UNP JOURNAL OF STATISTICS AND DATA SCIENCE
Volume 3 Nomor 4 Tahun 2025 Halaman 430-436.
ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

Logit and Complementary Log-Log Modeling in the Case of Factors Affecting Heart Failure Disease

Iga Mawarni*, Asyifa Dwi Ayshah, Dhiyaa Fitri Yafe, Fadhilah Fitri
Departemen Statistsika, Universitas Negeri Padang, Padang, Indonesia
*Corresponding author: igamawarni3103@gmail.com

Submitted : 28 Agustus 2025
Revised : 26 Oktober 2025
Accepted : 11 November 2025

ABSTRACT

Heart failure is one of the leading causes of morbidity and mortality globally. Heart disease is a disease caused by plaque that builds up in the coronary arteries that supply oxygen to the heart muscle. Research on heart failure disease aims to find out what factors affect heart failure disease and how much influence it has. This test was conducted using logistic regression method with logit modeling and complementary log-log modeling in analyzing data of 918 patients with heart failure disease. This study also takes which modeling is the best. The results of this analysis indicate that Age, Gender, Blood Sugar, and Chest Pain have significant effects on the likelihood of Heart Failure. Specifically, higher blood sugar levels and the presence of chest pain were found to increase the probability of heart failure, while gender and age showed varying effects across different age groups. Based on the model comparison, the Logit model demonstrated better fit and predictive accuracy than the Complementary Log-Log model, as reflected by its lower AIC value 897.43.

Keywords: *Complementary Log-Log, Heart Failure Disease, Logit*

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

Dalam dunia medis, gagal jantung dikenal sebagai penyakit dengan angka mortalitas yang cukup tinggi (Raharja et al., 2024). Di Asia Tenggara pada tahun 2023, Indonesia termasuk negara dengan jumlah kematian tertinggi yang mana mencapai sekitar 17.9 juta orang meninggal karena gagal jantung, jumlah tersebut mewakili 32% dari semua kematian (Widadi et al., 2024). Salah satu penyebab penyakit jantung adalah terbentuknya plak dalam arteri koroner yang bertugas mengalirkan oksigen menuju otot jantung (Saraswati, 2024). Selain itu, kondisi gagal jantung juga dipengaruhi oleh berbagai faktor lain yang berhubungan dengan gaya hidup, seperti kebiasaan merokok, konsumsi alkohol, stress, dan lain-lain (Wala et al., 2024).

World Health Organisasion (WHO) telah menetapkan pengukuran kualitas hidup sebagai salah satu indikator penting dalam bidang kesehatan (Nursita & Pratiwi, 2025). Penting untuk mengenali faktor-faktor yang memicu Penyakit Gagal Jantung, karena pemahaman terhadap faktor tersebut akan memudahkan pelaksanaan strategi pencegahan. Di antara berbagai metode analisis, Regresi Logistik Biner dapat digunakan untuk menguji faktor-faktor yang berperan pada penyakit ini.

Regresi logistik biner adalah salah satu teknik regresi yang memodelkan keterkaitan antara variabel prediktor (Dewi & Pratiwi, 2021). Pada analisis ini, variabel respon bersifat biner atau kategorikal, biasanya terbagi ke dalam dua kategori, yaitu 0 dan 1 (ya atau tidak) (Tampil et al., 2017). Dalam penelitian dengan data kategorik, beberapa model yang sering digunakan meliputi regresi logistik biner, probit, dan complementary log-log (Nugraha & Fatimah, 2016). Complementary log-log adalah metode yang memodelkan hubungan antara variabel dikotomus yang ditentukan berdasarkan ambang batas tertentu dengan variabel penjelas (Nuradilla, 2024). Penelitian ini berfokus dalam mengidentifikasi faktor-faktor yang memengaruhi dan menilai besarnya pengaruh tersebut.

II. METODE PENELITIAN

A. Analisis Regresi Logistik Biner

Regresi logistik biner merupakan metode analisis dengan variabel dependen (y) berbentuk biner, yang hanya memiliki dua kemungkinan hasil yaitu berhasil atau gagal (McFadden, 2022). Sementara itu, variabel independennya (x) dapat berupa kategorik maupun kontinu (Amelia et al., 2020).

DOI: https://doi.org/10.24036/ujsds/vol3-iss4/421

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

Fungsi peluang logistik biner :

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p)} \quad (1)$$

Model logistik biner :

$$g(\boldsymbol{x}) = logit[\pi(\boldsymbol{x})] = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p \quad (2)$$

Keterangan :
$g(x)$: hasil transformasi dalam bentuk logit
$\pi(x)$: probabilitas terjadinya peristiwa "sukses" $[P(Y = 1|x)]$
x : variabel bebas, $x = (x_1, x_2, \dots, x_p)$
p : jumlah variabel bebas
β_0 : intercept
β_j : koefisien variabel bebas ke-j, $j = 1, 2, \dots p$

B. Analisis Regresi Complementari Log-Log

Regresi complementary log-log digunakan untuk menganalisis hubungan antara variabel respon dikotomus dengan variabel prediktor, baik bersifat kategorik maupun kontinu (Ratnasari & Putri, 2015).
Fungsi peluang complementari log – log :

$$\pi(x) = 1 - \exp\left[-\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p)\right] \quad (3)$$

Model complementari log-log :

$$g(\boldsymbol{x}) = log[-\log(1 - \pi(\boldsymbol{x}))] = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p \quad (4)$$

Keterangan :
$g(x)$: Hasil transformasi dalam bentuk cloglog
$\pi(x)$: Probabilitas terjadinya peristiwa "sukses" $[P(Y = 1|x)]$
x : Variabel bebas, $x = (x_1, x_2, \dots, x_p)$
p : jumlah variabel bebas
β_0 : Intercept
β_j : Koefisien variabel bebas, $j = 1, 2, \dots p$

C. Pengujian Parameter secara Simultan dan Parsial

Pengujian simultan merupakan evaluasi dampak koefisien logistik secara menyeluruh, yang kerap menggunakan uji rasio Likelihood atau uji G. Sementara itu, pengujian parsial untuk menilai pengaruh koefisien secara individual kerap menggunakan uji Wald (Hendayana, 2013).

Tabel 1. Uji Kebaikan Model dan Hipotesis

Uji	H_0	H_1	Kriteria Penolakan
Uji Simultan / Uji G	Model Tidak Sesuai	Model Sesuai	G > $X^2_{(alfa,\ k)}$
Uji Parsial / Uji W	Peubah ke-i Tidak Berpengaruh	Peubah ke-i Berpengaruh	W > $X^2_{(alfa,\ 1)}$

D. Uji Kesesuaian Model

Kesesuaian model diuji dengan menggunakan uji Hosmer dan Lemeshow (2000).
Hipotesis :
H_0 : Model sesuai
H_1 : Model tidak sesuai

©Department of Statistics, Universitas Negeri Padang
DOI: https://doi.org/10.24036/ujsds/vol3-iss4/421

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

Statistik Uji :

$$\hat{C} = \sum_{k=1}^{g} \frac{(O_k - n'_k\bar{\pi}_k)^2}{n'_k\bar{\pi}_k(1-\bar{\pi}_k)} \sim X^2_{(g-2)} \quad (7)$$

Dengan kriteria penolakan tolak H_0 apabila $\hat{C} > X^2_{(g-2)}$ atau *p-value* < α. Statistik uji $\hat{C}$ ini mengikuti distribusi *chi-square* (X^2) dengan derajat bebas $(g-2)$, dengan g adalah banyaknya kelompok/grup.

E. Ketepatan Klasifikasi

Tabel 2. Akurasi

observasi	prediksi	
	0	**1**
0	a	b
1	c	d

Tabel Klasifikasi berfungsi untuk menilai ketepatan klasifikasi model regresi logistik dengan memanfaatkan nilai akurasi. Sensitivitas mengukur kemampuan suatu pengujian untuk mengidentifikasi observasi yang berhasil dari semua observasi yang memang berhasil. Sementara itu, spesifisitas adalah kemampuan pengujian untuk menunjukkan observasi yang gagal dari seluruh observasi yang benar-benar gagal. Nilai akurasi diatas 0.5 (> 0.5) menandakan tingkat keakuratan yang cukup memadai (Lin et al., 2010).

$$\text{Akurasi} = \frac{\text{True Positive} + \text{True Negative}}{Total} \quad (8)$$

F. Jenis dan Sumber Data

Penelitian ini menggunakan data sekunder yang diambil dari platform Kaggle, data dapat dilihat melalui link : https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction. data tersebut berjumlah 12 Variabel dan 918 pengamatan, namun pada penelitian ini hanya menggunakan 5 variabel dari 12 variabel tersebut.

G. Variabel yang Digunakan

Variabel dependen adalah penyakit gagal jantung (1=mengalami, 0=tidak mengalami). Variabel independen meliputi Usia, Jenis Kelamin, Kadar Gula Darah dan Nyeri Dada.

Tabel 3. Variabel Penelitian

Variabel	Nama Variabel	Kategori	Kode
Y	Penyakit Gagal Jantung	Mengalami	1
		Tidak Mengalami	0
X_1	Usia	28 – 52 tahun	1
		53 – 77 tahun	0
X_2	Jenis Kelamin	Perempuan	1
		Laki – Laki	0
X_3	Gula Darah	>120 mg/dl	1
		≤120 mg/dl	0
X_4	Nyeri Dada	Mengalami	1
		Tidak Mengalami	0

III. HASIL DAN PEMBAHASAN

A. Pemodelan Data

Pemodelan data menggunakan dua model estimasi parameter, yaitu Model Logit dan Complementary Log-Log. Pemodelan ini dilakukan untuk mengetahui pengaruh variabel prediktor terhadap kejadian gagal jantung serta membandingkan kinerja kedua model.

©Department of Statistics, Universitas Negeri Padang
DOI: https://doi.org/10.24036/ujsds/vol3-iss4/421

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

Tabel 4. Koefisien Pemodelan

Pemodelan	Konstanta	X_1	X_2	X_3	X_4
Logit	-0.327	-0.696	-1.355	1.439	2.320
Std. Error Logit	0.1425	0.1682	0.2150	0.2132	0.1849
Complementary log-log	-0.625	-0.396	-0.0932	0.761	1.358
Std. Error Cloglog	0.0987	0.1063	0.1535	0.1220	0.1048

B. Uji Parameter

Uji Simultan dilakukan dengan Statistik uji G, sedangkan uji parsial dengan Statistik uji W.

Tabel 5. Uji Parameter

Peubah		Logit	Complementary log-log	P-value logit	p-value cloglog
Intercept	Wald	5.258	40.056	0.0218	2.46e-10
Umur	Wald	17.098	13.853	3.55e-05	0.000197
Jenis Kelamin	Wald	39.690	36.857	2.97e-10	1.27e-09
Gula Darah	Wald	45.535	38.863	1.50e-11	4.56e-10
Nyeri Dada	Wald	157.377	168.013	<2e-16	<2e-16
	G	374.703	358.034		

Berdasarkan hasil dari pengujian diatas, secara simultan pemodelan Logit dan pemodelan Complementary log-log menghasilkan keputusan yang sama yaitu dengan $G > \chi^2_{(0.05,\ 4)} = 9.487729$ dengan demikian, dapat disimpulkan bahwa setidaknya ada satu variabel pediktor yang mempengaruhi terjadinya Penyakit Gagal Jantung. Sedangkan secara parsial dengan pemodelan Logit dan juga Complementary log-log semua variabel signifikan dengan $W > \chi^2_{(0.05,\ 4)} = 3.841$ kesimpulannya adalah, semua variabel berpengaruh terhadap Penyakit Gagal Jantung.

C. Pemilihan Model Terbaik

Pemilihan model terbaik dilakukan berdasarkan AIC value. Hasilnya menunjukkan bahwa model logit lebih baik dibandingkan complementary log-log dikarenakan model logit memiliki AIC value yang lebih kecil.

Tabel 6. AIC value

Pemodelan	AIC value
Logit	897.43
Complementary Log - Log	914.1

Berikut model regresi dengan fungsi hubung menggunakan pemodelan Logit :

$$\hat{g} = -0.327 - 0.696X_1 - 1.355X_2 + 1.439X_3 + 2.32X_4$$

D. Perbandingan Ketepatan Klasifikasi Model

Perandingan ini akan menunjukkan seberapa baik yang dihasilkan dari pemodelan Logit dan pemodelan Complementary Log-Log mampu mengklasifikasikan Penyakit Gagal Jantung.

Tabel 7. Tabel Perbandingan

Logit			Akurasi(%)
Observasi	**Predicted**		
	0	**1**	
Tidak mengalami Penyakit Gagal Jantung (0)	334	76	**78.13 %**

©Department of Statistics, Universitas Negeri Padang
DOI: https://doi.org/10.24036/ujsds/vol3-iss4/421

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

Mengalami Penyakit Gagal Jantung (1)	128	380	

Complemenray Log-Log			**Akurasi(%)**
Observasi	**Predicted**		
	0	**1**	
Tidak mengalami Penyakit Gagal Jantung (0)	334	76	**78.13 %**
Mengalami Penyakit Gagal Jantung (1)	128	380	

Tingkat keakuratan prediksi menggunakan kedua model menghasilkan persentase yang sama, dimana baik pemodelan Logit maupun Cloglog dapat memprediksi 78.13% data. Hasil dari kedua pemodelan sangat mirip dikarenakan distribusi data yang seimbang dan tidak terlalu ekstream.

E. Perbandingan Odds Rasio

Odds rasio ini digunakan untuk mengetahui besar kecenderungan peubah penjelas mempengaruhi peubah respon.

Tabel 8. Odds Rasio Cloglog

Peubah	**Kategori**	**Odds Ratio**
Intercept		0.721
Usia	28 – 52 tahun 53 – 77 tahun	0.498
Jenis Kelamin	Perempuan Laki – Laki	0.258
Gula Darah	>120 mg/dl ≤120 mg/dl	4.215
Nyeri Dada	Mengalami Tidak Mengalami	10.167

Dari output diatas dapat kita lihat bahwa seseorang yang berusia 28 sampai 52 tahun mempunyai kecenderungan 0.498 kali lebih rendah menderita jantung dari pada seseorang yang berusia 53 sampai 77 tahun. Perempuan mempunyai keenderungan 0.258 kali lebih rendah menderita Penyakit Gagal Jantung jika dibandingkan dengan Laki – laki. Seseorang dengan gula darah > 120 mg/dl mempunyai kecenderungan 4.215 kali lebih tinggi menderita Penyakit Gagal Jantung jika dibandingkan dengan seseorang dengan gula darah < 120 mg/dl. Dan seseorang yang mengalami nyeri dada mempunyai keenderungan 10.167 kali lebih tinggi menderita Penyakit Gagal Jantung jika dibandingkan dengan seseorang yang tidak mengalami nyeri dada.

F. Perbandingan dengan Penelitian Sebelumnya

Hasil penelitian ini menjunjukkan bahwa Perempuan mempunyai keenderungan 0.258 kali lebih rendah menderita Penyakit Gagal Jantung jika dibandingkan dengan Laki – laki, sejalan dengan penelitian oleh (Sihombing, 2022) yang juga menunjukkan bahwa seorang Pria memiliki peluang lebih tinggi sebesar 3.3% terkena penyakit jantung dibandingkan wanita.

Secara klinis, hasil penelitian ini mengidentifikasi bahwa pasien dengan gula darah tinggi dan memiliki nyeri dada perlu mendapatkan pemantauan jantung yang lebih ketat untuk mencegah komplokasi gagal jantung. Karna dapat kita ketahui, secara biologis kadar gula tinggi dapat mempercepat proses aterosklerosis dan menurunkan elastisitas pembuluh darah yang pada akhirnya meningkatkan beban kerja jantung dan risiko gagal jantung.

©Department of Statistics, Universitas Negeri Padang
DOI: https://doi.org/10.24036/ujsds/vol3-iss4/421

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

IV. KESIMPULAN

Berdasarkan hasil penelitian, model regresi logistik logit merupakan pemodelan terbaik berdasarkan AIC value. Bentuk dari model regresi logit yaitu :

$$\hat{g} = -0.327 - 0.696X_1 - 1.355X_2 + 1.439X_3 + 2.32X_4$$

Dengan pemanfaatan odds rasio didapatkan kesimpulan bahwa seseorang yang berusia 28 sampai 52 tahun mempunyai kecenderungan 0.498 kali lebih rendah menderita jantung dari pada seseorang yang berusia 53 sampai 77 tahun. Perempuan mempunyai keenderungan 0.258 kali lebih rendah menderita Penyakit Gagal Jantung jika dibandingkan dengan Laki – laki. Seseorang dengan gula darah > 120 mg/dl mempunyai kecenderungan 4.215 kali lebih tinggi menderita Penyakit Gagal Jantung jika dibandingkan dengan seseorang dengan gula darah < 120 mg/dl. Dan seseorang yang mengalami nyeri dada mempunyai keenderungan 10.167 kali lebih tinggi menderita Penyakit Gagal Jantung jika dibandingkan dengan seseorang yang tidak mengalami nyeri dada.

DAFTAR PUSTAKA

Amelia, A., Mulyani, F., & Nabilla, U. (2020). Analisis Faktor yang Mempengaruhi Status Penerimaan Beras Keluarga Miskin Menggunakan Regresi Logistik Biner di Kecamatan Langsa Barat. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *14*(2), 175–186.

Dewi, A. F., & Pratiwi, R. (2021). Analisis regresi logistik biner pada pengaruh harga, kualitas pelayanan dan promosi terhadap kepuasan pelanggan dalam menggunakan jasa layanan grab di kabupaten lamongan. *Inferensi*, *4*(2), 77–84.

Hendayana, R. (2013). Penerapan metode regresi logistik dalam menganalisis adopsi teknologi pertanian. *Informatika Pertanian*, *22*(1), 1–9.

Lin, T. C., Yang, H.-C. P., & Chen, T.-H. (2010). Enhancement of the predictive accuracy in logistic regression model by optimal threshold. *Journal of Statistics and Management Systems*, *13*(3), 501–513.

McFadden, D. (2022). Instability in mixed logit demand models. *Journal of Choice Modelling*, *43*, 100353.

Nugraha, J., & Fatimah, I. (2016). Modeling of photocatalytic activity of ZnO/AC by using linear probability model, logit and complementary log transformation. *Procedia Engineering*, *148*, 1112–1120.

Nuradilla, S. (2024). Analysis of Academic Hardiness Factors Affecting Student Emotional Exhaustion in Malang Using Logit and Probit Models. *Jurnal Matematika, Statistika Dan Komputasi*, *20*(3), 606–622.

Nursita, H., & Pratiwi, A. (2025). Peningkatan Kualitas Hidup Pada Pasien Gagal Jantung: A Narrative Review Article. *Jurnal Berita Ilmu Keperawatan*, *18*(1), 10–21.

Raharja, A. R., Pramudianto, A., & Muchsam, Y. (2024). Penerapan Algoritma Decision Tree dalam Klasifikasi Data 'Framingham'Untuk Menunjukkan Risiko Seseorang Terkena Penyakit Jantung dalam 10 Tahun Mendatang. *Technol. J*, *1*(1).

Ratnasari, V., & Putri, R. C. (2015). Pemodelan logit, probit dan complementary log-log pada studi kasus partisipasi perempuan dalam pembangunan ekonomi di Kalimantan Selatan. *Jurnal Sains Dan Seni ITS*, *4*(2), 15676.

Saraswati, D. (2024). Inovasi Pelayanan Kesehatan: Deteksi Dini Penyakit Jantung Koroner melalui Posbindu PTM. *Jurnal Kesehatan Dan Kebidanan Nusantara*, *2*(1), 10–16.

Sihombing, P. R. (2022). Aplikasi Pemodelan Logit, Probit dan Clog-Log pada Regresi Binomial (Studi Kasus: Pemodelan Penyakit Jantung). *Jurnal Multidisiplin Madani*, *2*(6), 2599–2610.

Tampil, Y., Komaliq, H., & Langi, Y. (2017). Analisis Regresi Logistik Untuk Menentukan Faktor-Faktor Yang

©Department of Statistics, Universitas Negeri Padang
DOI: https://doi.org/10.24036/ujsds/vol3-iss4/421

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa FMIPA Universitas Sam Ratulangi Manado. *D\'Cartesian: Jurnal Matematika Dan Aplikasi*, *6*(2), 56–62.

Wala, J., Herman, H., & Umar, R. (2024). Implementasi K-Means Clustering pada Pengelompokan Pasien Penyakit Jantung. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *9*(3), 205–216.

Widadi, S. Y., Ramdani, H. T., & Ibrahim, D. Y. (2024). Gambaran Kualitas Hidup dan Perawatan Diri Pada Pasien Gagal Jantung Kongestif Rawat Inap Di Rumah Sakit dr Slamet Garut. *Innovative: Journal Of Social Science Research*, *4*(3), 5739–5763.

©Department of Statistics, Universitas Negeri Padang
DOI: https://doi.org/10.24036/ujsds/vol3-iss4/421