ANALISIS KLASIFIKASI MENGGUNAKAN REGRESI LOGISTIK BINER UNTUK ANALISIS FAKTOR-FAKTOR YANG MEMENGARUHI STATUS PERAWATAN PASIEN BERDASARKAN *ELECTRONIC HEALTH RECORD* (EHR)

(Skripsi)

Oleh

RAHMA VISTA ARISTAWATI
NPM 1917031052

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2026

ABSTRACT

CLASSIFICATION ANALYSIS USING BINARY LOGISTIC REGRESSION TO IDENTIFY FACTORS INFLUENCING PATIENTS CARE STATUS BASED ON ELECTRONIC HEALTH RECORD (EHR)

By

RAHMA VISTA ARISTAWATI

Electronic Health Record (EHR) provide patients health information in digital form and can be used as a basis for determining patients care status (inpatient or outpatient). The effective utilization of EHR data requires analytical methods that can support fast and objective decision-making. This study aims to classify patients' care status based on EHR data using binary logistic regression. The variables considered include hemoglobin level, leukocyte count, platelet count, age, and sex. Parameter estimation was performed using the Maximum Likelihood Estimation (MLE) method, with model adequacy evaluated using the AIC and the deviance test. The results indicate that hemoglobin, leukocyte count, platelet count, and sex have significant effects on patients' care status, whereas age is not significant and was excluded from the final model. The resulting model achieves a classification accuracy of 72.19%.

Keywords: Electronic Health Records, classification, binary logistic regression, Maximum Likelihood Estimation, AIC, deviance test.

ABSTRAK

ANALISIS KLASIFIKASI MENGGUNAKAN REGRESI LOGISTIK BINER UNTUK ANALISIS FAKTOR-FAKTOR YANG MEMENGARUHI STATUS PERAWATAN PASIEN BERDASARKAN *ELECTRONIC HEALTH RECORD* (EHR)

Oleh

RAHMA VISTA ARISTAWATI

Electronic Health Record (EHR) menyediakan informasi kesehatan pasien secara digital dan dapat dimanfaatkan sebagai dasar pertimbangan dalam menentukan status perawatan pasien (rawat inap atau rawat jalan). Pemanfaatan data EHR yang optimal menuntut adanya metode analisis yang mampu membantu proses pengambilan keputusan secara cepat dan objektif. Penelitian ini bertujuan untuk mengklasifikasikan status perawatan pasien berdasarkan data EHR menggunakan metode regresi logistik biner. Variabel yang digunakan meliputi kadar hemoglobin, jumlah leukosit, jumlah trombosit, usia, dan jenis kelamin. Estimasi parameter dilakukan dengan metode *Maximum Likelihood Estimation* (MLE), dengan evaluasi kesesuaian model menggunakan AIC dan uji deviance. Hasil penelitian menunjukkan bahwa variabel hemoglobin, leukosit, trombosit, dan jenis kelamin berpengaruh signifikan terhadap status perawatan pasien, sedangkan usia tidak signifikan dan dikeluarkan dari model akhir. Model yang dihasilkan mencapai tingkat akurasi klasifikasi sebesar 72,19%.

Kata Kunci: *Electronic Health Record*, klasifikasi, regresi logistik biner, *Maximum Likelihood Estimation*, AIC, uji deviance.

ANALISIS KLASIFIKASI MENGGUNAKAN REGRESI LOGISTIK BINER UNTUK ANALISIS FAKTOR-FAKTOR YANG MEMENGARUHI STATUS PERAWATAN PASIEN BERDASARKAN *ELECTRONIC HEALTH RECORD* (EHR)

Oleh

RAHMA VISTA ARISTAWATI

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA

pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2026

Judul Skripsi	:	**ANALISIS KLASIFIKASI MENGGUNAKAN REGRESI LOGISTIK BINER UNTUK ANALISIS FAKTOR-FAKTOR YANG MEMENGARUHI STATUS PERAWATAN PASIEN BERDASARKAN *ELECTRONIC HEALTH RECORD* (EHR)**
Nama Mahasiswa	:	Rahma Vista Aristawati
Nomor Pokok Mahasiswa	:	1917031052
Jurusan	:	Matematika
Fakultas	:	Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing

Prof. Drs. Mustofa Usman, M.A., Ph.D.	**Dr. Muslim Ansori, S.Si., M.Si.**
NIP. 19570101191984041001	NIP. 197202271998021001

2. Ketua Jurusan Matematika

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 197403162005011001

MENGESAHKAN

1. Tim Penguji

Ketua : **Prof. Drs. Mustofa Usman, M.A., Ph.D.**

Sekertaris : **Dr. Muslim Ansori, S.Si., M.Si.**

Penguji
Bukan Pembimbing : **Dr. Dian Kurniasari, S.Si., M.Sc.**

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Eng. Heri Satria, S.Si., M.Si.
NIP. 197110012005011002

Tanggal Lulus Ujian Skripsi :**26 Januari 2026**

SURAT PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan dibawah ini:

Nama	: Rahma Vista Aristawati
Nomor Pokok Mahasiswa	: 1917031052
Jurusan	: Matematika
Judul Skripsi	: **ANALISIS KLASIFIKASI MENGGUNAKAN REGRESI LOGISTIK BINER UNTUK ANALISIS FAKTOR-FAKTOR YANG MEMENGARUHI STATUS PERAWATAN PASIEN BERDASARKAN *ELECTRONIC HEALTH RECORD* (EHR**

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan apabila dikemudian hari hasil penelitian atau tugas akhir saya merupakan salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku. Semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya tulis ilmiah Universitas Lampung.

Bandar Lampung, 6 Februari 2026
Yang membuat pernyataan,

Rahma Vista Aristawati

RIWAYAT HIDUP

Penulis bernama lengkap Rahma Vista Aristawati dilahirkan di Pringsewu pada 25 Agustus 2002. Penulis merupakan anak pertama dari dua bersaudara, dari pasangan Bapak Kalidi dan Ibu Ismiasih.

Pendidikan formal pertama penulis ditempuh di Taman Kanak Kanak Sekar Wangi pada tahun 2007-2008, kemudian melanjutkan pendidikan dasar di SD Negeri 1 Sukaraja pada tahun 2008-2014. Setelah itu, penulis melanjutkan pendidikan di SMP Negeri 1 Gadingrejo pada tahun 2014-2017, dan meneruskan ke jenjang menengah atas di SMA Negeri 1 Gedong Tataan pada tahun 2017-2019. Pada tahun 2019, penulis diterima sebagai mahasiswa Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung melalui jalur Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN).

Selama menempuh pendidikan di perguruan tinggi, penulis aktif di beberapa kegiatan diantaranya:

1. Pada tahun 2019 menjadi anggota magang di UKMF Natural FMIPA Unila
2. Pada Januari - Desember 2020 menjadi anggota pengurus bidang kaderisasi UKMF Natural FMIPA Unila.
3. Pada Januari - Desember 2021 menjadi Kepala Bidang Kaderisasi UKMF Natural FMIPA Unila.
4. Pada Januari - Februari 2022 melaksanakan Kuliah Kerja Nyata di Desa Pekondoh, Kecamatan Way Lima, Kabupaten Pesawaran, Provinsi Lampung
5. Pada Juni - Agustus 2022 melaksanakan Kerja Praktik di KSP Kopdit Gentiaras Pringsewu, Provinsi Lampung.

KATA INSPIRASI

“Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya.”
(Q.S. Al-Baqarah:256)

“Dan bersabarlah kamu, sesungguhnya janji Allah adalah benar.”
(Q.S. Ar-Rum:60)

“Hidup bukan saling mendahului, bermimpilah sendiri-sendiri.”
(Baskara Putra-Hindia)

PERSEMBAHAN

Bismillahirrahmanirrahim

Dengan mengucapkan puji syukur kepada Allah SWT, penulis mempersembahkan karya sederhana ini kepada:

Kepada Orang Tua Tercinta dan Adik Tersayang

Terima kasih yang sebesar-besarnya penulis sampaikan kepada kedua orang tua tercinta atas segala pengorbanan, doa yang tidak pernah putus, dan kasih sayang yang tak terhingga. Tak lupa, terima kasih kepada adik tersayang atas semangat dan keceriaan yang diberikan kepada penulis.

Keluarga Besar

Yang memberikan kebersamaan, kehangatan, dan dukungan kepada penulis.

Sahabat-sahabat Terbaikku

Yang selalu membersamai, mendukung, dan mendoakan dalam setiap proses perjalanan hidup penulis.

Dosen Pembimbing dan Pembahas

Dosen-dosen yang telah memberikan ilmu, bimbingan, serta pengalaman berharga kepada penulis.

Almamater Tercinta Universitas Lampung

SANWACANA

Puji syukur penulis panjatkan kehadirat Allah SWT atas segala rahmat, nikmat, dan karunia-Nya kepada penulis sehingga penulis dapat menyelesaikan skripsi yang berjudul "Analisis Klasifikasi Menggunakan Regresi Logistik Biner Untuk Analisis Faktor yang Memengaruhi Status Perawatan Pasien Berdasarkan *Electronic Health Record* (EHR)". Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar strata satu (S1) di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

Penyusunan skripsi ini tidak terlepas dari bantuan, kerja sama, bimbingan dan doa dari berbagai pihak. Oleh karena itu, pada kesempatan ini, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Drs. Mustofa Usman, M.A., Ph.D., selaku Dosen Pembimbing I, atas kesediaan beliau dalam meluangkan waktu, tenaga, dan pikiran, serta bimbingan dan arahan yang berharga selama proses penyusunan skripsi ini.
2. Bapak Dr. Muslim Ansori, S.Si., M.Si., selaku Dosen Pembimbing II, atas bimbingan, masukan, dan saran yang diberikan kepada penulis selama proses penyusunan skripsi ini
3. Ibu Dr. Dian Kurniasari, S.Si., M.Sc., selaku Dosen Pembahas, atas kritik, saran, dan masukan yang membangun demi penyempurnaan skripsi ini.
4. Ibu Widiarti, S.Si., M.Si., selaku pembimbing akademik, atas nasihat dan arahan yang diberikan kepada penulis selama menempuh Pendidikan di Jurusa Matematika.
5. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku Ketua Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung

6. Bapak Dr. Eng. Heri Satria, S.Si., M.Si. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.
7. Seluruh dosen dan staff Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.
8. Ayah dan mamak tercinta, atas doa, kasih sayang, dukungan, motivasi, dan pengorbanan yang tiada henti demi kemudahan dan kelancaran penulis.
9. Adikku tercinta, Lia serta keluarga besar penulis yang selalu memberikan doa, kehangatan keluarga, dan dukungan selama ini kepada penulis.
10. Teman-teman emang jaya store (Clara, Rizqa, Nurje), yang telah menjadi teman perjalanan penulis sejak awal masa perkuliahan, saling berbagi cerita, kebersamaan dan semangat yang sangat berarti bagi penulis.
11. Argun, Ulfi, Dea, dan Anggun, yang telah menemani penulis, menjadi tempat bercerita dan memberikan dorongan serta banyak membantu penulis dalam menghadapi proses penyusunan skripsi
12. Temen seperjuangan skripsi Siti, Karima, Fitdes, dan Mia yang telah saling menyemangati dan menjadi teman diskusi selama proses penulisan skripsi ini.
13. Teman-teman Matematika 2019
14. Semua pihak yang terlibat dalam proses penulisan skripsi ini yang tidak dapat disebutkan satu persatu

Akhir kata, semoga Allah membalas segala kebaikan yang telah diberikan oleh semua pihak. Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan, namun besar harapan penulis agar skripsi ini dapat memberikan manfaat bagi pembaca.

Bandar Lampung, 6 Februari 2026
Penulis

Rahma Vista Aristawati

DAFTAR ISI

Halaman

DAFTAR TABEL

Tabel Halaman

DAFTAR GAMBAR

Gambar Halaman

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Perkembangan teknologi informasi telah membawa perubahan besar dalam sistem pelayanan kesehatan di seluruh dunia. Salah satu tantangan utama dalam era digital saat ini adalah bagaimana mengelola dan memanfaatkan data kesehatan yang terus meningkat secara signifikan. Perkembangan ini mendorong pencatatan data kesehatan dari sistem manual menjadi lebih terstruktur dan digital melalui *Electronic Health Records* (EHR) atau Rekam Medis Elektronik (RME).

Electronic Health Records (EHR) adalah sistem informasi digital berupa catatan digital data kesehatan yang merekam kondisi kesehatan pasien dan tindakan medis yang dilakukan, serta dapat diakses oleh pihak pelayanan kesehatan (Fadhal, 2022). Sistem ini memungkinkan pengumpulan berbagai jenis data klinis yang mencakup rincian riwayat medis seperti hasil pemeriksaan diagnostik, obat yang dikonsumsi, rencana perawatan, riwayat penyakit, catatan alergi, pemeriksaan radiologi, dan informasi lainnya. Semua informasi yang tercatat dalam EHR sangat berharga untuk mendukung pengambilan keputusan klinis. Salah satu tantangan yang dihadapi oleh institusi pelayanan kesehatan, khususnya rumah sakit adalah menentukan dengan cepat pasien yang memerlukan perawatan inap atau cukup ditangani secara rawat jalan. Keputusan ini penting untuk mengoptimalkan alokasi sumber daya seperti ruang rawat, tenaga medis, dan waktu pelayanan, sehingga pelayanan kesehatan yang diberikan dapat mendukung perbaikan kualitas

pengobatan, lebih efisien, dan tepat sasaran, serta mengurangi adanya kesalahan medis.

Penentuan apakah pasien membutuhkan rawat inap atau cukup ditangani dengan rawat jalan sangat dipengaruhi oleh berbagai faktor seperti hasil uji laboratorium, usia, serta kondisi fisik pasien. Dengan ketersediaan data EHR, klasifikasi status perawatan pasien dapat diformulasikan menjadi kasus statistika biner. Salah satu metode yang dapat menangani permasalahan klasifikasi ini adalah adalah regresi logistik biner yang termasuk dalam kelompok analisis regresi dengan pendekatan analisis multivariabel. Analisis regresi adalah salah satu alat analisis yang bertujuan untuk mengetahui pengaruh suatu variabel terhadap variabel lain (Prasetyo, R.A., dan Helma, 2022). Variabel dalam analisis regresi terbagi menjadi 2, yaitu variabel terikat dan variabel bebas. Variabel dependen dalam analisis regresi ada yang bersifat dikotomus (biner) dengan nilai 0 atau 1. Analisis regresi yang cocok digunakan untuk variabel dependen yang bersifat biner adalah analisis regresi logistik.

Regresi logistik merupakan analisis regresi yang digunakan untuk menjelaskan hubungan antara variabel respon (terikat) yang bersifat kualitatif dan variabel prediktor (bebas), baik yang bersifat nominal atau ordinal, maupun interval atau rasio (Avini, dkk, 2022). Regresi logistik memiliki beberapa jenis berdasarkan jenis skala variabel terikatnya yaitu regresi logistik biner, regresi logistik multinomial, dan regresi logistik ordinal. Regresi logistik biner adalah metode regresi untuk menelaah hubungan antara variabel terikat (Y) dengan satu atau beberapa variabel bebas (X), dimana variabel terikatnya berupa data kualitatif dikotomi atau memiliki 2 kategori yaitu untuk menyatakan keberadaan suatu karakteristik dilambangkan dengan nilai 1 dan untuk menyatakan ketidakberadaan suatu karakteristik dilambangkan dengan nilai 0 (Ripai, dkk, 2022).

Beberapa metode sudah diterapkan untuk mengklasifikasikan data *Electronic Health Records* (EHR). Dalam penelitian yang dilakukan oleh Sadikin dan

Nurhaida (2021), digunakan beberapa metode klasifikasi seperti *Decision Tree* dengan tingkat akurasi sebesar 63.39%, *Gaussian Naïve Bayes* dengan tingkat akurasi sebesar 68.95%, dan *Random Forest* dengan tingkat akurasi tertinggi sebesar 71.58%. Sedangkan, pada tahun 2022 Rosita dkk melakukan prediksi apakah pasien memerlukan rawat inap atau cukup rawat jalan dengan menerapkan berbagai algoritma *machine learning*. Hasilnya menunjukkan bahwa metode Neural Network memberikan performa terbaik dengan tingkat akurasi mencapai 74,47%.

Penelitian sebelumnya dengan menggunakan metode Regresi Logistik Biner telah dilakukan oleh Bawono, 2019 untuk mengetahui metode mana yang paling baik antara Regresi Logistik Biner dan Naïve Bayes dalam mengklasifikasi debitur berdasarkan kualitas kredit nasabah dengan menggunakan nilai akurasi sebagai parameter perbandingan. Hasil penelitian menunjukkan bahwa metode Regresi Logistik Biner mencapai tingkat akurasi yang lebih tinggi, yakni sebesar 99,47% yang berarti metode Regresi Logistik Biner lebih baik dalam klasifikasi, sedangkan pada metode Naïve Bayes nilai akurasinya sebesar 96,55%.

Berdasarkan latar belakang dan hasil penelitian yang diuraikan, dapat disimpulkan bahwa regresi logistik biner merupakan metode yang relevan untuk digunakan dalam klasifikasi data *Electronic Health Record* (EHR). Metode ini dipilih sebagai pendekatan statistik klasik yang bersifat inferensial, berbeda dari pendekatan machine learning yang berfokus pada akurasi prediksi. Oleh karena itu, penelitian ini bertujuan untuk menerapkan metode regresi logistik biner dalam membangun model klasifikasi status perawatan pasien berdasarkan data rekam medis elektronik, dengan judul, “Analisis Klasifikasi Menggunakan Regresi Logistik Biner Untuk Analisis Faktor-Faktor yang Memengaruhi Status Perawatan Pasien Berdasarkan *Electronic Health Record* (EHR)”

1.2 Tujuan

Adapun tujuan yang ingin dicapai dalam penelitian ini antara lain:

1. Membangun model klasifikasi menggunakan regresi logistik biner untuk mengklasifikasikan status perawatan pasien berdasarkan data *Electronic Health Record* (EHR).
2. Mengidentifikasi variabel-variabel yang berpengaruh signifikan terhadap status perawatan pasien.
3. Mengetahui ketepatan klasifikasi dari model regresi logistik biner terhadap status perawatan pasien

1.3 Manfaat Penelitian

Adapun manfaat yang ingin diperoleh dari penelitian ini adalah sebagai berikut:

1. Meningkatkan pemahaman dan wawasan mengenai penerapan Regresi Logistik Biner dalam pemodelan klasifikasi
2. Dapat mengetahui keakuratan metode Regresi Logistik Biner dalam melakukan klasifikasi
3. Dapat menjadi sumber keilmuan dan referensi yang baik bagi pembaca

II. TINJAUAN PUSTAKA

2.1 Regresi Logistik Biner

Regresi logistik adalah suatu metode analisis data dalam statistika yang dapat menjelaskan hubungan antara peubah penjelas (X) dan peubah respon (Y) yang tidak dapat dijelaskan dengan model regresi linear biasa karena merupakan regresi non-linear yang digunakan untuk mendeskripsikan hubungan antara variabel X dan Y bersifat tidak linear, sebaran distribusi Y tidak normal, dan keragaman respon tidak konstan atau heteroskedastisitas (Agresti, 1996). Berdasarkan variabel responnya, ada tiga macam regresi logistik yaitu regresi logistik biner, regresi logistik multinomial, dan regresi logistik nominal. Pada regresi logistik biner variabel responnya berskala nominal dan terdiri dari dua kategori (biner). Regresi logistik multinomial variabel responnya juga berskala nominal dengan lebih dari dua kategori, dan regresi logistik ordinal variabel responnya merupakan data dengan skala ordinal (memiliki urutan tertentu).

Menurut Hosmer & Lemeshow (2000), regresi logistik biner adalah metode analisis statistika yang digunakan untuk menganalisis hubungan antara peubah respon dengan skala pengukuran dua kategori (*dichotomus*) dengan satu atau lebih peubah penjelas berskala kategori atau kontinu. Variabel respon berupa data kualitatif dikotomi, dimana nilai 1 menyatakan keberadaan suatu karakteristik dan nilai 0 menyatakan keberadaan suatu karakteristik. Hasil dari variabel respon (Y) misalnya yaitu "berhasil" dan "gagal" yang dinotasikan dengan $Y = 1$ dan $Y = 0$. Dimana $Y = 1$ mewakili kemungkinan "berhasil" dengan probabilitas

$$\Pr(Y = 1|\boldsymbol{x}) = \pi(\boldsymbol{x}) \qquad (2.1)$$

dan Y = 0 mewakili kemungkinan "gagal" dengan probabilitas

$$\Pr(Y = 0|\boldsymbol{x}) = 1 - \pi(\boldsymbol{x}) \tag{2.2}$$

dimana variabel respon (Y) mengikuti distribusi Bernoulli untuk setiap observasi tunggal.

Model regresi logistik yang terdiri dari dua atau lebih variabel bebas, dikenal dengan model multivariabel. Rata-rata bersyarat dari Y ketika X memiliki nilai tertentu adalah $\pi(\boldsymbol{x}) = E(\, Y \mid \boldsymbol{x}\,)$, dapat dilihat dari rumus

$$E(\, Y \mid \boldsymbol{x}\,) = (1)\big(\pi(\boldsymbol{x})\big) + (0)\big(1 - \pi(\boldsymbol{x})\big) = \pi(\boldsymbol{x}) \tag{2.3}$$

Model umum regresi logistik biner (dikotomus) dengan prediktor $x_1, x_2, \dots, x_k$ dinotasikan dalam persamaan (2.4) berikut

$$\pi(\boldsymbol{x}) = \frac{e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)}}{1 + e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)}} \tag{2.4}$$

Keterangan:

$\pi(\boldsymbol{x})$ = peluang berhasil dengan nilai probabilitas $0 < \pi(\boldsymbol{x}) < 1$

k = banyaknya variabel bebas

β_i = nilai parameter ke- i, dengan $i = 1, 2, \dots, k$

x_j = variabel penjelas, dengan $j = 1, 2, \dots, k$

Persamaan (2.4) merupakan fungsi non-linear, untuk mendapatkan fungsi yang linear maka fungsi tersebut perlu dilakukan transformasi ke dalam bentuk logit agar hubungan antara variabel bebas dan variabel terikat dapat terlihat (Hosmer & Lemeshow, 2000). Transformasi model menggunakan transformasi logit, yaitu $\mathrm{g}\big(\pi(\boldsymbol{x})\big) = \ln\left(\frac{\pi}{1-\pi}\right)$ mengubah interval $[-\infty, \infty]$ menjadi $[0, 1]$. Fungsi $\pi(\boldsymbol{x})$ ketika dilakukan transformasi logit sebagai berikut:

$$\pi(\boldsymbol{x}) = \frac{e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)}}{1 + e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)}}$$

$$\big(\pi(\boldsymbol{x})\big)\big(1 + e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)}\big) = e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)}$$

$$\big(\pi(\boldsymbol{x})\big) + \big(\pi(\boldsymbol{x})\, e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)}\big) = e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)}$$

$$\pi(\boldsymbol{x}) = e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)} - \big(\pi(\boldsymbol{x})\, e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)}\big)$$

$$\pi(\boldsymbol{x}) = \big(1 - \pi(\boldsymbol{x})\big)\, e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)}$$

Pada tahap ini, bentuk $\frac{\pi(x)}{1-\pi(x)}$ merupakan odds, yaitu rasio peluang terjadinya kejadian terhadap tidak terjadinya kejadian.

$$\frac{\pi(\boldsymbol{x})}{1-\pi(\boldsymbol{x})} = e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)}$$

Mengambil logaritma natural dari odds diperoleh:

$$\ln\left(\frac{\pi(\boldsymbol{x})}{1-\pi(\boldsymbol{x})}\right) = \ln\left(e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)}\right)$$

$$= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

dimana $\ln\left(\frac{\pi(x)}{1-\pi(x)}\right)$ disebut log-odds atau fungsi logit, karena merupakan logaritma dari odds. Sehingga diperoleh persamaan (2.5) yang lebih sederhana yang merupakan fungsi linear, yaitu:

$$\mathrm{g}\big(\pi(\boldsymbol{x})\big) = \ln\left(\frac{\pi(\boldsymbol{x})}{1-\pi(\boldsymbol{x})}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \qquad (2.5)$$

dengan $\mathrm{g}\big(\pi(\boldsymbol{x})\big)$ dikatakan sebagai fungsi logit model regresi logistik biner dengan k variabel bebas.

2.2 Multikolinearitas

Multikolinearitas merupakan suatu kondisi dimana terdapat korelasi atau hubungan antara variabel-variabel bebas (Indriyani, dkk, 2020). Model regresi yang baik yaitu tidak memiliki korelasi antar variabelnya. Adanya korelasi yang tinggi dapat menyebabkan hasil estimasi yang tidak akurat dan penduga yang bias.

Indikasi adanya multikolinearitas dapat dilihat dari besarnya koefisien korelasi antarvariabel bebas. Menurut Hair et al (2019), apabila terdapat pasangan variabel independen dengan koefisien korelasi absolut lebih besar dari 0,80, maka hal tersebut mengindikasikan adanya multikolinearitas yang kuat.

Selain dilihat dari besarnya koefisien korelasi, menurut Gujarati (1995), gejala multikolinearitas dalam model regresi dapat di uji menggunakan nilai *Variance Inflation Factor* (VIF) dari masing-masing variabel bebas (independen) terhadap variabel terikat (dependen). Perhitungan nilai VIF dicari menggunakan rumus (2.6) berikut.

$$VIF_{(c)} = \frac{1}{(1 - R_c^2)} \; ; \; j = 1, 2, \dots, k \tag{2.6}$$

dimana,

c = indeks variabel bebas variabel

R_c^2 = koefisien determinasi yang diperoleh dari variabel bebas (X_c) yang diregresikan dengan variabel bebas lainnya.

Suatu model regresi dikatakan bebas multikolinearitas ketika model memiliki nilai VIF < 10 dan nilai VIF > 10 mengindikasikan adanya multikolinearitas dalam model.

2.3 Estimasi Parameter

Untuk menduga parameter yang tidak diketahui pada model regresi logistik, salah satu metode yang digunakan adalah *Maximum Likelihood Estimation* (MLE). Metode ini merupakan metode untuk mengestimasi nilai parameter β dengan cara memaksimalkan fungsi likelihood sehingga didapatkan penaksir parameter dengan kemungkinan maksimum (Purba, 2020). Proses awal untuk menggunakan metode *Maximum Likelihood Estimation* (MLE) adalah dengan membentuk fungsi likelihood, yang menggambarkan fungsi probabilitas dari data yang digunakan sebagai fungsi dari penduga parameter (Hosmer and Lemeshow, 2000).

Dalam model regresi logistik, variabel respon mengikuti distribusi Bernoulli dengan fungsi probabilititas dari pasangan variabel pada pengamatan ke-i $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ dengan $i = 1, 2, \dots, n$ secara umum dinyatakan dalam persamaan (2.7) berikut.

$$f(y_i) = \pi(\boldsymbol{x}_i)^{y_i}[1 - \pi(\boldsymbol{x}_i)]^{1-y_i} \; ; \; Y = 0{,}1 \tag{2.7}$$

dengan $\pi(\boldsymbol{x}_i) = \frac{e^{\left(\sum_{j=0}^{k} \beta_j x_{ij}\right)}}{1+e^{\left(\sum_{j=0}^{k} \beta_j x_{ij}\right)}}$

dimana jika $j = 0$, maka nilai dari $\boldsymbol{x}_{ij} = \boldsymbol{x}_{i0} = 1$. Nilai variabel respon $(\boldsymbol{Y}_i)$ diasumsikan saling bebas satu sama lain, sehingga fungsi likelihood untuk parameter $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)$ diperoleh dalam persamaan berikut.

$$l(\boldsymbol{\beta}) = \prod_{i=1}^{n} \pi(\boldsymbol{x}_i)^{y_i}[1 - \pi(\boldsymbol{x}_i)]^{1-y_i} \tag{2.8}$$

dimana,

$\boldsymbol{\beta}$ = vektor parameter yang akan diasumsi

$\pi(\boldsymbol{x}_i)$ = prediksi probabilitas

y_i = nilai sebenarnya dari variabel respon, dengan nilai (0 atau 1)

Karena setiap pengamatan diasumsikan saling bebas atau independen, maka fungsi likelihood total dapat dihitung dengan menggabungkan fungsi likelihood dari masing-masing pengamatan, yaitu sebagai berikut.

$$\begin{aligned}
l(\boldsymbol{\beta}) &= \prod_{i=1}^{n} [\pi(\boldsymbol{x}_i)^{y_i}[1 - \pi(\boldsymbol{x}_i)]^{1-y_i}] \\
&= \left\{\prod_{i=1}^{n} \pi(\boldsymbol{x}_i)^{y_i}\right\} \left\{\prod_{i=1}^{n} \left(\left(1 - \pi(\boldsymbol{x}_i)\right)\left(1 - \pi(\boldsymbol{x}_i)\right)^{-y_i}\right)\right\} \\
&= \left\{\prod_{i=1}^{n} \left(1 - \pi(\boldsymbol{x}_i)\right)\right\} \left\{\prod_{i=1}^{n} \pi(\boldsymbol{x}_i)^{y_i}\right\} \left\{\prod_{i=1}^{n} \left(1 - \pi(\boldsymbol{x}_i)\right)^{-y_i}\right\} \\
&= \left\{\prod_{i=1}^{n} \left(1 - \pi(\boldsymbol{x}_i)\right)\right\} \left\{\prod_{i=1}^{n} e^{\left(\ln\left(\frac{\pi(\boldsymbol{x}_i)}{1-\pi(\boldsymbol{x}_i)}\right)^{y_i}\right)}\right\} \\
&= \left\{\prod_{i=1}^{n} \left(1 - \pi(\boldsymbol{x}_i)\right)\right\} \left\{e^{\left(\sum_{i=1}^{n} y_i \ln\left(\frac{\pi(\boldsymbol{x}_i)}{1-\pi(\boldsymbol{x}_i)}\right)\right)}\right\} \\
&= \left\{\prod_{i=1}^{n} \left(\frac{1}{1 + e^{\sum_{j=0}^{k} \beta_j x_{ij}}}\right)\right\} \left\{e^{\left(\sum_{j=0}^{k} \beta_j \left(\sum_{i=1}^{n} y_i x_{ij}\right)\right)}\right\}
\end{aligned}$$

Agar mempermudah hitungan matematis, fungsi likelihood yang dinyatakan dengan $l(\boldsymbol{\beta})$ dioptimalkan dalam bentuk log-likelihood yang dinyatakan dengan ln $l(\boldsymbol{\beta})$ sebagai berikut.

$$L(\boldsymbol{\beta}) = \ln\big(l(\boldsymbol{\beta})\big)$$

$$= \ln\left[\left\{\prod_{i=1}^{n}\left(\frac{1}{1+e^{\sum_{j=0}^{k}\beta_j x_{ij}}}\right)\right\}\left\{e^{\left(\sum_{j=0}^{k}\beta_j\left(\sum_{i=1}^{n}y_i x_{ij}\right)\right)}\right\}\right]$$

$$= \left\{\ln\left(e^{\left(\sum_{j=0}^{k}\beta_j\left(\sum_{i=1}^{n}y_i x_{ij}\right)\right)}\right)\right\}\left\{\ln\left(\prod_{i=1}^{n}\left(1+e^{\sum_{j=0}^{k}\beta_j x_{ij}}\right)^{-1}\right)\right\}$$

$$= \sum_{j=0}^{k}\beta_j\left(\sum_{i=1}^{n}y_i x_{ij}\right) - \sum_{i=1}^{n}\ln\left(1+e^{\sum_{j=0}^{k}\beta_j x_{ij}}\right)$$

Nilai parameter $\boldsymbol{\beta}$ yang memaksimumkan $L(\boldsymbol{\beta})$ didapatkan melalui pendekatan diferensiasi dengan membuat turunan pertama dari $L(\boldsymbol{\beta})$ terhadap $\boldsymbol{\beta}$ sama dengan nol, sebagai berikut.

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{\partial}{\partial \boldsymbol{\beta}}\left[\sum_{j=0}^{k}\beta_j\left(\sum_{i=1}^{n}y_i x_{ij}\right) - \sum_{i=1}^{n}\ln\left(1+e^{\sum_{j=0}^{k}\beta_j x_{ij}}\right)\right] =$$

$$= \sum_{i=1}^{n}y_i x_{ij} - \sum_{i=1}^{n}x_{ij}\left(\frac{e^{\sum_{j=0}^{k}\beta_j x_{ij}}}{1+e^{\sum_{j=0}^{k}\beta_j x_{ij}}}\right) = 0$$

Sehingga diperoleh persamaan,

$$\sum_{i=1}^{n}y_i x_{ij} - \sum_{i=1}^{n}x_{ij}\,\pi(\boldsymbol{x}_i) = 0 \qquad (2.9)$$

$$\sum_{i=1}^{n}x_{ij}\left(y_i - \pi(\boldsymbol{x}_i)\right) = 0$$

dengan $j = 0, 1, 2, \ldots, k$

Menurut Hosmer dan Lemeshow (2000) estimasi varians dan kovarians dikembangkan menggunakan teori *Maximum Likelihood Estimation* (MLE) dari koefisien parameternya. Berdasarkan teori ini, estimasi varians kovarians diperoleh melalui turunan kedua dari fungsi log-likelihood $L(\boldsymbol{\beta})$ yaitu sebagai berikut.

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_j \partial \boldsymbol{\beta}_s} = -\sum_{i=1}^{n} x_{ij} x_{is} \pi_i (1 - \pi_i) \tag{2.10}$$

dengan $j, s = 0, 1, 2, \ldots, k$ dan π_i menunjukkan $\pi(\boldsymbol{x}_i)$.

Matriks varians-kovarians dari parameter estimasi diperoleh melalui invers matriks (2.11) beriikut.

$$\text{Cov}\left(\widehat{\boldsymbol{\beta}}\right) = \left(\mathbf{X}'\widehat{\mathbf{V}}\mathbf{X}\right)^{-1} \tag{2.11}$$

dimana $\mathbf{X}$ merupakan matriks $\left(n \times (m+1)\right)$ dari setiap pengamatan dan matriks $\mathbf{X}$ dinyatakan sebagai berikut.

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

dan $\widehat{\mathbf{V}}$ merupakan matriks diagonal $(n \times n)$ dengan diagonal utamanya adalah $\left(\hat{\pi}_i(1 - \hat{\pi}_i)\right)$ dan matriks $\widehat{\mathbf{V}}$ dinyatakan sebagai berikut.

$$\widehat{\mathbf{V}} = \begin{bmatrix} \hat{\pi}_1(1 - \hat{\pi}_1) & 0 & 0 & 0 \\ 0 & \hat{\pi}_1(1 - \hat{\pi}_1) & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \hat{\pi}_n(1 - \hat{\pi}_n) \end{bmatrix}$$

Estimasi standar error $\left(\text{SE}\left(\left(\widehat{\boldsymbol{\beta}}\right)\right)\right)$ dari koefisien parameter yang diestimasi diperoleh dengan mengambil akar kuadrat dari diagonal utama matriks varians-kovarians.

Model regresi logistik memiliki persamaan yang non-linear terhadap parameter $\boldsymbol{\beta}$, sehingga solusi bagi $\hat{\beta} = \hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ tidak dapat dinyatakan secara eksplisit, maka diperlukan metode numerik untuk mendapatkan estimasi yang akurat. Metode iterasi Newton Rapshon dapat digunakan untuk memperoleh nilai estimasi $\boldsymbol{\beta}$ dengan menghitung turunan pertama dari fungsi $L(\boldsymbol{\beta})$. Iterasi merupakan metode umum dalam perhitungan estimasi dari parameter $\boldsymbol{\beta}$ (Hosmer & Lemeshow, 2000).

Langkah-langkah iterasi Newton Rapshon dapat dijelaskan sebagai berikut:

1. Mulailah dengan menentukan nilai dugaan awal parameter β_0 untuk diestimasi
2. Membentuk vektor gradien $\mathbf{q}$ yang berisi turunan pertama dari fungsi log-likelihood $\big(L(\beta)\big)$ terhadap parameter β. Turunan pertamanya yaitu $\frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$

$$\mathbf{q} = \left(\frac{\partial L(\beta)}{\partial \beta_0}, \frac{\partial L(\beta)}{\partial \beta_1}, \dots, \frac{\partial L(\beta)}{\partial \beta_k}\right)$$

3. Membentuk matriks Hessian $\mathbf{H}$ yang berisi turunan kedua dari fungsi log-likelihood $\big(L(\beta)\big)$ terhadap parameter β. Turunan keduanya yaitu $\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_j \partial \boldsymbol{\beta}_s}$

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 L(\beta)}{\partial {\beta_0}^2} & \frac{\partial^2 L(\beta)}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 L(\beta)}{\partial \beta_0 \partial \beta_2} & \cdots & \frac{\partial^2 L(\beta)}{\partial \beta_0 \partial \beta_k} \\ \frac{\partial^2 L(\beta)}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 L(\beta)}{\partial {\beta_1}^2} & \frac{\partial^2 L(\beta)}{\partial \beta_1 \partial \beta_2} & \cdots & \frac{\partial^2 L(\beta)}{\partial \beta_1 \partial \beta_k} \\ \frac{\partial^2 L(\beta)}{\partial \beta_0 \partial \beta_2} & \frac{\partial^2 L(\beta)}{\partial \beta_1 \partial \beta_2} & \frac{\partial^2 L(\beta)}{\partial {\beta_2}^2} & \cdots & \frac{\partial^2 L(\beta)}{\partial \beta_2 \partial \beta_k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 L(\beta)}{\partial \beta_0 \partial \beta_k} & \frac{\partial^2 L(\beta)}{\partial \beta_1 \partial \beta_k} & \frac{\partial^2 L(\beta)}{\partial \beta_2 \partial \beta_k} & \cdots & \frac{\partial^2 L(\beta)}{\partial {\beta_k}^2} \end{bmatrix}$$

4. Masukkan nilai β_0 ke dalam vektor $\mathbf{q}$ dan matriks $\mathbf{H}$ sehingga didapatkan vektor $\mathbf{q}_t(\beta_0)$ dan matriks $\mathbf{H}_t(\beta_0)$
5. Mulai lakukan estimasi parameter β dari iterasi $t = 0$ dengan menggunakan formula iterasi Newton Rapshon (2.12) berikut

$$\beta_{t+1} = \beta_t - \big(\mathbf{H}_t(\beta_0)\big)^{-1} \mathbf{q}_t(\beta_0) \tag{2.12}$$

dengan $t = 1, 2, \dots$, sampai konvergen dan pada setiap iterasi berlaku

$$\frac{\partial^2 L(\beta)}{\partial \beta_j \beta_s} = -\sum_{i=1}^{n} x_{ij} x_{is} \pi_{i(t)} \big(1 - \pi_{i(t)}\big) \tag{2.13}$$

$$\frac{\partial L(\beta)}{\partial \beta_0} = \sum_{i=1}^{n} \big(y_i - \pi_{i(t)}\big) x_{ij} = 0 \tag{2.14}$$

$$\pi_{i(t)} = \frac{e^{\left(\sum_{j=0}^{k} \beta_{j(t)} x_{ij}\right)}}{1 + e^{\left(\sum_{j=0}^{k} \beta_{j(t)} x_{ij}\right)}} \tag{2.15}$$

6. Jika nilai dugaan parameter belum konvergen, maka ulangi langkah 5 dan teruskan hingga iterasi ke-$t = t + 1$.

Iterasi diberhentikan jika $\left\|\beta_{(t+1)} - \beta_{(t)}\right\| \leq \varepsilon$ yang dimana ε adalah sebuah bilangan yang sangat kecil. Estimasi parameter akhir yang diperoleh adalah $\beta_{(t+1)}$ pada iterasi terakhir ketika kriteria konvergensi sudah terpenuhi

2.4 Pengujian Parameter Model

Pengujian parameter model merupakan pengujian yang bertujuan untuk menguji koefisien β dalam model tersebut memiliki signifikansi atau tidak. Dengan menguji signifikansi koefisien β maka dapat diketahui pengaruh variabel bebas terhadap variabel terikatr dalam model. Pengujian parameter model meliputi pengujian secara simultan dan pengujian secara parsial.

2.4.1 Uji Simultan

Pengujian secara simultan dilakukan untuk memeriksa signifikansi koefisien β secara keseluruhan dengan menggunakan uji *Ratio Likelihood* (Hosmer, dkk, 2013).

- Hipotesis

  $H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$

  H_1 : minimal ada satu $\beta_j \neq 0$, dengan $j = 1,2,3, \ldots p$

- Statistik uji:

$$G = -2 \ln \left[\frac{likelihood\ tanpa\ variabel\ bebas}{likelihood\ dengan\ variabel\ bebas} \right]$$

$$G = -2 \ln \left[\frac{\left(\frac{n_1}{n}\right)^{n_1} \left(\frac{n_0}{n}\right)^{n_0}}{\prod_{i=1}^{n} \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{(1-y_i)}} \right] \qquad (2.16)$$

dimana,

n_0 = Jumlah pengamatan dengan kategori $y = 0$

n_1 = Jumlah pengamatan dengan kategori $y = 1$

n = Jumlah total pengamatan $(n_0 + n_1)$

p = Banyaknya parameter

- Daerah keputusan

  Statistik uji G mengikuti distribusi Chi-square dengan taraf signifikansi (α) dan derajat bebas (db) $= p$ yang merupakan jumlah prediktor dalam model, maka diperoleh keputusan

  Tolak H_0 jika nilai $G > X^2{}_{(a,p)}$

  Terima H_0 jika nilai $G < X^2{}_{(a,p)}$

- Kesimpulan

  Jika tolak H_0, maka model yang mengandung variabel bebas (independen) signifikan secara keseluruhan terhadap model

2.4.2 Uji Parsial

Pengujian secara parsial dilakukan untuk memeriksa signifikansi koefisien β secara individu dengan menggunakan uji Wald (Hosmer, dkk, 2013). UJI Wald bertujuan untuk melihat pengaruh masing-masing koefisien β terhadap variabel respon. Uji ini membandingkan parameter hasil (dugaan β) melalui maximum likelihood dengan standar error dari parameter tersebut.

- Hipotesis

  $H_0 : \beta_j = 0$

  $H_1 : \beta_j \neq 0$, dengan $j = 1,2,3, \dots p$

- Statistik uji:

$$W = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \tag{2.17}$$

dimana

$\hat{\beta}_j$ = Penduga bagi β_j

$SE(\hat{\beta}_j)$ = Penduga galat baku bagi β_j

- Daerah Keputusan

  Statistik uji W mengikuti sebaran normal baku dengan taraf signifikansi (α), maka diperoleh keputusan

  Tolak H_0 jika nilai $W > Z_{(a/2)}$

  Tidak tolak H_0 jika nilai $W < Z_{(a/2)}$

- Kesimpulan

  Jika tolak H_0 maka variabel bebas secara parsial memiliki pengaruh yang signifikan terhadap variabel terikat

2.5 Pengujian Kesesuaian Model

Menurut Ghozali (2018), uji kesesuaian model atau yang dikenal sebagai *goodness of fit test* merupakan sebuah uji yang bertujuan untuk menguji hipotesis nol yang menyatakan bahwa data empiris sesuai dengan model, yang berarti tidak ada perbedaan signifikan antara model dan data sehingga model dapat dikatakan *fit*. Pengujian ini dilakukan dengan uji Hosmer-Lemeshow dengan hipotesis yang digunakan yaitu:

H_0: Model sudah sesuai (tidak ada perbedaan signifikan antara model dan data)

H_1: Model tidak sesuai (ada perbedaan signifikan antara model dan data)

Dengan statistik ujinya adalah

$$\hat{C} = \sum_{k=1}^{g} \frac{(o_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)} \tag{2.18}$$

Keterangan:

g = Banyaknya grup

k = Indeks grup, $k = 1,2, \dots, g$

o_k = Jumlah nilai variabel respon pada grup ke- k

n'_k = Jumlah subjek pada grup ke- k

$\bar{\pi}_k$ = Rata-rata estimasi probabilitas

Statistik uji Hosmer-Lemeshow mengikuti distribusi Chi-square dengan taraf signifikansi (α) dan derajat bebas (db) $= g - 2$, maka aturan keputusannya adalah tolak H_0 jika nilai $\hat{C} > X^2{}_{(g-2)}$ atau $p - value < \alpha$ dan terima H_0 jika nilai $\hat{C} < X^2{}_{(g-2)}$ atau $p - value \geq \alpha$, sehingga dengan menolak H_0 maka disimpulkan model tidak sesuai dan terdapat perbedaan signifikan antara model dan data.

2.6 Interpretasi Parameter

Interpretasi koefisien parameter digunakan untuk memahami hubungan fungsional antar variabel respon (terikat) dengan variabel prediktor (bebas) dan menggambarkan bagaimana setiap perubahan pada variabel terikat disebabkan oleh variabel bebas. Menurut Hosmer dan Lemeshow (2000), koefisien yang diestimasi dari variabel bebas menyatakan *slope* atau tingkat perubahan variabel terikat ketika variabel bebas berubah satu unit.

Interpretasi koefisien parameter pada regresi logistik dengan varaibel bebas bersifat biner dapat menggunakan Odds ratio atau kecenderungan rasio. Odds ratio merupakan rasio peluang untuk kejadian ketika $x = 1$ dibandingkan dengan $x = 0$ (Jika x adalah variabel biner, diasumsikan bernilai 0 dan 1). Odds ratio dilambangkan dengan Ψ dan dituliskan dalam persamaan (2.19) berikut.

$$\Psi = \frac{\frac{\pi(1)}{1-\pi(1)}}{\frac{\pi(0)}{1-\pi(0)}} \tag{2.19}$$

dimana,

- Untuk $x = 1$, digunakan

$$\frac{\pi(1)}{1-\pi(1)} = \frac{\frac{exp\{\beta_0+\beta_j\}}{1+exp\{\beta_0+\beta_j\}}}{1-\frac{exp\{\beta_0+\beta_j\}}{1+exp\{\beta_0+\beta_j\}}} = \frac{exp\{\beta_0+\beta_j\}}{1} = exp\{\beta_0+\beta_j\}$$

- Untuk $x = 0$, digunakan

$$\frac{\pi(0)}{1-\pi(0)} = \frac{\frac{exp\{\beta_0\}}{1+exp\{\beta_0\}}}{1-\frac{exp\{\beta_0\}}{1+exp\{\beta_0\}}} = \frac{\frac{exp\{\beta_0\}}{1+exp\{\beta_0\}}}{\frac{1+exp\{\beta_0\}-exp\{\beta_0\}}{1+exp\{\beta_0\}}} = exp\{\beta_0\}$$

maka didapatkan persamaan (2.20) berikut

$$\begin{aligned}\Psi &= \frac{\frac{\pi(1)}{1-\pi(1)}}{\frac{\pi(0)}{1-\pi(0)}} \\ &= \frac{exp\{\beta_0+\beta_j\}}{exp\{\beta_0\}} \\ &= exp\{\beta_0+\beta_j-\beta_0\} \\ &= exp\{\beta_j\} \end{aligned} \quad (2.20)$$

2.7 Evaluasi Model

Evaluasi model bertujuan untuk menilai sejauh mana model regresi logistik mampu mengklasifikasikan data dengan benar berdasarkan kategori variabel respon. Kesalahan pengklasifikasian dapat diidentifikasi melalui tabel keputusan klasifikasi (confusion matrix), yang menggambarkan jumlah pengamatan yang diklasifikasikan dengan benar dan salah oleh model (Hosmer & Lemeshow, 2013). Menurut Sharda, dkk (2021) struktur umum confusion matrix seperti pada Tabel 1 berikut.

Tabel 1. Ketepatan Klasifikasi (Confusion Matrix)

Hasil Observasi	**Taksiran Klasifikasi**	
	$\boldsymbol{v_1}$	$\boldsymbol{v_2}$
$\boldsymbol{y_1}$	n_{11}	n_{12}
$\boldsymbol{y_2}$	n_{21}	n_{22}

Keterangan:

n_{11} = Banyaknya pengamatan dari $y_1$tepat dikategorikan v_1

n_{12} = Banyaknya pengamatan dari $y_1$tepat dikategorikan v_2

n_{21} = Banyaknya pengamatan dari $y_2$tepat dikategorikan v_1

n_{22} = Banyaknya pengamatan dari $y_2$tepat dikategorikan v_2

Berdasarkan confusion matrix tersebut, kinerja model dievaluasi menggunakan beberapa ukuran, yaitu akurasi, sensitivitas, dan spesifisitas. Akurasi mengukur proporsi keseluruhan pengamatan yang diklasifikasikan dengan benar, sensitivitas menunjukkan kemampuan model dalam mengenali kategori positif (y_1), sedangkan spesifisitas mengukur kemampuan model dalam mengenali kategori negatif (y_2) (Hosmer & Lemeshow, 2013). Selain itu, salah satu cara penting untuk mengklasifikasikan suatu objek adalah dengan mengukur taraf dari errornya atau tingkat kesalahannya (Johnson & Wichern, 2007). Tingkat kesalahan tersebut akan dihitung menggunakan APER. *Apparent Error Rate* (APER) merupakan proporsi sampel yang diklasifikasikan secara tidak benar. Berikut adalah persamaan untuk perhitungan masing-masing ukuran evaluasi.

$$Akurasi = \frac{n_{11} + n_{22}}{n} \times 100\% \tag{2.21}$$

$$Sensitivitas = \frac{n_{11}}{n_{11} + n_{12}} \times 100\% \tag{2.22}$$

$$Spesifisitas = \frac{n_{22}}{n_{21} + n_{22}} \times 100\% \tag{2.23}$$

$$APER = \frac{n_{21} + n_{12}}{n} \times 100\% \tag{2.24}$$

dimana n menyatakan jumlah keseluruhan pengamatan

Menurut Arisandi & Dewi (2023) menyatakan bahwa nilai evaluasi model yang diperoleh dari hasil perhitungan confusion matrix dapat diklasifikasikan ke dalam kategori tingkat performa model yang ditunjukkan pada Tabel 2 berikut.

Tabel 2 Kategori Tingkat Performa Model

Nilai (%)	**Kategori**
90.01 – 100.00	Sangat baik
80.01 – 90.00	Baik
70.01 – 80.00	Cukup baik
60.01 – 70.00	Buruk
≤ 60.00	Sangat buruk

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada semester ganjil tahun ajaran 2025/2026 bertempat di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yaitu dataset *Electronic Health Record* (EHR) yang merupakan *open access* data dari situs Kaggle dengan judul *"Patient Treatment Classification – Electronic Health Record Dataset"*. Kumpulan data ini berisi hasil rekam medis pasien yang dikumpulkan dari sebuah rumah sakit swasta di Indonesia, yang mencatat hasil pemeriksaan laboratorium harian pasien selama bulan Januari 2019. Link data EHR ini adalah https://www.kaggle.com/datasets/saurabhshahane/patient-treatment-classification.

Dataset terdiri dari 6 atribut yang mencakup data bersifat numerik dan kategorik, Variabel terikat (dependen) merupakan variabel yang dijadikan akibat dari variabel bebas (independen). Dalam penelitian ini variabel terikat (Y) yang digunakan adalah variabel *source* yaitu status perawatan pasien, sedangkan variabel- variabel bebas (X) yang digunakan dalam analisis tercantum pada tabel berikut.

Tabel 3. Variabel Data Penelitian

Variabel	Fitur	Keterangan	Satuan
X_1	HAEMOGLOBINS	Kadar hemoglobin (protein dalam sel darah merah)	g/dL
X_2	LEUCOCYTE	Jumlah leukosit (sel darah putih dalam darah)	ribu/μL
X_3	TROMBOCYTE	Jumlah trombosit (sel pembekuan darah)	ribu/μL
X_4	AGE	Usia pasien	tahun
X_5	SEX	Jenis kelamin pasien	boolean
Y	SOURCE	Status perawatan pasien: *in care* (rawat inap) atau *out care* (rawat jalan)	boolean

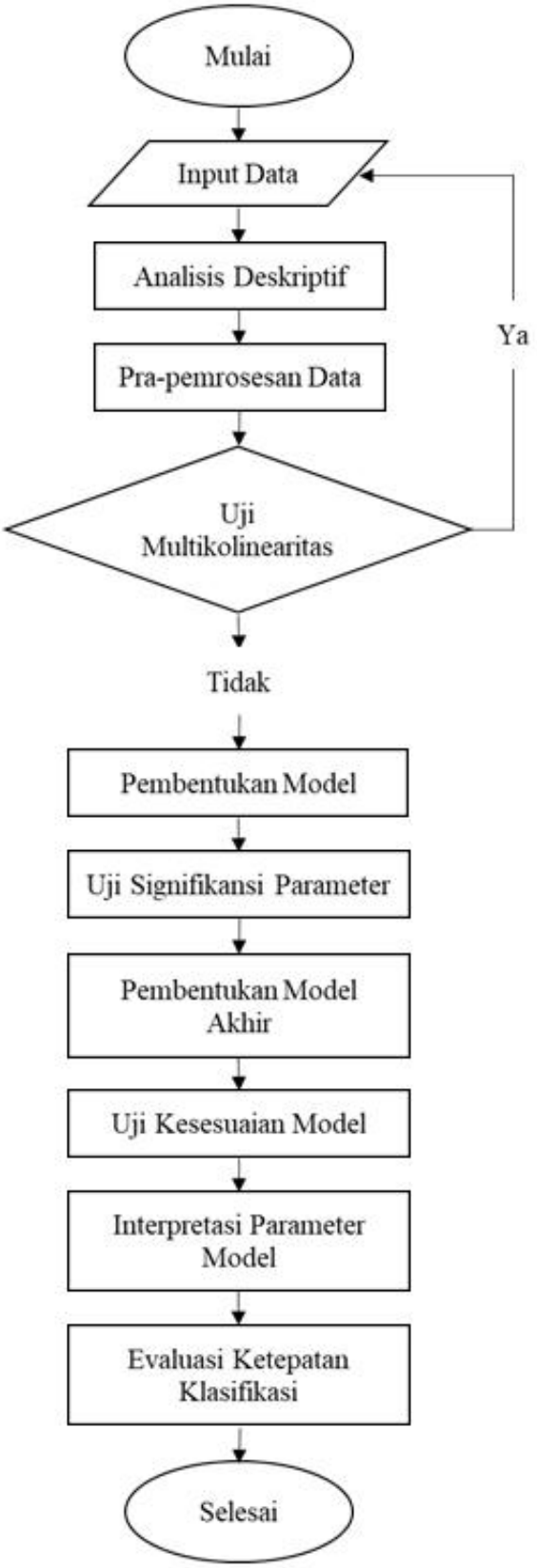
3.3 Metode Penelitian

Penelitian ini dilakukan dengan mengandalkan informasi yang sudah tersedia dalam literatur yang ada, seperti buku, karya ilmiah, dan jurnal. Untuk mencapai tujuan penelitian, penulis menggunakan perangkat lunak SAS Studio dalam proses perhitungan. Adapun langkah-langkah yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Melakukan penginputan data *Electronic Health Record* (EHR) yang diperoleh dari website Kaggle ke dalam perangkat lunak SAS Studio untuk dianalisis
2. Mendeskripsikan karakteristik data dengan statistika deskriptif untuk semua variabel, yaitu:
   - Varibel numerik yaitu, kadar hemoglobin sebagai X_1, jumlah leukosit sebagai X_2, jumlah trombosit sebagai X_3, dan usia pasien sebagai X_4,
   - Variabel kategorik yaitu, jenis kelamin pasien sebagai X_5 dan status perawatan pasien sebagai Y
3. Melakukan pra-pemrosesan data
   - Melakukan pembersihan data (*data cleaning*) untuk memeriksa *missing value* dan ditangani bila ada

- Melakukan transformasi variabel kategorik ke dalam bentuk variabel dummy agar dapat diolah

4. Melakukan pemeriksaan multikolinearitas untuk mengecek korelasi antar variabel bebas dengan menghitung nilai VIF, pastikan tidak ada korelasi tinggi antar variabel
5. Membangun model regresi logistik biner menggunakan metode maximum likelihood estimation (MLE), sehingga diperoleh fungsi peluang kejadian
6. Pengujian signifikansi parameter dengan uji simultan menggunakan Ratio Likelihood Test (Uji G) dan uji parsial dengan Wald Test
7. Membentuk model akhir regresi logistik biner yang hanya terdiri atas variabel-variabel yang signifikan
8. Menguji kesesuaian model menggunakan uji Hosmer-Lemeshow untuk mengetahui model yang terbentuk sesuai atau tidak dengan data pengamatan
9. Melakukan interpretasi parameter model dengan menggunakan Odds Ratio, sehingga hasil dapat ditafsirkan dengan mudah dalam konteks medis
10. Mengukur performa model dengan evaluasi ketepatan menggunakan tabel ketepatan klasifikasi dan menghitung kesalahan klasifikasi dari setiap model dengan menggunakan APER untuk menilai seberapa baik model dalam memprediksi status perawatan pasien



Gambar 1. Diagram Alir Metode Penelitian

V. KESIMPULAN

Berdasarkan hasil analisis regresi logistik biner terhadap data *Electronic Health Record* (EHR) untuk mengklasifikasikan status perawatan pasien (rawat inap atau rawat jalan) dengan variabel bebas hemoglobin, leukosit, trombosit, usia, dan jenis kelamin pasien, diperoleh beberapa kesimpulan sebagai berikut:

Model regresi logistik biner yang terbentuk memenuhi kriteria kesesuain model berdasarkan nilai AIC dan uji deviance. Model akhir menunjukkan bahwa status perawatan pasien dipengaruhi oleh kadar hemoglobin, jumlah leukosit, jumlah trombosit, dan jenis kelamin.

Hasil pengujian secara simultan menunjukkan bahwa seluruh variabel dalam model berpengaruh signifikan terhadap status perawatan pasien. Secara parsial, variabel hemoglobin, leukosit, trombosit, dan jenis kelamin berpengaruh signifikan terhadap peluang pasien dirawat inap, sedangkan variabel usia tidak menunjukkan pengaruh signifikan dan dikeluarkan dari model akhir.

Berdasarkan hasil estimasi probabilitas untuk kategori 1 (pasien dirawat inap) diperoleh bahwa peningkatan kadar hemoglobin dan jumlah trombosit berkaitan dengan penurunan peluang pasien dirawat inap, sedangkan peningkatan jumlah leukosit berkaitan dengan peningkatan peluang pasien dirawat inap. Selain itu, pasien laki-laki memiliki peluang rawat inap lebih tinggi dibandingkan pasien perempuan pada tingkat variabel yang sama. Namun, interpretasi klinis dari hasil ini perlu mempertimbangkan konteks medis yang luas.

Model regresi logistik biner yang dihasilkan memiliki tingkat ketepatan klasifikasi sebesar 72,19%, yang menunjukkan bahwa model mampu mengklasifikasikan status perawatan pasien apakah akan dirawat inap atau dirawat jalan dengan tingkat ketepatan yang cukup baik.

DAFTAR PUSTAKA

Agresti, A. 1996. *An Introduction to Categorical Data Analysis*. Toronto: John Wiley and Sons Inc.

Arisandi, R., & Dewi, A, L. 2023. Analisis Faktor Risiko Gagal Jantung Dengan Regresi Logistik Berbasis loMT. Jurnal Gaussian. **12**(4): 549-559.

Avini, Patunduk, K. W., Sumarni, Harbianti, Pratiwi, A., & Hidayat, R. 2022. Analisis Model Cox Proportional Hazard dan Regresi Logistik sebagai Upaya Pencegahan Covid-19 di Kota Palopo. *Inferensi*. **5**(2): 105-114.

Bawono, B., Utami, T. W., & Nur, I. M. 2019. Perbandingan Metode Regresi Logistik Biner dan Naive Bayes Dalam Klasifikasi Debitur Berdasarkan Kualitas Kerdit Nasabah. *Jurnal Ilmiah*.

Fadhal, M. 2022. Penggunaan Electronic Health Record (EHR) Dalam Keperawatan Jiwa: Literature Review. *Jurnal Ilmiah STIKES Citra Delima Bangka Belitung*. **5**(2): 113-124.

Ghozali, I. 2018. *Aplikasi Analisis Multivariat dengan Program SPSS*. Ed. ke-5. Universitas Diponegoro. Semarang.

Gujarati, D. 1995. *Ekonomi Dasar*. Diterjemahkan oleh Sumarno Zain. Erlangga. Jakarta.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. 2019 . *Multivariate Data Analysis.* 8th ed. Cengage Learning.

Hosmer, D.W. & Lemeshow, S. 2000. *Applied Logistics Regression*. 2th Edition. John Wiley & Sons. New York.

Hosmer, D.W., Lemeshow, S., & Sturdivant, X. R. 2013. *Applied Logistics Regression*. 3rd Edition. John Wiley & Sons. New York .

Indriyani, S., Raupong, & Anisa. 2020. Estimasi Parameter Regresi Logistik Biner Dengan Metode Partial Least Squares. 1-8.

Johnson, R. A. & Wichern, D. W. 2007. *Applied Multivariate Statistical Analysis*. Sixth Edition. Pearson Prentice Hall, New Jersey.

Kaggle. 2019. *Patient Treatment Classification – Electronic Health Record Dataset.* https://www.kaggle.com/datasets/saurabhshahane/patient-treatment-classification. Diakses pada 14 November 2024.

Martias, L. D. 2021. Statistika Deskriptif Sebagai Kumpulan Informasi. *Jurnal Ilmu Perpustakaan dan Informasi*. **16**(1): 40-59.

Purba, S. A. 2020. Estimasi Parameter Data Berdistribusi Normal Menggunakan Maksimum Likelihood Berdasarkan Newton Rapshon. *Jurnal Sains Dasar*. **9**(1): 16-18.

Prasetyo, R.A., dan Helma. 2022. Analisis Regresi Linear Berganda Untuk Melihat Faktor Yang Berpengaruh Terhadap Kemiskinan Di Provinsi Sumatera Barat. *Journal Of Mathematics UNP*. **7**(2): 62-68.

Ripai, M., Hayati, U., Widyawati, W., Susana, H., & Fathurrohma. 2022. Pengklasifikasian Surat Pemberitahuan Pajak Daerah Menggunakan Metode Regresi Logistik Biner Untuk Mengetahui Patuh dan Tidak Patuh Dalam Pembayaran Pajak Daerah. *Kopertip.* **6**(1): 27-33.

Rosita, R., Pertiwi, D.A.A., & Khairunnisa, O.G. 2022. Prediction of hospital intensive patients using neural network algorithm. *Journal of Soft Computing Exploration*. **3**(1): 8–11.

Sadikin, M. & Nurhaida, I. 2021. Exploratory Study of Some Machine Learning Techniques to Classify the Patient Treatment. *International Journal of Advanced Computer Science and Applications*. **12**(2): 380-387.

Sharda, R., Delen, D., & Turban, E. 2021. *Analytics, Data Science, & Artificial Intelligence Systems for Decision Support*. Pearson.