi sebesar 99.8% dan nilai kappa yang mendekati 1 artinya model mampu mengklasifikasikan masalah gizi balita di Indonesia dengan tepat.

DAFTAR PUSTAKA

- Bukhori, A., dan Pratiwi, N. (2018). Implementasi Metode Decision Tree dengan Algoritma ID3 dan C4.5 untuk Mengklasifikasikan Partisipasi Perempuan Nikah dalam Kegiatan Ekonomi Rumah Tangga di DIY.
- Defiyanti, S. dan D. L. C. P. (2019). Perbandingan Kinerja Algoritma ID3 Dan C4.5 Dalam Klasifikasi Spam-Mail Sofi. Gunadarma University Repository.
- Faisal, M. R., Mangkurat, U. L., Nugrahadi, D. T., & Mangkurat, U. L. (2019). Belajar Data Science : Klasifikasi dengan Bahasa Pemrograman R.
- Han, J., dan Kamber, M. (2006). *Data Mining Concepts and Techniques Second Edition*. California: Morgan Kaufmann.
- Hand, D., Mannila, H. and Smyth, P.; 2001. *Principles of Data Mining*. MIT Press.
- Jiawei Han, Jian Pei, H. T. (2023). *Data Mining Concepts And Techniques (Beth LoGiudice (ed.); 4th ed.)*. Katey Birtcher.
- Kementrian Kesehatan (KEMENKES RI). (2021). Hasil Studi Status Gizi Indonesia (SSGI) Tingkat Nasional, Provinsi, dan Kabupaten/Kota.
- Lasarudin, A., Gani, H., dan Tomayahu, M. (2022). Perbandingan Metode Naïve Bayes dan C4 . 5 Klasifikasi Status Gizi Bayi Balita.
- Pane, H. W., Tasnim, Sulfiyanti, Puspita, H. R., Hastuti, P., Apriza, Sianturi, P. E., Rifai, A., & Hulu, V. T. (2022). Gizi Kesehatan (J. Simarmata (ed.); 1st ed., Vol. 1). Yayasan Kita Menulis.
- Sifaunajah, A., dan Wahyuningtyas, R. D. (2022). Penggunaan Algoritma ID3 untuk Klasifikasi Data Calon Peserta Didik. CSRID Journal.
- S. Youn, D. Mcleod, *A Comparative Study for Email Classification. Proceedings of International Joint Conferences on Computer, Information, System Sciences and Engineering, Bridgeport, CT, (2006)*.