

# Classification of Factors Affecting Preeclampsia in Pregnant Women at RSUP. Dr. M. Djamil Padang using the CART Algorithm

Aulia Yuswita, Dina Fitria\*, Dony Permana, dan Admi Salma  
Departemen Statistika, Universitas Negeri Padang, Kota Padang, Negara Indonesia  
\*Corresponding author: [dinafitria@fmipa.unp.ac.id](mailto:dinafitria@fmipa.unp.ac.id)

Submitted : 10 Februari 2025  
Revised : 18 Februari 2025  
Accepted : 28 Februari 2025

## ABSTRACT

*Preeclampsia is a pregnancy-specific condition characterized by hypertension and proteinuria that occurs after 20 weeks of gestation. This condition is caused by various factors that can influence its occurrence in pregnant women, such as age, parity, history of hypertension, obesity, and kidney disorders. This study aims to identify the risk factors influencing preeclampsia based on the diagnosis of 200 patients in 2022 at Dr. M. Djamil Central General Hospital Padang. Each variable was analyzed and classified using a decision tree method with the CART (Classification and Regression Tree) algorithm. The CART algorithm is binary in nature and capable of analyzing both categorical and continuous response variables, handling data with missing values, and producing an interpretable decision tree structure. The results of the study indicate that parity is the primary risk factor for preeclampsia. The model developed using the CART algorithm was evaluated with a confusion matrix, yielding an accuracy of 54%, a precision of 33.3% in classifying patients with mild preeclampsia (PER), and a recall of 23.8% in classifying patients with severe preeclampsia (PEB).*

**Keywords:** CART, Classification, preeclampsia, Decision tree, AKI



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

Angka Kematian ibu (AKI) digunakan untuk mengukur tingkat kesehatan perempuan. AKI mencakup kematian perempuan selama kehamilan atau dalam 42 hari setelah melahirkan, yang disebabkan oleh hal-hal terkait kehamilan, seperti komplikasi medis, bukan karena kecelakaan atau cedera (BPS,2023). AKI juga menjadi salah satu fokus dalam tujuan *Sustainable Development Goals* (SDGs), yang bertujuan untuk memastikan kehidupan sehat dan kesejahteraan bagi semua orang di segala usia (Arikah,2020). SDGs menargetkan penurunan AKI menjadi 70 per 100.000 kelahiran hidup pada tahun 2030. Preeklampsia merupakan salah satu penyakit hipertensi dalam kehamilan yang sering ditemui pada ibu hamil, dengan angka kejadian 2-8% kehamilan di seluruh dunia. Kejadian preeklampsia di Indonesia adalah 128.273 kasus per tahun atau berkisaran 5,3% dari seluruh ibu hamil (Alomedika, 2023).

Preeklampsia merupakan salah satu penyebab utama kematian ibu di Indonesia dengan prevalensi yang terus meningkat. WHO memperkirakan bahwa preeklampsia lebih sering terjadi di negara-negara berkembang, dengan prevalensi berkisaran antara 1,8% hingga 16,7% (Alomedika, 2023). Preeklampsia adalah kondisi spesifik yang terjadi selama kehamilan, ditandai dengan tekanan darah tinggi ( $\geq 140/90$ ), pembekakan tubuh (*edema*), dan adanya protein dalam urin (*proteinuria*). Kondisi ini biasanya muncul setelah usia kehamilan 20 minggu (trimester kedua). Penderita preeklampsia dibagi menjadi dua kategori, preeklampsia ringan (*mild preeclampsia*) dan preeklampsia berat (*severe preeclampsia*).

Banyak faktor yang memungkinkan berperan dalam terjadinya preeklampsia yang dimiliki ibu hamil. Penyakit preeklampsia dipengaruhi oleh faktor risiko. Faktor risiko dari preeklampsia menurut Katsiki dkk (2017) merupakan faktor maternal dan faktor kehamilan. Faktor maternal diantaranya usia, paritas, riwayat hipertensi, obesitas, dan gangguan ginjal. Adapun faktor kehamilan diantaranya kehamilan anggur (*molahilatidosa*), penumpukan cairan yang

berlebih antara dua atau lebih ruang tubuh janin (*hydrops fetalis*) dan kehamilan ganda. Faktor risiko tersebut seharusnya diperhatikan secara serius agar dampak dari AKI dapat menurun.

Klasifikasi merupakan pengolahan data dalam jumlah besar dengan algoritma tertentu sehingga hasilnya tepat dan akurat (Wibisono dan Fahrurrozi, 2020). Klasifikasi adalah metode analisis data yang digunakan untuk memprediksi kategori atau kelas dari suatu sampel. Beberapa metode klasifikasi yang umum digunakan antara lain pohon keputusan (*decision tree*), *naïve bayes*, jaringan syaraf tiruan, *k-nearest neighbour*, *support vector machine* dan lain-lain. Dalam klasifikasi, terdapat berbagai algoritma, seperti CART (*Classification and Regression Trees*), ID3 (*Iterative Dichotomiser*), C4.5, C5.0 dan lainnya. Penelitian ini menggunakan metode klasifikasi dengan algoritma CART.

Algoritma CART memiliki struktur pohon biner, dimana tiap simpul harus memiliki dua cabang. Algoritma CART dikembangkan oleh Leo Breiman, Jerome H. Freidman, Richard A. Olshen dan Charles J. Stone pada tahun 1984. algoritma CART dikenal sebagai *Binary Recursive Partitioning* karena pada setiap simpul data akan akan dibagi menjadi dua simpul anak dan pembagian ini akan terus dilakukan sampai tidak bisa lagi dibagi..

Algoritma CART bertujuan untuk mengelompokkan dan memisahkan data kedalam kelompok yang tepat berdasarkan kesamaan yang ada, sehingga dapat digunakan pada klasifikasi atau prediksi dengan akurat. Algoritma CART memiliki beberapa kelebihan, diantaranya tidak memerlukan asumsi distribusi pada variabel independen karena termasuk metode nonparametrik. Selain itu, CART memudahkan pengambilan keputusan pada data dengan sifat multivariabel dan menghasilkan output yang sederhana serta mudah diinterpretasikan. Oleh karena itu, untuk artikel ini dilakukan analisis apa faktor risiko utama penyakit preeklampsia dengan cara mengklasifikasikan kelas menggunakan pohon keputusan.

## II. METODE PENELITIAN

### A. Sumber Data dan Variabel Penelitian

Penelitian ini menggunakan data sekunder, yang diperoleh dari data rekap medis pasien preeklampsia yang dirawat di RSUP Dr. M. Djamil Padang Tahun 2022. Variabel digunakan yaitu usia ibu, paritas, obesitas, riwayat penyakit, pendidikan, status pekerjaan dan diagnosa preeklampsia (preeklampsia ringan dan preeklampsia berat).

### B. Algoritma CART

Algoritma CART dikembangkan oleh Leo Breiman, Jerome H. Freidman, Richard A. Olshen dan Charles J. Stone, yaitu metode statistik nonparametrik yang digunakan untuk analisis klasifikasi. Tujuan algoritma CART adalah untuk mengklasifikasikan objek kedalam atau lebih kelompok. Teknik CART dikenal dengan istilah *Binary Recursive Partitioning*, yaitu proses pembagian data yang terus dilakukan hingga tidak bisa lagi dibagi.

Berikut adalah langkah-langkah penerapan Algoritma CART.

#### 1. Pembentukan pohon klasifikasi

##### a. Penentuan Pemilah

Menurut Breiman dkk (1993) setiap pemilah bergantung pada nilai dari satu variabel independen. Pemilahan pemilah digunakan berdasarkan *indeks gini*. *Indeks gini* yang merupakan salah satu kriteria untuk menentukan titik pemecah terbaik.

$$i(t) = 1 - \sum_{i \neq j} p(i|t)p(j|t) \quad (1)$$

Dimana:

$p(i|t)$  : peluang  $i$  pada simpul  $t$

$p(j|t)$  : peluang  $j$  pada simpul  $t$

Probabilitas pengamatan yang masuk pada simpul kanan dan kiri dihitung sebagai berikut.

$$P_L = \frac{\text{calon simpul kiri}}{\text{data latih}} \quad (2)$$

$$P_R = \frac{\text{calon simpul kanan}}{\text{data latih}} \quad (3)$$

Untuk menentukan kriteria *goodness of split* ( $\phi(s, t)$ ) adalah berikut.

$$(\phi(s, t)) = \Delta i(s, t) = i(t) - P_L i(t_L) - P_R i(t_R) \quad (4)$$

Dimana:

$i(t)$  : fungsi keheterogenitas pada simpul  $t$

$i(t_L)$  : fungsi keheterogenitas pada simpul anak kiri

$i(t_R)$  : fungsi heterogenitas pada simpul anak kanan  
 $P_L$  : proporsi pengamatan menuju simpul kanan  
 $P_R$  : proporsi pengamatan menuju simpul kiri  
 Pemilah yang memberikan nilai *goodness of split* terbesar merupakan pemilahan utama.

- b. Penentuan simpul terminal  
 Penentuan simpul terminal yang diketahui jika jumlah data dalam simpul terminal tersebut minimal 5. Jika hal tersebut terpenuhi, maka tidak akan dipecah lebih lanjut.
- c. Penandaan label kelas  
 Proses pemberian label kelas pada simpul akhir dilakukan dengan memilah kelas yang paling banyak jumlahnya, sebagai berikut.

$$P(j_0|t) = \max_j p(j|t) = \max_j \frac{N_j(t)}{N(t)} \tag{5}$$

Dimana:

$p(j|t)$  : proporsi kelas  $j$  pada simpul  $t$   
 $N_j(t)$  : jumlah pengamatan kelas  $j$  pada *terminal node*  $t$   
 $N(t)$  : jumlah total amatan pada *terminal node*  $t$

2. Pemangkasan Pohon Klasifikasi

Pemangkasan pohon merupakan langkah penting dalam pembuatan pohon keputusan, terutama mencegah *overfitting* yang bisa terjadi ketika pohon keputusan terlalu besar dan terlalu rumit, namun jika pohon keputusan dibatasi dengan batas tertentu, hal ini bisa menyebabkan *underfitting*. Menurut Breiman dkk (1993), ukuran pohon yang ideal dapat diperoleh dengan memangkas pohon hingga mencapai ukuran *cost complexity minimum* pada persamaan berikut.

$$R_\alpha(T) = R(T) + \alpha|T| \tag{6}$$

Dimana:

$R(T)$  : proporsi kesalahan pada sub pohon (*resubstitution estimate*)  
 $\alpha$  : parameter kompleksitas (*complexity parameter*)  
 $|T|$  : ukuran banyak simpul terbanyak terminal pohon  $T$

3. Penentuan Pohon Klasifikasi Optimal

Semakin besar ukuran pohon maka semakin kompleks struktur data yang terbentuk. Oleh karena itu, perlu dilakukan pemangkasan untuk mendapatkan pohon optimal yang lebih sederhana namun tetap menghasilkan nilai prediksi yang akurat. Pohon optimal dihasilkan dari pemangkasan.

C. Ukuran Ketetapan Klasifikasi

Menurut Johnson dkk (2007) *confusion matrix* adalah table yang menunjukkan jumlah data uji yang salah dan benar diklasifikasikan. Jadi, *confusion matrix* adalah metode yang digunakan untuk mengukur seberapa akurat model klasifikasi. Nilai *confusion matrix* ditampilkan pada Tabel 1.

Table 1. Confusion Matrix

Classification		Predicted Class	
		Preeklampsia Berat	Preeklampsia Ringan
Actual Class	Preeklampsia Berat	True Positive (TP)	False Positive (FP)
	Preeklampsia Ringan	False Negative (FN)	True Negative (TN)

(Sumber yang diperoleh dari: Gorunescu, F. 2011)

Metode *confusion matrix* digunakan untuk menghitung akurasi, *precision* dan *recall* dari algoritma CART berdasarkan tabel *confusion matrix*.

a. Akurasi

Akurasi merupakan pengukuran keakuratan model dalam memprediksi data dengan benar dari total prediksi yang dilakukan. Akurasi diperoleh dari perbandingan jumlah amatan klasifikasi dengan jumlah keseluruhan amatan. Semakin tinggi presentase nilai akurasi, maka semakin besar juga kedekatan nilai prediksi dan nilai aktual.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (7)$$

b. Presisi

Presisi merupakan pengukuran keakuratan model dalam memprediksi kelas positif dari total prediksi positif yang dilakukan. Nilai presisi didapatkan dari perbandingan jumlah prediksi benar positif terhadap jumlah total amatan positif.

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (8)$$

c. Recall

Recall merupakan pengukuran dengan menemukan seberapa baik suatu model dalam mengidentifikasi kelas negatif. Nilai recall diperoleh dari perbandingan jumlah prediksi benar positif dengan jumlah total amatan positif.

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (9)$$

### III. HASIL DAN PEMBAHASAN

#### A. Statistika Deskriptif

Jumlah data yang digunakan adalah 200 sampel, terdiri dari 6 variabel faktor risiko penyakit preeklampsia dan 1 variabel sebagai diagnosa penyakit preeklampsia rawat inap pada RSUP Dr. M. Djamil Padang terdapat pada Table 2.

Table 2. Satatistik Deskriptif

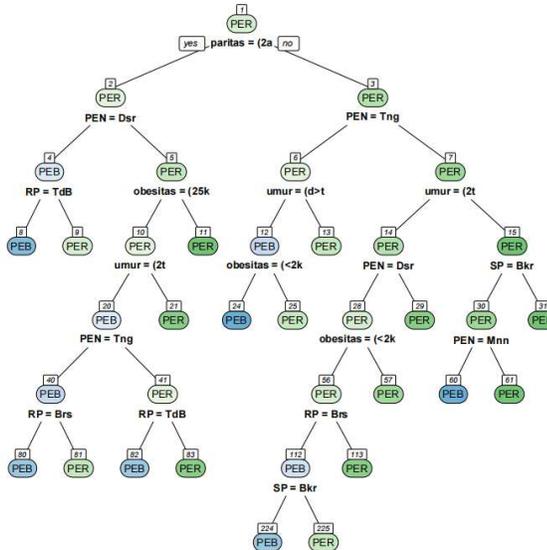
Variabel	Kategori	Jumlah
Umur ( $X_1$ )	(<20 dan >35 tahun)	52
	(20-35 tahun)	148
Paritas ( $X_2$ )	(1 dan >3 anak)	99
	(2-3 anak)	101
Obesitas ( $X_3$ )	( $\geq 25 \text{ kg/m}^2$ )	111
	(< $25 \text{ kg/m}^2$ )	89
Riwayat Penyakit ( $X_4$ )	Tidak Berisiko	80
	Berisiko	120
Pendidikan ( $X_5$ )	Dasar	125
	Menengah	35
	Tinggi	40
Status Pekerjaan ( $X_6$ )	Bekerja	131
	Tidak Bekerja	69
Diagnosa Preeklampsia ( $Y$ )	PEB	76
	PER	124

Berdasarkan Table 2, menunjukkan distribusi pasien preeklampsia berdasarkan variabel risiko seperti umur, paritas, obesitas, riwayat penyakit, pendidikan dan status pekerjaan, yang mana dari 200 pasien preeklampsia terdapat 62% atau 124 pasien terkena Preeklampsia Ringan dengan 148 dari 200 pasien preeklampsia berusia (20-35 tahun) yang dianggap berisiko. Mayoritas pasien yang terkena preeklampsia adalah ibu yang tidak berisiko paritas sebesar 101 pasien dan sebanyak 111 pasien berisiko obesitas dan sebanyak 120 pasien berisiko memiliki riwayat penyakit dan sebanyak 125 pasien mempunyai pendidikan yang dasar terhadap pengetahuan tentang penyakit preeklampsia dan sebanyak 131 pasien ibu hamil yang bekerja.

#### B. Pembentukan Pohon Klasifikasi

Data yang sudah disiapkan sebanyak 200 data penyakit preeklampsia, tahap berikutnya yaitu membagi data menjadi dua diantaranya data latih sebesar 150 dan data uji sebanyak 50. Algoritma CART dianalisis menggunakan *software RStudio*. Pembentukan pohon keputusan pada CART dilakukan dengan menghitung nilai *indeks gini* pada persamaan 1 sampai persamaan 3, dilanjutkan menghitung nilai *goodness of split* pada persamaan 4 tiap pemilahnya, yang mana pemilah dengan nilai tertinggi menjadi pemilah utama. Perhitungan pada persamaan 1 dan 4 akan terus berlangsung hingga semua dalam simpul menjadi homogen. Setelah itu, simpul akan diberi label berdasarkan kelas

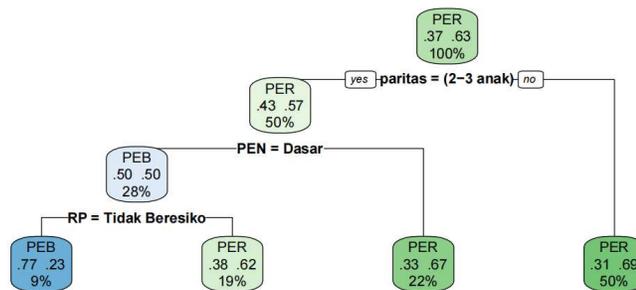
dengan jumlah data yang ada. Pada analisis CART, pembagian data dapat digunakan dengan membandingkan nilai latih dan nilai uji yaitu perbandingan 75%:25%. Berikut merupakan hasil pohon klasifikasi dari algoritma CART yang ditampilkan pada Gambar 1.



Gambar 1. Pohon Klasifikasi Algoritma CART

Pada Gambar 1, pohon klasifikasi algoritma CART menunjukkan bahwa paritas adalah variabel utama yang mempengaruhi penyakit preeklampsia rawat inap di RSUP Dr. M. Djamil Padang. Paritas memiliki nilai *indeks gini* tertinggi, sehingga menjadi *root node* (simpul akar). Pohon keputusan ini terdiri dari 1 simpul akar sebagai faktor risiko utama, 17 *internal node* (simpul dalam) sebagai variabel, dan 19 *terminal node* (simpul terminal). Variabel pada simpul akar memiliki nilai *Goodness of split* tertinggi seperti paritas yang dikategorikan tidak berisiko. Variabel lain seperti umur, obesitas, pendidikan, riwayat penyakit dan status pekerjaan berada pada *internal node*, sedangkan *terminal node* menunjukkan hasil diagnosa preeklampsia berat dan preeklampsia ringan.

Untuk mendapatkan pohon yang optimal, kita perlu memangkas pohon klasifikasi yang terlalu besar. Pemangkasan dilakukan pada pohon yang memiliki nilai parameter kompleksitas terendah dengan persamaan 6, diperoleh sebesar 0.0121. Setelah itu, pohon klasifikasi yang telah dipangkas berdasarkan nilai parameter kompleksitas terendah menunjukkan hasil yang optimal, sebagaimana disajikan pada Gambar 2.



Gambar 2. Pohon Complexity Parameter Minimum Optimal

Pada Gambar 2, pohon dengan parameter kompleksitas terendah yang optimal memiliki 1 *root node*, 2 simpul internal dan 4 simpul terminal. Berdasarkan pohon klasifikasi optimal, paritas merupakan faktor risiko utama dalam klasifikasi pasien preeklampsia di RSUP Dr. M. Djamil Padang. Pengujian akurasi dari pohon keputusan yang optimal dilakukan dengan menggunakan 50 data uji. Hasilnya dapat dilihat pada Table 3.

**Table 3.** Uji Klasifikasi CART

Klasifikas		Prediksi	
		Preeklampsia Berat	Preeklampsia Ringan
Aktual	Preeklampsia Berat	5	7
	Preeklampsia Ringan	16	22

Pada Table 3, hasil ketetapan klasifikasi menunjukkan bahwa sebanyak 5 pasien preeklampsia diprediksi mengalami preeklampsia berat (*recall*). Sementara itu, sebanyak 16 pasien preeklampsia di prediksi mengalami preeklampsia ringan (*presisi*). Perhitungan ini dilakukan berdasarkan persamaan 7, 8 dan 9.

$$Accuracy = \frac{5+22}{5+7+22+16} \times 100\% = 54\%$$

$$Precision = \frac{4}{5+7} \times 100\% = 33,3\%$$

$$Recall = \frac{5}{5+16} \times 100\% = 23,8\%$$

Hasil perhitungan yang diperoleh, nilai akurasi data prediski sebesar 54%, jadi, dapat disimpulkan bahwa pohon keputusan yang optimal berhasil mengklasifikasikan data sebesar 54%, nilai *precision* yang sebesar 33,3% untuk ketetapan klasifikasi pasien penyakit preeklampsia ringan dan nilai *recall* diperoleh untuk mengukur ketetapan klasifikasi pasien penyakit preeklampsia berat sebesar 23,8%.

#### IV. KESIMPULAN

Berdasarkan analisis klasifikasi dengan pohon keputusan algoritma CART pada data pasien penyakit preeklampsia di RSUP Dr. M. Djamil Padang, faktor risiko utama yang ditemukan adalah paritas dengan kategori tidak berisiko. Hasil klasifikasi menggunakan *confusion matrix* menunjukkan nilai akurasi sebesar 54%, presisi 33,3% untuk klasifikasi pasien preeklampsia ringan (PER), dan *recall* 23,8% untuk klasifikasi preeklampsia berat (PEB). Maka kesimpulan dari penelitian pada pasien preeklampsia yang merupakan faktor risiko utama preeklampsia adalah paritas. Dengan hasil penelitian tersebut dapat dilakukan oleh para tenaga medis dengan cara skrining awal, terutama pada ibu hamil dalam konsultasi kehamilan pertama dan memberikan intervensi dini untuk menekankan pentingnya pemeriksaan kehamilan secara rutin agar mengurangi risiko terjadinya preeklampsia.

#### DAFTAR PUSTAKA

- Arikah, T., Rahardjo, T. B. W., & Widodo, S. (2020). Kejadian Hipertensi pada Ibu Hamil. *Jurnal Penelitian Dan Pengembangan Kesehatan Masyarakat Indonesia*, 1(2), 115-124.
- Breiman, L., Friedman, J.H. Olsen, R.A. dan Stone, C.J. (1993). *Classification and regression tree*, New York: Chapman and Hall.
- CW, F. D., Emilia, R., & Indrayatna, F. (2023, August). Klasifikasi Tingkat Pencemaran Udara Kota Jakarta Tahun 2021 Menggunakan Algoritma Decision Tree. In *Prosiding Seminar Nasional Statistika Aktuaria* (Vol. 2, pp. 127-131).
- Fadillah, N. N., Syafrandi, S., Amalita, N., & Permama, D. (2024). Classification of Unemployment at West Sumatra Province in 2021 using Algorithm Classification and Regression Tree. *UNP Journal of Statistics and Data Science*, 2(2), 179-186.
- Luthffia, A. (2023). Epidemiologi Preeklampsia. *Alomedika.com*. Diakses pada 14 April 2025, dari [\[https://reset.alomedika.com/penyakit/obstetrik-dan-ginekologi/preeklampsia/perubahan-sistem-organ-dan-organ-pada-pre-eklampsia\]](https://reset.alomedika.com/penyakit/obstetrik-dan-ginekologi/preeklampsia/perubahan-sistem-organ-dan-organ-pada-pre-eklampsia)
- Pratiwi, D. (2020). Faktor Maternal yang mempengaruhi kejadian preeklampsia pada kehamilan. *Jurnal Medika Hutama*, 2(01), 402-6.
- Qadrini, L., Seppewali, A., & Aina, A. (2021). Decision tree dan adaboost pada klasifikasi penerima program bantuan sosial. *Jurnal inovasi penelitian*, 2(7), 1959-1966.

- Ramadhon, R. N., Ogi, A., Agung, A. P., Putra, R., Febrihartina, S. S., & Firdaus, U. (2024). Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak Aktif pada Data Bank. *Karimah Tauhid*, 3(2), 1860-1874.
- Rahmayuni, I. (2014). Perbandingan performansi algoritma c4. 5 dan cart dalam klasifikasi data nilai mahasiswa prodi teknik komputer politeknik negeri padang. *Jurnal TeknoIf*, 2(1).
- Rina. (2023). Memahami Confusion Matrix: Accuracy, Precision, Recall, Specificity, dan F1-Score untuk Evaluasi Model Klasifikasi. *Medium.com*. Diakses pada 27 Januari 2025, dari [<https://esairina.medium.com/memahami-confusion-matrix-accuracy-precision-recall-specificity-dan-f1-score-610d4f0db7cf>]
- Statistik, B. P. (2023). Hasil long form sensus penduduk 2020. *Jakarta (ID): BPS*.
- World Health Organization. (2023). *Trends in maternal mortality 2000 to 2020: estimates by WHO, UNICEF, UNFPA, World Bank Group and UNDESA/Population Division*. World Health Organization.