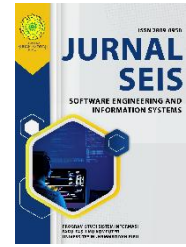




e-ISSN: 2809-0950



Analisis Pengaruh Ketidakseimbangan Data terhadap Kinerja Model Klasifikasi Penyakit Jantung

Maulana Muhammad Jogo Samodro^{1*)}, Eralangga Mandala Sakti²⁾, Bagas Mukti Kumudasmoro³⁾,
Muhammad Fahmi Mubarak Nahdli⁴⁾

^{1,2,3}Fakultas Ilmu Komputer, Universitas Safin Pati, Fakultas Teknik dan Ilmu Komputer, ⁴Universitas
Pancasakti Tegal

email: jogo_samodro@gmail.com, eralangga_mandala@usp.ac.id, bagas_mukti@usp.ac.id,
muhhammadfahmimubarak@upstegal.ac.id

Abstract

Heart disease is one of the leading causes of mortality worldwide, highlighting the need for data-driven predictive models to support risk analysis. However, medical data modeling often faces challenges such as class imbalance and weak predictive signals. This study aims to evaluate the performance of a Logistic Regression model in predicting heart attack events by comparing imbalanced and balanced datasets, as well as different training-testing data split ratios of 80:20 and 90:10. Model performance was assessed using accuracy, precision, recall, F1-score, and confusion matrix metrics. The results show that models trained on imbalanced data achieved higher accuracy values (0.69 for 80:20 and 0.70 for 90:10), but exhibited biased performance between classes, particularly low recall for the minority class (0.50 and 0.51, respectively). After data balancing, overall accuracy decreased; however, the model demonstrated more balanced performance with improved recall and F1-score for the minority class (recall of 0.64 and F1-scores of 0.69 for 80:20, and 0.66 for 90:10). These findings indicate that accuracy alone does not objectively reflect model performance on imbalanced datasets. Therefore, appropriate evaluation metrics and representative datasets are crucial in developing reliable heart disease risk prediction models.

Keywords: logistic regression, heart disease, imbalanced data, classification

Abstrak

Penyakit jantung menjadi salah satu penyebab kematian, sehingga diperlukan model prediksi berbasis data medis untuk mendukung analisis risiko. Pemodelan data medis sering menghadapi permasalahan distribusi kelas yang tidak seimbang dan lemahnya sinyal prediktif. Penelitian ini dilakukan untuk mengevaluasi kinerja model *Logistic Regression* dalam memprediksi kejadian serangan jantung dengan membandingkan penggunaan data tidak seimbang dan data seimbang, serta variasi proporsi data latih dan data uji sebesar 80:20 dan 90:10. Kinerja model kemudian dievaluasi berdasarkan nilai *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*. Hasil penelitian menunjukkan bahwa pada data tidak seimbang, model menghasilkan nilai *accuracy* lebih tinggi yaitu: 80:20 adalah 0,69%; 90:10 adalah 0,70% namun menunjukkan ketimpangan performa antar kelas, khususnya rendahnya *recall* pada kelas minoritas yaitu, 80:20 adalah 0,50; dan 90:10 adalah 0,51. Setelah dilakukan penyeimbangan data, nilai *accuracy* menurun, tetapi performa model menjadi lebih seimbang dengan peningkatan nilai *recall* dan *F1-score* kelas minoritas yaitu 80:20 *recall* dan *F1-score* adalah 0,64 dan 0,69; 90:10 *recall* dan *F1-score* adalah 0,64 dan 0,66. Temuan ini menunjukkan bahwa *accuracy* tidak selalu merepresentasikan kinerja model secara objektif pada dataset tidak seimbang. Berdasarkan hasil tersebut, penggunaan metrik evaluasi yang tepat dan dataset yang representatif sangat penting dalam pengembangan model prediksi risiko penyakit jantung.

Keywords: logistic regression, penyakit jantung, data tidak seimbang, klasifikasi

PENDAHULUAN

Jantung merupakan organ vital bagi makhluk hidup. Organ ini memiliki fungsi sebagai pompa

dalam mengalirkan darah ke seluruh jaringan tubuh manusia (Triyono et al., 2022). Serangan pada organ ini dapat mengakibatkan aliran darah tidak efektif, yang dapat membuat kualitas hidup serta

ketergantungan pada obat-obatan dan dapat mengancam jiwa (Bonek-Wytrych et al., 2024)(Ghani et al., 2016). Berdasarkan data WHO pada tahun 2024, terdapat 17,9 juta orang meninggal akibat dari penyakit jantung ini. Angka tersebut berada pada 32% penyebab kematian secara global di dunia (WHO, 2025).

Situasi yang sama berada di Indonesia. Berdasarkan data Kementerian Kesehatan dan Riset Kesehatan Dasar estimasi angka kematian mencapai 651,481 orang yang disebabkan oleh penyakit jantung (Indonesia, 2023). Berdasarkan data-data tersebut, penyakit ini menjadi salah satu penyebab utama kematian dan menjadi masalah kesehatan global yang memerlukan perhatian serius. Prevalensi penyakit jantung terus mengalami meningkat dengan adanya pola makan, perubahan gaya hidup, dan faktor risiko metabolik seperti hipertensi, diabetes, serta dislipidemia (Kemenkes, 2023). Oleh karena itu, upaya deteksi dini dan analisis faktor risiko penyakit jantung menjadi sangat penting untuk menekan angka morbiditas dan mortalitas (Anwar, 2025).

Perkembangan *machine learning* telah membuka peluang baru dalam bidang kesehatan (Jogo Samodro et al., 2024)(Fadlil et al., 2025)(Jogo et al., 2023), khususnya dalam pengolahan data medis untuk mendukung pengambilan keputusan klinis (Nasution et al., 2025). Algoritma klasifikasi seperti Logistic Regression, telah banyak digunakan untuk memprediksi risiko penyakit jantung berdasarkan data klinis dan demografis (Chen, 2025)(Nasution et al., 2025)(Ath et al., 2022). Beberapa penelitian melaporkan tingkat akurasi yang tinggi ketika menggunakan dataset medis yang kaya akan informasi klinis dan memiliki hubungan kausal yang kuat antara fitur dan label (Zhang et al., n.d.)(Wan, 2025).

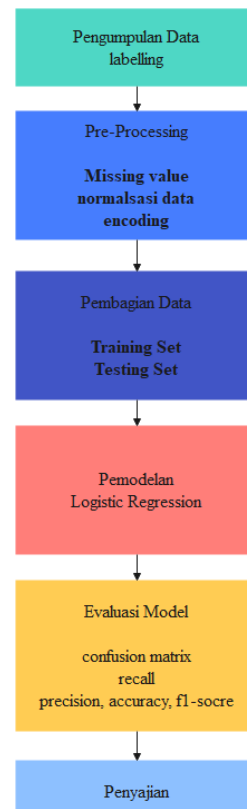
Beberapa penelitian sebelumnya telah menunjukkan bahwa penerapan metode pembelajaran mesin, termasuk Logistic Regression, mampu memberikan kinerja yang cukup baik dalam memprediksi penyakit jantung (Okolie et al., 2025). Fokus sebagian besar penelitian tersebut pada peningkatan nilai *accuracy* tanpa memperhatikan dampak distribusi kelas yang tidak seimbang terhadap performa model, khususnya pada kelas minoritas yang secara klinis sangat penting. Beberapa penelitian telah mencoba mengatasi permasalahan ketidakseimbangan data dengan menerapkan teknik penyeimbangan seperti SMOTE, yang terbukti meningkatkan nilai recall dan F1-score pada kelas minoritas (Aryuni et al., 2023). Selain itu, studi-studi terdahulu umumnya menerapkan satu skema pembagian data latih dan data uji, sehingga belum memberikan gambaran yang komprehensif mengenai pengaruh variasi rasio data terhadap stabilitas dan objektivitas kinerja model. Oleh

karena itu, penelitian ini dilakukan untuk mengisi celah tersebut dengan mengevaluasi kinerja Logistic Regression pada kondisi data tidak seimbang dan data seimbang, serta membandingkan rasio pembagian data latih dan data uji sebesar 80:20 dan 90:10 menggunakan berbagai metrik evaluasi, termasuk *accuracy*, *recall*, dan *F1-score*.

Berdasarkan permasalahan tersebut, tujuan penelitian ini untuk menganalisis kinerja model klasifikasi penyakit jantung dengan pendekatan *machine learning* yang menekankan pada evaluasi yang objektif dan realistis. Fokus penelitian ini tidak hanya pada pencapaian akurasi yang tinggi, tetapi juga pada pemahaman keterbatasan dataset serta implikasi statistik dari proses *preprocessing* dan penyeimbangan data terhadap performa model.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen berbasis *machine learning*. Tahapan penelitian meliputi pengumpulan data, *preprocessing*, pemodelan, serta evaluasi kinerja model klasifikasi. *Flowchart* penelitian ini termuat tertera pada Gambar 1.



Gambar 1. *Flowchart*

1. Pengumpulan data

Tahap pengumpulan data merupakan langkah awal dalam penelitian ini. Data penelitian berasal dari dataset penyakit jantung yang berisikan demografi dan klinis pasien, seperti usia, jenis

kelamin, tekanan darah, status merokok, kadar kolesterol, diabetes, serta riwayat jantung. Dataset merupakan data sekunder yang diperoleh dalam bentuk *.csv. Data ini merepresentasikan kondisi pasien yang beragam.

2. Pre-processing

Tahap *pre-processing* bertujuan untuk meningkatkan kualitas data sebelum digunakan dalam pemodelan. Proses yang dilakukan meliputi:

2.1 Pembersihan Data

Menangani *missing values* dan memastikan tidak terdapat data duplikat atau tidak konsisten.

2.2 Transformasi Data Kategorikal

Variabel kategorikal seperti status merokok (*never, past, yes*) dan jenis kelamin (*male, female*) untuk dapat diproses algoritma *machine learning* diubah ke dalam bentuk numerik menggunakan teknik *ordinal encoding* atau *label mapping*.

2.3 Normalisasi Data Numerik

Atribut numerik seperti tekanan darah, kolesterol, dan gula darah dinormalisasi menggunakan metode *StandardScaler* untuk menyamakan skala data, khususnya karena algoritma Logistic Regression sensitif terhadap perbedaan skala.

2.4 Penyeimbangan Data (Data Balancing)

Data pada dataset tidak memiliki keseimbangan dalam distribusi kelas pada variabel target. Teknik penyeimbangan data dilakukan untuk mengurangi bias model terhadap kelas mayoritas pada target.

3. Pembagian Data

Setelah tahap pre-processing, data set dibagi menjadi dua bagian, yaitu data latih dan data uji. Data dibagi menjadi 80:20 dan 90:10. Pembagian data dilakukan menggunakan teknik *stratified sampling* untuk memastikan proporsi kelas pada data latih dan data uji tetap seimbang. Data latih digunakan untuk membangun model, sedangkan data uji digunakan untuk mengevaluasi kinerja model secara objektif.

4. Pemodelan

Tahap pemodelan dilakukan dengan menerapkan algoritma *machine learning* untuk mengklasifikasikan penyakit jantung. Model yang digunakan adalah Logistic Regression.

5. Evaluasi Model

Kinerja dan keandalan model klasifikasi diukur menggunakan evaluasi model. Tahapan ini untuk mengetahui tingkat ketepatan,

keseimbangan dalam klasifikasi. Metrik evaluasi yang digunakan meliputi:

5.1 **Accuracy**, untuk melihat tingkat ketepatan prediksi secara keseluruhan

5.2 **Precision**, untuk mengukur ketepatan prediksi kelas positif

5.3 **Recall (Sensitivity)**, untuk menilai kemampuan model dalam mendeteksi pasien dengan penyakit jantung

5.4 **F1-Score**, sebagai keseimbangan antara precision dan recall

Evaluasi tidak hanya berfokus pada nilai akurasi, tetapi juga pada kemampuan model dalam mendeteksi kelas minoritas, yang penting dalam konteks medis. Berikut adalah rumus *accuracy, precision, recall, dan f1-score* (Jogo et al., 2023).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1 - Score = 2 \times \frac{Precision \times recall}{Precision+recall} \tag{4}$$

Keterangan:

TP (*True Positive*): Data positif yang diprediksi positif

TN (*True Negative*): Data negatif yang diprediksi negatif

FP (*False Positive*): Data negatif yang diprediksi positif

FN (*False Negative*): Data positif yang diprediksi negatif

6. Penyajian

Tahap penyajian dilakukan untuk menampilkan hasil penelitian dalam bentuk tabel evaluasi model. Tahapan ini dilakukan proses analisis berdasarkan data evaluasi yang diperoleh.

HASIL DAN PEMBAHASAN

1. Preprocessing

Dataset yang digunakan telah tahapan *preprocessing* untuk memastikan kualitas data sebelum dilakukan pemodelan. Tahapan tersebut meliputi pembersihan data, transformasi variabel kategorikal menjadi numerik, normalisasi atribut numerik, serta penanganan ketidakseimbangan kelas pada variabel target.

Hasil *preprocessing* menghasilkan dataset akhir dengan 14 atribut numerik, yang mencakup variabel demografis dan medis, seperti usia, tekanan darah sistolik dan diastolik, kadar kolesterol, kadar gula darah puasa, status merokok, hasil EKG, serta riwayat penyakit jantung sebelumnya. Tabel 1 merupakan atribut yang telah berada dalam format

numerik sehingga dapat digunakan secara langsung oleh algoritma *machine learning*.

Tabel 1. Atribut

Atribut	Nilai
Age	25-90
Gender	Male-female
Diabetes	0-1
obesity	0-1
Smoking_status	0-1
Blood_pressure_systolic	61-199
Blood_pressure_diastolic	37-127
Fasting_blood_sugar	70-230
Cholesterol_hdl	8-93
cholesterol_ldl	-19-282
triglycerides	50-380
EKG_result	0-1
Previous_heart_disease	0-1
Heart_attack	0-1

2. Distribusi Kelas

Analisis distribusi kelas pada variabel target `heart_attack` menunjukkan adanya ketidakseimbangan kelas, di mana jumlah data

pasien tanpa serangan jantung lebih dominan dibandingkan pasien dengan serangan jantung. Jumlah terkena serangan jantung sebanyak 63501 dan tidak terkena sebanyak 94854 dengan total keseluruhan data 158355. Kondisi ini berpotensi menyebabkan model bias terhadap kelas mayoritas dan menghasilkan nilai akurasi yang tidak merepresentasikan kemampuan prediksi yang sebenarnya.

Teknik penyeimbangan data dilakukan untuk mengatasi permasalahan perbedaan jumlah data dengan menggunakan metode *resampling*. Setelah proses penyeimbangan, jumlah data pada masing-masing kelas menjadi relatif seimbang, sehingga model dapat dilatih secara lebih adil terhadap kedua kelas.

3. Hasil Pemodelan

Algoritma logistic regression digunakan untuk menguji model klasifikasi. Hasil klasifikasi model tertera dalam Tabel 2.

Tabel 2. Hasil Klasifikasi

<i>train = 80%; testing = 20%</i>					
Logistic Regression Accuracy: 0.69					
	precision	recall	f1-score	support	
0	0.71	0.82	0.76		18971
1	0.66	0.50	0.57		12700
accuracy			0.69		31671
macro avg	0.68	0.66	0.67		31671
weighted avg	0.69	0.69	0.69		31671
<i>train = 90%; testing = 10%</i>					
Logistic Regression Accuracy: 0.70					
	precision	recall	f1-score	support	
0	0.72	0.82	0.77		9486
1	0.66	0.51	0.58		6350
accuracy			0.70		15836
macro avg	0.69	0.67	0.67		15836
weighted avg	0.69	0.70	0.69		15836
<i>balancing data, train = 90%; testing = 10%</i>					
Logistic Regression Accuracy: 0.67					
	precision	recall	f1-score	support	
0	0.66	0.71	0.68		6351
1	0.68	0.64	0.66		6350
Accuracy			0.67		12701

<i>macro avg</i>	0.67	0.67	0.67	12701
<i>weighted avg</i>	0.67	0.67	0.67	12701
<i>Balancing data, train = 80%; testing = 20%</i>				
<i>Logistic Regression Accuracy: 0.68</i>				
	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
0	0.66	0.71	0.69	12701
1	0.69	0.64	0.66	12700
<i>Accuracy</i>			0.68	25401
<i>macro avg</i>	0.68	0.68	0.68	25401
<i>weighted avg</i>	0.68	0.68	0.68	25401

Skenario pertama, model *Logistic Regression* dilatih menggunakan data yang tidak seimbang dengan perbandingan data latih dan data uji sebesar 80:20. Hasil pengujian menunjukkan nilai *accuracy* sebesar 0,69. Meskipun nilai tersebut tergolong cukup baik, evaluasi lebih lanjut terhadap metrik per kelas mengungkap adanya ketimpangan performa. Model memiliki nilai *recall* yang tinggi pada kelas 0 sebesar 0,82, namun nilai *recall* pada kelas 1 relatif rendah, yaitu 0,50. Temuan ini mengindikasikan bahwa model cenderung bias terhadap kelas mayoritas dan kurang optimal dalam mendeteksi kelas minoritas.

Eksperimen selanjutnya dilakukan dengan meningkatkan proporsi data latih menjadi 90% dan data uji menjadi 10%. Pada skenario ini, nilai *accuracy* model mengalami peningkatan marginal menjadi 0,70. Namun demikian, pola ketidakseimbangan performa masih tetap terjadi, di mana *recall* kelas 1 hanya meningkat menjadi 0,51. Hal ini menunjukkan bahwa peningkatan jumlah data latih pada kondisi data tidak seimbang tidak memberikan dampak signifikan terhadap kemampuan model dalam mengenali kelas minoritas.

Penyeimbangan data dilakukan untuk mengatasi permasalahan tersebut, sebelum tahap pelatihan model. Pada data seimbang dengan pembagian data latih dan data uji sebesar 90% dan 10%, diperoleh nilai *accuracy* sebesar 0,67. Meskipun nilai *accuracy* mengalami penurunan dibandingkan dengan skenario data tidak seimbang, model menunjukkan performa yang lebih seimbang antar kelas. Nilai *recall* untuk kelas 0 dan kelas 1 sebesar 0,71 dan 0,64, dengan nilai *macro F1-score* sebesar 0,67. Hasil ini mengindikasikan peningkatan kemampuan model dalam mengklasifikasikan kelas minoritas secara lebih proporsional.

Eksperimen terakhir dilakukan pada data seimbang dengan perbandingan data latih dan uji 80:20. Hasil pengujian menunjukkan nilai *accuracy*

sebesar 0,68, dengan nilai *recall* kelas 0 sebesar 0,71 dan kelas 1 sebesar 0,64. Konsistensi nilai *macro F1-score* sebesar 0,68 menunjukkan bahwa model memiliki stabilitas performa yang baik dan tidak menunjukkan kecenderungan bias terhadap salah satu kelas.

Berdasarkan keseluruhan hasil eksperimen, dapat disimpulkan bahwa penggunaan *accuracy* sebagai satu-satunya metrik evaluasi pada dataset yang tidak seimbang berpotensi memberikan gambaran performa yang menyesatkan. Meskipun data tidak seimbang menghasilkan nilai *accuracy* yang lebih tinggi, model cenderung gagal mengenali kelas minoritas dengan baik. Sebaliknya, penerapan teknik penyeimbangan data menghasilkan performa yang lebih adil dan representatif, khususnya dalam meningkatkan nilai *recall* dan *F1-score* pada kelas minoritas.

Setelah proses penyeimbangan data, model tidak lagi bergantung pada dominasi kelas mayoritas dan dipaksa untuk mempelajari pola yang lebih bermakna dari atribut medis yang tersedia. Kondisi ini menyebabkan penurunan nilai *accuracy*, namun mencerminkan kinerja model yang lebih realistis dan objektif dalam merepresentasikan kemampuan prediktif sebenarnya.

Hasil penelitian ini juga mengindikasikan bahwa hubungan antara atribut medis yang digunakan dengan kejadian serangan jantung pada dataset relatif lemah. Dengan kata lain, dataset memiliki rasio sinyal terhadap *noise* yang rendah, sehingga model mengalami kesulitan dalam membentuk batas keputusan yang jelas antara kedua kelas.

Fenomena serupa telah dilaporkan dalam penelitian sebelumnya, di mana dataset berbasis simulasi risiko atau label sintesis cenderung menghasilkan performa klasifikasi yang terbatas, bahkan ketika menggunakan algoritma non-linear seperti *Random Forest*. Hal ini menunjukkan bahwa

keterbatasan performa lebih dipengaruhi oleh kualitas dan kekuatan prediktif data dibandingkan dengan kompleksitas algoritma yang digunakan.

Dalam konteks dataset medis yang tidak seimbang dan memiliki sinyal prediktif yang lemah, evaluasi model tidak dapat hanya bergantung pada nilai *accuracy*. Nilai metrik evaluasi *precision*, *recall*, *F1-score*, serta analisis *confusion matrix* menjadi sangat penting untuk memperoleh gambaran kinerja model yang lebih komprehensif.

Hasil evaluasi menunjukkan bahwa meskipun nilai *accuracy* relatif rendah, model masih mampu mengidentifikasi sebagian pola risiko penyakit jantung. Namun demikian, performa tersebut belum memadai untuk digunakan sebagai alat diagnosis klinis, dan lebih sesuai untuk tujuan analisis risiko atau studi eksploratif.

Secara keseluruhan, penelitian ini menunjukkan bahwa peningkatan kualitas *preprocessing* dan penyeimbangan data tidak selalu menghasilkan peningkatan nilai *accuracy*. Sebaliknya, langkah-langkah tersebut justru mampu mengungkap keterbatasan informasi prediktif yang terkandung dalam dataset. Dengan demikian, hasil penelitian ini menegaskan pentingnya:

1. Evaluasi model menggunakan metrik yang sesuai dengan karakteristik data.
2. Kehati-hatian dalam menafsirkan nilai *accuracy* pada dataset yang tidak seimbang.
3. Pemilihan dataset klinis yang representatif dan berkualitas untuk pengembangan sistem prediksi medis.

KESIMPULAN

Penyakit jantung sebagai salah satu penyebab utama kematian memerlukan model prediksi berbasis data medis yang andal untuk mendukung analisis risiko. Hasil penelitian menunjukkan bahwa penggunaan data tidak seimbang menghasilkan nilai *accuracy* yang lebih tinggi, yaitu sebesar 0,69 pada rasio data latih dan uji 80:20 serta 0,70 pada rasio 90:10. Namun, performa model pada kondisi ini tidak merata, ditunjukkan oleh rendahnya nilai *recall* pada kelas minoritas, yaitu sebesar 0,50 untuk rasio 80:20 dan 0,51 untuk rasio 90:10. Setelah dilakukan penyeimbangan data, nilai *accuracy* mengalami penurunan, tetapi kinerja model menjadi lebih seimbang dengan peningkatan nilai *recall* kelas minoritas menjadi 0,64 pada kedua rasio pembagian data. Selain itu, nilai *F1-score* kelas minoritas juga meningkat menjadi 0,69 pada rasio 80:20 dan 0,66 pada rasio 90:10. Temuan ini menegaskan bahwa

accuracy tidak selalu mencerminkan kinerja model secara objektif pada dataset tidak seimbang. Oleh karena itu, penggunaan metrik evaluasi yang komprehensif dan pengelolaan data yang tepat sangat penting dalam pengembangan model prediksi risiko penyakit jantung.

TERIMA KASIH

Ucapan terima kasih diberikan kepada Universitas Safin Pati dalam memberikan dana penelitian internal. Rasa terimakasih disampaikan kepada semua pihak yang sudah membantu terlaksananya penelitian, sehingga peneliti dapat mencapai kesuksesan penelitiannya.

DAFTAR PUSTAKA

- Akter, Simon Bin, Akter, Sumya, Hasan, R., Hasan, M. M., Eisenberg, D., Azim, R., Fresneda Fernandez, J., & Pias, T. S. (2025). Optimizing stability of heart disease prediction across imbalanced learning with interpretable Grow Network. *Computer Methods and Programs in Biomedicine*, 265, 108702. <https://doi.org/10.1016/j.cmpb.2025.108702>
- Anwar, A. H. (2025). *SISTEMATIC REVIEW FAKTOR RESIKO PENYAKIT JANTUNG*. 02(01), 57–69.
- Aryuni, M., Adiarto, S., Miranda, E., Madyatmadja, E. D., Albert, V. D. S., & Sestomi, E. (2023). Imbalanced Learning in Heart Disease Categorization: Improving Minority Class Prediction Accuracy Using the SMOTE Algorithm. *INTERNATIONAL JOURNAL of FUZZY LOGIC and INTELLIGENT SYSTEMS*, 23(2), 140–151. <https://doi.org/10.5391/IJFIS.2023.23.2.140>
- Ath, S., Al, T., Darmawan, D., Fahmi, N., Hakim, A., Qibtiya, M. Al, & Syafei, N. S. (2022). *Jurnal Teknologi Terpadu HYBRID MACHINE LEARNING MODEL UNTUK MEMREDIKSI PENYAKIT JANTUNG DENGAN METODE LOGISTIC REGRESSION DAN RANDOM*. 8(1), 40–46.
- Bonek-Wytrych, G., Sierka, O., Szynal, M., & Dąbek, J. (2024). Quality of Life of Patients with Heart Failure Due to Myocardial Ischemia. *Reviews in Cardiovascular Medicine*, 25(9), 342. <https://doi.org/10.31083/j.rcm2509342>
- Chen, Z. (2025). *Heart Disease Prediction Models Performance Analysis based on Logistic Regression, Random Forest and XGBoost*. 153, 115–124.
- Fadlil, A., Perdana, L., Pujiyanta, A., Herman, H., Fathurrahman, H. I. K., & Samodro, M. M. J. (2025). Implementation of Dysarthria Identification Using MFCC and Multilayer

- Perceptron Algorithm. *SSRG International Journal of Electrical and Electronics Engineering*, 12(1), 32–46. <https://doi.org/10.14445/23488379/IJEEEE-V12I1P105>
- Ghani, L., Dewi, M., Novriani, H., Penelitian, P., & Daya, S. (2016). *Faktor Risiko Dominan Penyakit Jantung Koroner di Indonesia*. 153–164.
- Indonesia, K. K. R. (2023). *PROFIL KESEHATAN INDONESIA 2023*.
- Jogo, M. M. S., Biddinika, M. K., & Fadlil, A. (2023). Klasifikasi Penyakit Diabetes dengan Algoritma Decision Tree dan Naïve Bayes. *RESISTOR (Elektronika Kendali Telekomunikasi Tenaga Listrik Komputer) Vol.*, 6(2), 113–118.
- Jogo Samodro, M. M., Biddinika, M. K., & Fadlil, A. (2024). Optimal Feature Selection in Diabetes Classification Using the MLP Algorithm. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 18(2). <https://doi.org/10.22146/ijccs.94575>
- Kemenkes. (2023). *Surveki Kesehatan Indonesia*.
- Nasution, N., Nasution, F. B., & Hasan, M. A. (2025). *Predicting Heart Disease Using Machine Learning : An Evaluation of Logistic Regression , Random Forest , SVM , and KNN Models on the UCI Heart Disease Dataset*. 9(2), 140–150.
- Okolie, A., Obunadike, C., Okoro, S. C., Olufemi, I. B., Nwoke, P., & Akwabeng, P. M. (2025). Heart Disease Prediction: A Logistic Regression Approach. *Open Journal of Applied Sciences*, 15(11), 3534–3552. <https://doi.org/10.4236/ojapps.2025.1511229>
- Triyono, D., Liani, R., Utami, A. W., Tristiyanti, S., Supriatna, A., Surabaya, K. P., & Bandung, S. B. (2022). *PENYAKIT JANTUNG KORONER DI INDONESIA : PERAN*. 17(1), 86–94.
- Wan, S. (2025). *Machine learning approaches for cardiovascular disease prediction: A review*. (1).
- WHO. (2025). Cardiovascular diseases (CVDs). *World Health Organization*.
- Zhang, P., Wu, L., Zou, T., Zou, Z., Tu, J., & Gong, R. (n.d.). *Machine Learning for Early Prediction of Major Adverse Cardiovascular Events After First Percutaneous Coronary Intervention in Patients With Acute Myocardial Infarction : Retrospective Cohort Study Corresponding Author : 8*. <https://doi.org/10.2196/48487>