

KLASIFIKASI PENGAMBILAN KEPUTUSAN TINDAKAN OPERASI SESAR MENGUNAKAN ALGORITMA CLASSIFICATION AND REGRESSION TREES

Nyayu Audy Fildzah Daniela*, Anita Desiani, Irmeilyana

Fakultas MIPA, Universitas Sriwijaya

E-mail Korespondensi : anita_desiani@unsri.ac.id

History Artikel

Diterima : 09 Juli 2024 Disetujui : 30 September 2024 Dipublikasikan : 26 Oktober 2024

Abstract

Data mining is a method that can be used to see patterns in data sets whose results can be used for decision making. One method of the data mining process is classification. To make data mining classification there is one algorithm that can be used, namely Classification and Regression Trees (CART). This study aims to classify the decision making of cesarean section using the Classification and Regression Trees (CART). Of the 80 cesarean operation data in the UCI dataset, the data was tested by dividing the data into training data and test data. The percentage of split dataset used includes 90% training data 10% test data, 80% training data 20% test data, 70% training data 30%, 60% training data 40% test data, 50% training data 50% test data, and 85% training data 15% test data . It was found that the implementation of the CART algorithm for the classification of the Caesarean dataset had the highest accuracy of 75%. These results indicate that the CART algorithm can be used to classify the decisions of cesarean section

Keywords: *Caesarean, Classification, Classification and Regression Trees*

Abstrak

Data mining adalah sebuah metode yang dapat digunakan untuk melihat pola pada kumpulan data yang hasilnya dapat digunakan untuk pengambilan keputusan. Salah satu metode dari proses data mining adalah klasifikasi. Untuk membuat klasifikasi data mining ada salah satu algoritma yang dapat digunakan yaitu Classification and Regression Trees (CART). Penelitian ini bertujuan untuk mengklasifikasikan pengambilan keputusan tindakan operasi sesar menggunakan algoritma Classification and Regression Trees (CART). Dari 80 data tindakan operasi sesar di dataset UCI dilakukan pengujian data dengan membagi data menjadi data latih dan data uji. Persentase split dataset yang digunakan antara lain 90% data latih 10% data uji, 80% data latih 20% data uji, 70% data latih 30%, 60% data latih 40% data uji, 50% data latih 50% data uji dan 85% data latih dan 15% data uji. Diperoleh hasil bahwa implementasi algoritma CART untuk klasifikasi dataset caesarean menghasilkan akurasi tertinggi 75%. Hasil tersebut menunjukkan bahwa algoritma CART dapat digunakan untuk klasifikasi pengambilan keputusan tindakan operasi sesar.

Kata Kunci: *Operasi sesar, Klasifikasi, Classification and Regression Trees.*

How to Cite: Nyayu Audy Fildzah Daniela., dkk (2024). Klasifikasi Pengambilan Keputusan Tindakan Operasi Sesar Menggunakan Algoritma Classification And Regression Trees. KOMPUTEK : Jurnal Teknik Universitas Muhammadiyah Ponorogo, Vol 8 (2): Halaman 54-63

© 2024 Universitas Muhammadiyah Ponorogo. All rights reserved

ISSN 2614-0985 (Print)

ISSN 2614-0977 (Online)

PENDAHULUAN

Operasi Caesar adalah prosedur pembedahan yang dilakukan dalam proses melahirkan untuk mencegah kematian ibu dan anak karena kondisi medis tertentu (Desyani et al., 2020). Jika berkaitan dengan nyawa maka pengambilan keputusan harus dilakukan dengan banyak pertimbangan dan penanganan yang cepat (Ismail, 2019). Kita tidak dapat memastikan bahwa semua dokter memiliki pengetahuan dan pengalaman yang sama, sehingga akan ada kemungkinan dapat membuat keputusan yang salah. Pengambilan keputusan yang salah dapat merugikan ibu, seperti infeksi, perdarahan, cedera viseral, dan gangguan kesehatan lainnya serta berpengaruh pada tindakan persalinan berikutnya (Desyani et al., 2020).

Data mining adalah sebuah metode yang dapat digunakan di sistem komputer untuk mengetahui pola dari sekumpulan data (Hasugian & Panjaitan, 2018). Data mining dapat digunakan untuk membantu pengambilan keputusan (Sivakami & Prabhu, 2020). Klasifikasi menjadi salah satu metode dalam data mining yang dapat membantu pengambilan keputusan (Prakoso & Tutik, 2017). Alur pengambilan keputusan dengan klasifikasi dilakukan dengan mengelompokkan data baru berdasarkan pola kemiripan antara data baru dengan kelompok data yang tersedia

(Oktanisa & Supianto, 2018). Oleh karena itu, pengambilan keputusan tentang apakah perlu dilakukan tindakan operasi sesar/tidak dapat menggunakan metode klasifikasi dalam data mining.

Klasifikasi dalam data mining dapat dibuat dengan suatu algoritma tertentu. Adapun algoritma yang dapat digunakan antara lain Naive bayes, KNN, SVM, Decision tree (C4.5, Id3, J48, CART) dan beberapa metode lainnya (Sumarno, 2017). Untuk penelitian ini akan menggunakan algoritma CART. Algoritma CART memiliki beberapa keunggulan seperti dapat digunakan untuk klasifikasi data numerik/ kategorik, menghasilkan pohon keputusan yang disesuaikan dengan target variabel, Jika variabel target berupa data kategorik maka outputnya adalah pohon klasifikasi. Jika variabel target berupa data numerik maka outputnya adalah pohon regresi, dan hasilnya lebih mudah diinterpretasikan (Suntoro, 2019). Dalam penelitian ini juga, hasil akurasi algoritma CART akan dibandingkan dengan hasil peneliti lain dengan algoritma selain CART. Hal ini dilakukan untuk menentukan algoritma yang paling baik digunakan untuk klasifikasi pengambilan keputusan tindakan operasi sesar.

METODE PENELITIAN

Pada penelitian ini, ada beberapa tahapan yaitu deskripsi data, praproses data

atau persiapan data, implementasi algoritma CART untuk dan evaluasi hasil.

(<https://archive.ics.uci.edu/ml/datasets/Caesarian+Section+Classification+Dataset>).

2.1 Deskripsi Data

Dataset yang digunakan pada penelitian ini diambil dari website

Dataset caesarean terdiri dari 5 atribut dengan total 80 data. Deskripsi data secara lengkap dapat dilihat pada tabel 1.

Tabel 1 Deskripsi Data Caesarean

No	Nama Atribut	Jenis	Kode Atribut	Keterangan
1.	Age	Int	22,26,28,27,32,36,33,23,20,29,25,37,24,18,30,40,31,19,21,35,17,38	Menerangkan usia ibu. Data hilang = 0.
2.	Delivery Number	Int	1,2,3,4	Menerangkan nomor kelahiran bayi Data hilang = 0 0= timely 1= premature 2= latecomer
3.	Delivery Time	Int	0,1,2	Data hilang = 0 0= low 1= normal 2= high
4.	Blood of Pressure	Int	0,1,2	Data hilang = 0 0= apt 1= inept
5.	Heart Problem	Int	0,1	Data hilang = 0 0= no 1= yes
6.	Caesarean	Int	0,1	Data hilang = 0

Berdasarkan tabel 1 diperoleh bahwa semua atribut bertipe integer, data sudah lengkap/tidak ada data hilang.

2.2 Praproses Data

Dataset dalam penelitian ini adalah dataset yang sudah siap digunakan. Dataset tidak perlu dilakukan normalisasi data. Langkah selanjutnya adalah membagi data menjadi data latih dan data uji serta dievaluasi pembagian data paling efektif untuk algoritma CART di penelitian ini. Adapun pembagian data yang dilakukan

dalam penelitian ini adalah membagi data menjadi 90% data latih 10% data uji, 80% data latih 20% data uji, 70% data latih 30% data uji, 60% data latih 40% data uji, 50% data latih 50% data uji, dan 85% data latih 15% data uji.

2.3 Implementasi Algoritma CART

Algoritma *CART* adalah suatu algoritma yang dapat menghasilkan pohon keputusan biner menggunakan data pelatihan dengan klasifikasi yang diketahui (Javed Mehedi Shamrat et al., 2022). Klasifikasi data

menggunakan *CART* dapat diketahui dengan menghitung nilai *IndexGini* dan *GiniGain* (Suntoro, 2019). Perhitungan *IndexGini* dilakukan untuk pembentukan cabang dan *GiniGain* untuk pembentukan node. Berikut langkah-langkah perhitungan manual algoritma *CART*

1. Siapkan *dataset*

2. Hitung nilai *IndexGini*

Perhitungan *IndexGini*. dapat dilihat pada persamaan (1) :

$$\text{IndexGini} = 1 - \sum_{i=1}^k p_i^2 \quad (1)$$

3. Hitung nilai *GiniGain*

Perhitungan *GiniGain* dapat dilihat pada persamaan (2) :

GiniGain =

$$\text{Gini}(A, S) \sum_{i=1}^n \frac{|S_i|}{|S|} \times \text{Gini}(S_i) \quad (2)$$

4. Membuat Node dan Cabang dari Nilai *GiniGain* Maksimal

5. Ulangi langkah (2) – langkah (4) hingga semua node terpartisi

Jika variabel target berupa data kategorik maka outputnya adalah pohon klasifikasi. Jika variabel target berupa data numerik maka outputnya adalah pohon regresi (Suntoro, 2019).

2.4 Evaluasi Hasil

Evaluasi data dilakukan untuk menentukan tingkat baik/buruknya kinerja algoritma pada klasifikasi data (Fitri et al., 2018). Pada penelitian ini, evaluasi data dilihat dari hasil confusion matrix. Confusion matrix adalah sebuah matriks yang menampilkan hasil kinerja dari

algoritma klasifikasi (Ardiansyah & Walim, 2018). Setiap baris dalam confusion matrix mewakili kelas yang sebenarnya, sedangkan setiap kolom mewakili kelas yang diprediksi (Muqorobin et al., 2020).

Dalam penelitian ini, konsep keterangan confusion matrix yaitu TP (Benar 1) artinya data sebenarnya diklasifikasikan 1 (dilakukan tindakan operasi sesar), sistem benar memprediksi data diklasifikasikan ke kelompok 1. TN (Benar 0) artinya data sebenarnya diklasifikasikan 0 (tidak dilakukan tindakan operasi sesar), sistem benar memprediksi data diklasifikasikan ke kelompok 0. FN (Salah 0) artinya data sebenarnya diklasifikasikan 0 (tidak dilakukan tindakan operasi sesar) tetapi sistem salah memprediksi, data diklasifikasikan ke kelompok 1. FP (Salah 1) artinya data sebenarnya diklasifikasikan 1 (dilakukan tindakan operasi sesar) tetapi sistem salah memprediksi, data diklasifikasikan ke kelompok 0.

Tabel 2 Confusion Matrix

Kelas	Prediksi 0	Prediksi 1
Klasifikasi 0	TN	FP
Klasifikasi 1	FN	TP

Dari kumpulan data latih yang diterapkan pada model, confusion matrix akan menghasilkan beberapa nilai yaitu Akurasi, Presisi, dan Recall. Pada penelitian ini, nilai-nilai tersebut menjadi tolak ukur

evaluasi data. Berikut persamaan untuk menentukan nilai akurasi, presisi, dan recall.

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \times 100\% \quad (3)$$

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \quad (4)$$

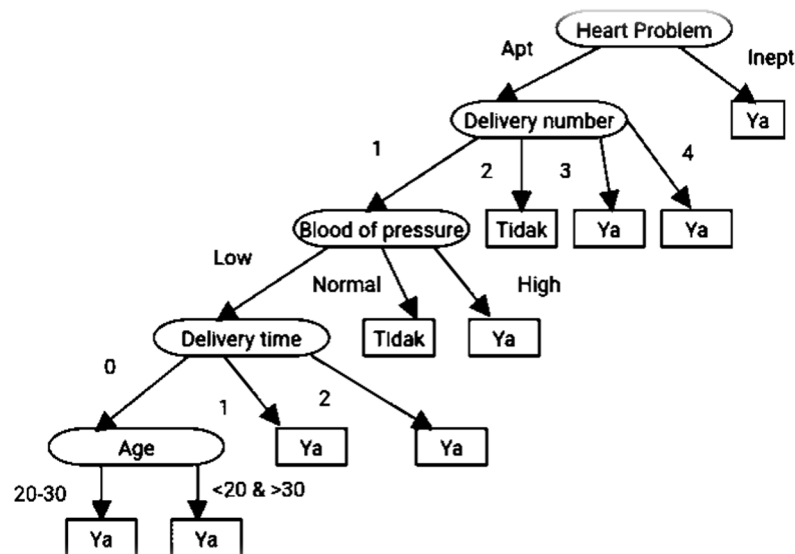
$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \quad (5)$$

HASIL DAN PEMBAHASAN

Berdasarkan deskripsi data, *dataset* yang dimiliki ialah *dataset* yang sudah siap/ tidak perlu dilakukan *preprocessing*.

Kemudian, sekumpulan data dalam dataset ini diklasifikasi dengan algoritma *CART*. Dari implementasi dengan *CART* diperoleh hasil bahwa nilai gini gain tertinggi hingga terendah secara berturut-turut adalah *Heart problem*, *Delivery number*, *Blood of pressure*, *Delivery time*, dan *Age*. Adapun pohon klasifikasi *CART* dapat dilihat pada gambar 1.

Berdasarkan gambar 1 diperoleh bahwa atribut paling berpengaruh dalam penentuan klasifikasi adalah atribut heart problem, sedangkan age tidak memberi pengaruh signifikan.



Gambar 1 Pohon klasifikasi CART

Kelompok-kelompok klasifikasinya adalah sebagai berikut:

- IF Heart problem (Inept) THEN Caesarean(ya/1).
- IF Heart problem (Apt) AND Delivery number (2) THEN Caesarean(tidak/0).

- IF Heart problem (Apt) AND Delivery number (3/4) THEN Caesarean(ya/1).
- IF Heart problem (Apt) AND Delivery number (1) AND Blood pressure (Normal) THEN Caesarean(tidak/0)

- IF Heart problem (Apt) AND Delivery number (1) AND Blood pressure (High) THEN Caesarean(ya/1)
- IF Heart problem (Apt) AND Delivery number (1) AND Blood pressure (Law) AND Delivery time (premature/latecomer) THEN Caesarean(ya/1)
- IF Heart problem (Apt) AND Delivery number (1) AND Blood pressure (Law) AND Delivery time (timely) And Age(20-30/ <20 dan >30) THEN Caesarean(ya/1)

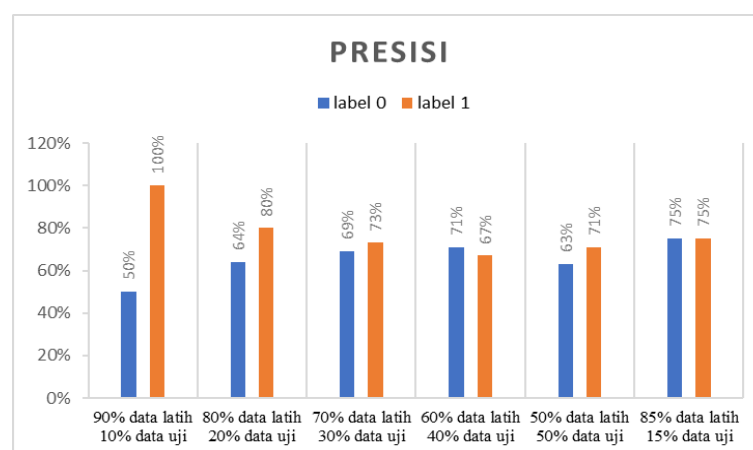
Kemudian dilakukan pengujian data yang hasilnya dapat dilihat pada tabel 3 berikut.

Tabel 3 Nilai Akurasi, Presisi, Recall

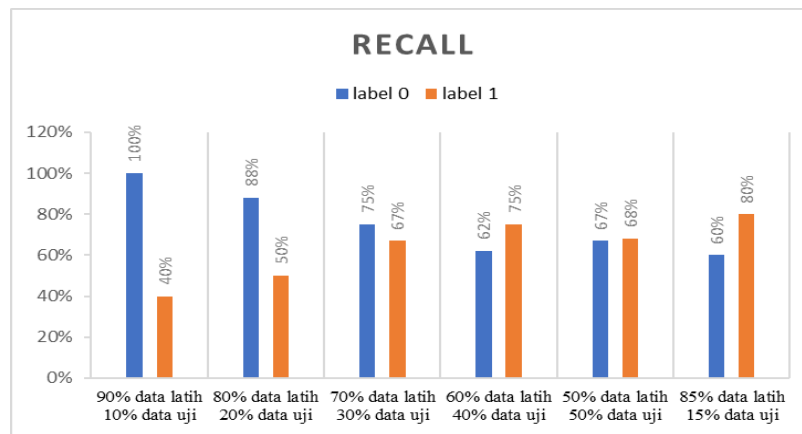
Pengujian Data	Akurasi	Presisi		Recall	
		0	1	0	1
90% data latih 10% data uji	62%	50%	100%	100%	40%
80% data latih 20% data uji	69%	64%	80%	88%	50%
70% data latih 30% data uji	71%	69%	73%	75%	67%
60% data latih 40% data uji	69%	71%	67%	62%	75%
50% data latih 50% data uji	68%	63%	71%	67%	68%
85% data latih 15% data uji	75%	75%	75%	60%	80%

Berdasarkan tabel 3, kita dapat melihat bahwa akurasi tertinggi diperoleh dari pengujian data dengan 85% data latih 15%

data uji. Akurasi terendah diperoleh dari pengujian data dengan 90% data latih 10% data uji.



Gambar 2 Nilai Presisi



Gambar 3 Nilai Recall

Berdasarkan gambar 2 dan gambar 3 diperoleh bahwa untuk nilai presisi terbaik yang bisa memprediksi klasifikasi tindakan caesarean ialah pada pengujian data dengan 85% data latih 15% data uji. Untuk nilai recall terbaik ialah pada pengujian data 70% data latih 30% data uji. Dari 5 pengujian data yang telah dilakukan, kita pilih hasil pengujian dengan rata rata akurasi, presisi, dan recall terbaik yaitu pengujian data dengan 85% data latih 15% data uji.

Tabel 4 Confusion Matrix

Label sebenarnya	Label prediksi	
	0	1
0	3	2
1	1	6

Adapun hasil confusion matrix dari pengujian tersebut dapat dilihat pada tabel 3.2. Dari 80 data diambil 15% data uji yaitu sebanyak 12 data. Berdasarkan tabel 4 kita peroleh nilai-nilai TN, TP, FN, FP berturut-turut adalah 3, 6, 1, 2. Artinya, dari 12 data, ada data diklasifikasikan dengan benar sebanyak 3 data dikategori 1 (dilakukan tindakan sesar) dan 6 data dikategori 0 (tidak dilakukan tindakan sesar). Data diklasifikasikan salah sebanyak 1 data dikategori 1 seharusnya 0 dan 2 data dikategori 0 seharusnya 1. Kemudian, hasil penelitian yang sudah diperoleh, dibandingkan dengan hasil penelitian lain sebagai berikut:

Tabel 5 Perbandingan hasil penelitian

Nama peneliti	Algoritma	Akurasi
Nyayu audy fildzah daniela (Alsharif et al., 2019)	CART	75%
	SVM	58.33%
	KNN	66,66%
	Decision tree	54,16%
	Naïve bayes	70,83%
	Random forest	54,16%
(Suciningsih et al., 2021)	C4.5	45%

Berdasarkan tabel 5, kita dapat melihat bahwa klasifikasi data menghasilkan

akurasi tertinggi sampai terendah berturut-turut menggunakan algoritma *CART*, *Naive*

bayes, KNN, SVM, Decision tree/Random forest, dan C4.5.

KESIMPULAN

Dari hasil penelitian diperoleh bahwa implementasi algoritma CART untuk klasifikasi dataset caesarean menghasilkan akurasi tertinggi 75%. Dengan akurasi tersebut maka algoritma CART termasuk baik digunakan untuk klasifikasi pengambilan keputusan tindakan operasi sesar. Selain itu, jika dibandingkan dengan algoritma Naive bayes, KNN, SVM, Decision tree/Random forest, dan C4.5, maka algoritma CART paling baik digunakan untuk klasifikasi pengambilan keputusan tindakan operasi sesar karena akurasi yang dihasilkan lebih tinggi.

Semoga penelitian ini dapat digunakan sebagai referensi dalam menyelesaikan permasalahan klasifikasi pengambilan keputusan tindakan operasi sesar dengan algoritma CART.

REFERENSI

- Alsharif, O. S. S., Elbayoudi, K. M., Aldrawi, A. A. S., & Akyol, K. (2019). Evaluation of Different Machine Learning Methods for Caesarean Data Classification. *International Journal of Information Engineering and Electronic Business*, 11(5), 19–23. <https://doi.org/10.5815/ijieeb.2019.05.03>
- Ardiansyah, D., & Walim, W. (2018). Algoritma c4.5 untuk klasifikasi calon peserta lomba cerdas cermat siswa smp dengan menggunakan aplikasi rapid miner. *Jurnal Inkofar*, 1(2), 5–12.
- Desyani, T., Saifudin, A., & Yulianti, Y. (2020). Feature Selection Based on Naive Bayes for Caesarean Section Prediction. *IOP Conference Series: Materials Science and Engineering*, 879(1). <https://doi.org/10.1088/1757-899X/879/1/012091>
- Fitri, S., Nurjanah, N., & Astuti, W. (2018). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa (Studi Kasus: Umtas). *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 9(1), 633–640. <https://doi.org/10.24176/simet.v9i1.2002>
- Hasugian, P. S., & Panjaitan, S. (2018). The Data Mining of Cell Phone Most Interested Using. *JURNAL INFOKUM*, Volume 7, No.1, Desember 2018, 7(1), 7–15.
- Ismail, M. (2019). Perlindungan Hukum bagi Dokter dalam Menangani Keadaan Medis Darurat Berdasarkan Implied Consent. *Islamadina : Jurnal Pemikiran Islam*, 53. <https://doi.org/10.30595/islamadina.v0i0.4380>
- Javed Mehedi Shamrat, F. M., Ranjan, R., Hasib, K. M., Yadav, A., & Siddique, A. H. (2022). Performance Evaluation Among ID3, C4.5, and CART Decision Tree Algorithm. *Lecture Notes in Networks and Systems*, 317(January), 127–142. https://doi.org/10.1007/978-981-16-5640-8_11
- Muqorobin, M., Kusrini, K., Rokhmah, S., & Muslihah, I. (2020). Estimation System For Late Payment Of School Tuition Fees. *International Journal of Computer and Information System (IJCIS)*, 1(1), 1–6. <https://doi.org/10.29040/ijcis.v1i1.5>
- Oktanisa, I., & Supianto, A. A. (2018). Perbandingan Teknik Klasifikasi Dalam Data Mining Untuk Bank a Comparison of Classification Techniques in Data Mining for. *Teknologi Informasi Dan Ilmu Komputer*, 5(5), 567–576. <https://doi.org/10.25126/jtiik20185958>

- Prakoso, S. A., & Tutik, E. T. (2017). Komparasi Algoritma C4.5 dengan Naive Bayes untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu di PTS “KZX.” *Journal of Chemical Information and Modeling*, 53(9), 1689–1699.
- Sivakami, M., & Prabhu, P. (2020). A Comparative Review of Recent Data Mining Techniques for Prediction of Cardiovascular Disease from Electronic Health Records. In D. J. Hemanth, S. Shakya, & Z. Baig (Eds.), *Intelligent Data Communication Technologies and Internet of Things* (pp. 477–484). Springer International Publishing.
- Suciningsih, I. G. A., Hidayat, M. A., & Hapsari, R. A. (2021). *Comparison analysis of naïve bayes and decision tree C4 . 5 for caesarean section prediction*. 46–52.
- Sumarno, H. (2017). Komparasi algoritma klasifikasi machine learning pada analisis sentimen film berbahasa Indonesia. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 1(1), 978–979. <https://jurnal.kaputama.ac.id/index.php/JTIK/article/view/141/156%0Ahttp://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/2647/991/%0Awww.ijtre.com%0Ahttps://doi.org/10.1080/01431161.2018.1433343%0Ahttp://ejournal-sl.undip.ac.id/index.php/gaussian%0A>
- Suntoro, J. (2019). *Data Mining: Algoritma dan Implementasi dengan Pemrograman PHP*. Elex Media Komputindo. <https://books.google.co.id/books?id=8NKdDwAAQBAJ>