ESTIMASI MODEL REGRESI LOGISTIK BINER NONPARAMETRIK DENGAN PENAKSIR LOKAL LINIER DALAM MENGANALISIS FAKTOR PENYEBAB *PENDING* KLAIM BPJS KESEHATAN

SKRIPSI

A. ANNISA MIFTAHUL SAKINAH

H051181003

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

OKTOBER 2022

ESTIMASI MODEL REGRESI LOGISTIK BINER NONPARAMETRIK DENGAN PENAKSIR LOKAL LINIER DALAM MENGANALISIS FAKTOR PENYEBAB *PENDING* KLAIM BPJS KESEHATAN

SKRIPSI

Diajukan sebagai salah satu syarat memperoleh gelar Sarjana Sains pada Program Stutdi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin

A. ANNISA MIFTAHUL SAKINAH

H051181003

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

OKTOBER 2022

LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

Estimasi Model Regresi Logistik Biner Nonparametrik dengan Penaksir Lokal Linier Dalam Menganalisis Faktor Penyebab *Pending* Klaim BPJS Kesehatan

adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun

Makassar, 12 Oktober 2022

METERAI TEMPEL 6F4AKX060468122

A. Annisa Miftahul Sakinah
NIM H051181026

ESTIMASI MODEL REGRESI LOGISTIK BINER NONPARAMETRIK DENGAN PENAKSIR LOKAL LINIER DALAM MENGANALISIS FAKTOR PENYEBAB *PENDING* KLAIM BPJS KESEHATAN

Disetujui Oleh:

Pembimbing Utama

Dr. Anna Islamiyati, S.Si, M.Si.
NIP. 19770808 200501 2 002

Pembimbing Pertama

Dra. Nasrah Sirajang, M.Si
NIP. 1962101 5198810 1 001

Ketua Program Studi

KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN UNIVERSITAS HASANUDDIN DEPARTEMEN STATISTIKA FAKULTAS MIPA

Dr. Nurtiti Sunusi, S.Si., M.Si.
NIP. 19720117 199703 2 002

Pada 12 Oktober 2022

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama	: A. Annisa Miftahul Sakinah
NIM	: H051181003
Program Studi	: Statistika
Judul Skripsi	: Estimasi Model Regresi Logistik Biner Nonparametrik dengan Penaksir Lokal Linier Dalam Menganalisis Faktor Penyebab *Pending* Klaim BPJS Kesehatan

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin

DEWAN PENGUJI

1. Ketua : Dr. Anna Islamiyati, S.Si, M.Si. (............................)
2. Sekretaris : Dra. Nasrah Sirajang, M.Si. (............................)
3. Anggota : Drs. Raupong, M.Si. (............................)
4. Anggota : Siswanto, S.Si., M.Si. (............................)

Ditetapkan di : Makassar

Tanggal : 12 Oktober 2022

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Segala puji hanya milik Allah *Subhanallahu Wa Ta'ala* atas segala limpahan rahmat dan hidayah-Nya yang telah diberikan kepada penulis sampai saat ini. Shalawat dan salam senantiasa tercurahkan kepada baginda Rasulullah *Shallallahu 'Alaihi Wa sallam*. *Alhamdulillaahirobbil'aalamiin*, berkat rahmat dan kemudahan yang diberikan oleh Allah *Subhanallahu Wa Ta'ala*, penulis dapat menyelesaikan tugas akhir yang berjudul "**Estimasi Model Regresi Logistik Biner Nonparametrik dengan Penaksir Lokal Linier Dalam Menganalisis Faktor Penyebab *Pending* Klaim BPJS Kesehatan**" sebagai salah satu syarat memperoleh gelar sarjana pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Dalam penyelesaian skripsi ini, penulis telah melewati perjuangan panjang dan pengorbanan yang tidak sedikit. Namun berkat rahmat dan izin-Nya serta dukungan dari berbagai pihak yang turut membantu sehingga akhirnya tugas akhir ini dapat terselesaikan di waktu yang terbaik menurut Allah SWT. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang setinggi-tingginya dan penghargaan yang tak terhingga kepada **Ayahanda Firdaus** dan **Ibunda Fujianti Hasan** yang telah membesarkan dan mendidik penulis dengan penuh kesabaran dengan limpahan cinta, kasih sayang dan doa yang tak henti dilangitkan kepada penulis, adik-adik tercinta penulis yaitu **A. Farid Murtadhi Pairing** dan **A. Qatrunnada** yang selalu menyemangati dan mendoakan penulis, serta seluruh keluarga besar penulis yang selalu mendoakan dan memberikan bantuan baik dalam bentuk moral ataupun material.

Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan juga penulis ucapkan kepada:

1. **Bapak Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.**, selaku Rektor Universitas Hasanuddin berserta seluruh jajarannya.
2. **Bapak Dr. Eng. Amiruddin, S.Si., M.Si**, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.

3. **Ibu Dr. Nurtiti Sunusi S.Si., M.Si.**, selaku Ketua Departemen Statistika, segenap Dosen Pengajar dan Staf yang telah membekali ilmu dan kemudahan kepada penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika.
4. **Ibu Dr. Anna Islamiyati, S.Si., M.Si.** selaku pembimbing utama sekaligus penasehat akademik penulis yang telah ikhlas meluangkan waktu dan pemikirannya untuk memberikan arahan, pengetahuan, motivasi dan bimbingan di tengah kesibukannya.
5. **Ibu Dra. Nasrah Sirajang, M.Si.** selaku pembimbing pertama penulis yang telah meluangkan waktunya di tengah kesibukan untuk memberikan arahan, pengetahuan, motivasi, bimbingan untuk penulis.
6. **Bapak Drs. Raupong, M.Si.** dan **Bapak Siswanto, S.Si., M.Si.** selaku tim penguji yang telah memberikan saran dan kritikan yang membangun dalam penyempurnaan penyusunan tugas akhir ini.
7. Sahabat tercinta penulis, **Alfiana Wahyuni**, **Fitra Damayanti**, **Naura Alfatiyya Arda**, **Marsya Anggun Prisila** dan **Nurul Rezki** yang selalu ada dalam setiap keadaan. Terimakasih telah menjadi sahabat terbaik yang senantiasa mendoakan, mendengarkan keluhan, memberikan dorongan, semangat dan motivasi sejak mahasiswa baru.
8. Sahabat "KOSER", **Akidah Amaliah** dan **Juni Wahdaniyah** yang juga selalu menemani penulis dalam setiap keadaan, memberikan motivasi, doa dan semangat dalam mengerjakan tugas akhir ini.
9. Teman-teman **INTEGRAL 2018**, terkhusus kepada **Yustika**, **Fadhil Al-Anshory**, **Fernando Toding Bua, Nur Anisa Syahbani Salim**, **Amalia Andrianingrum**, **Muh. Ishak**, **Abdul Jalil Saleh**, **Muh. Lutfi**, **Ardi S**, **Muh. Ainun Luthfi**, **Fuad Hamdi Bahar** dan **Ahmad Ilham B**. Terimakasih telah memberikan warna dalam dunia perkuliahan dan mengajarkan arti persaudaraan. Pengalaman berharga telah penulis dapatkan dari teman-teman selama berproses bersama.
10. Teman-teman **Statistika 2018** terima kasih atas kebersamaan selama menjalani pendidikan di Departemen Statistika. Terkhusus kepada **Sri Indriani Amil**, **Hajratul Ashwad K**, **Nur Hidayah L**, **Taufiq Akbar**, **Nur**

Anugrah Yusuf, **Adhiyaksa Prananda RS** dan **La Ade** yang selalu membantu dan menjadi sosok guru bagi penulis.

11. Keluarga besar **HIMASTAT FMIPA UNHAS** dan **HIMATIKA FMIPA UNHAS**, terimakasih atas seluruh pengalaman, pembelajaran serta telah menjadi keluarga penulis selama masa perkuliahan.
12. **Keluarga Mahasiswa FMIPA Unhas**, terimakasih untuk cerita, pengalaman dan ilmu yang sangat berharga selama penulis berproses di KMF. Terkhusus kepada **Pengurus BEM FMIPA Unhas Periode 2021/2022** semoga selalu dengan slogan "Takkan Pudar".
13. Teman-teman **KKN Unhas Gelombang 106 wilayah Sinjai 2,** terimakasih untuk setiap bentuk kebaikan teman-teman dan cerita KKN yang luar biasa.
14. Kepada seluruh pihak yang tidak dapat penulis sebutkan satu persatu, terima kasih setinggi-tingginya untuk segala dukungan dan partisipasi yang diberikan kepada penulis semoga bernilai ibadah di sisi Allah *Subhanahu Wa Ta'ala*.

Penulis menyadari bahwa masih banyak kekurangan dalam tugas akhir ini, untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga tulisan ini memberikan manfaat untuk pembaca.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Makassar, 12 Oktober 2022

A. Annisa Miftahul Sakinah

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIK

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama	: A. Annisa Miftahul Sakinah
NIM	: H051181003
Program Studi	: Statistika
Departemen	: Statistika
Fakultas	: Matematika dan Ilmu Pengetahuan Alam
Jenis Karya	: Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non-eksklusif** (***Non-exclusive Royalty- Free Right***) atas tugas akhir saya yang berjudul:

"Estimasi Model Regresi Logistik Biner Nonparametrik dengan Penaksir Lokal Linier Dalam Menganalisis Faktor Penyebab *Pending* Klaim BPJS Kesehatan"

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.
Dibuat di Makassar pada tanggal, 12 Oktober 2022

Yang menyatakan

(A. Annisa Miftahul Sakinah)

ABSTRAK

Regresi logistik biner nonparametrik dengan penaksir lokal linier merupakan metode regresi logistik biner dengan pendekatan nonparametrik yang menggunakan penaksir lokal linier dalam menganalisis hubungan antara satu variabel respon dengan satu atau lebih variabel prediktor yang bersifat kategorik atau kontinu. Penelitian ini bertujuan untuk mendapatkan estimasi model regresi logistik biner nonparametrik dengan penaksir lokal linier dan mengetahui faktor-faktor yang berpengaruh terhadap penyebab *pending* klaim BPJS Kesehatan di Rumah Sakit Universitas Hasanuddin. *Maximum Likelihood Estimation* digunakan untuk mengestimasi parameter regresi logistik dengan menggunakan iterasi Newton Raphson. Hasil yang diperoleh yaitu estimasi model regresi logistik biner nonparametrik dengan penaksir lokal linier yang bersesuaian dengan data pengajuan Rumah Sakit Universitas Hasanuddin ke BPJS Kesehatan adalah $\hat{g}(x) = -16{,}46242 - 0{,}98638(x_1 - 71,9612) + 1{,}63252(x_2 - 15{,}9999)$. Berdasarkan model yang diperoleh maka faktor-faktor yang mempengaruhi penyebab *pending* klaim BPJS Kesehatan adalah umur dan harga klaim.

Kata Kunci: Logistik Biner Nonparametrik, Penaksir Lokal Linier, *Pending* Klaim, BPJS Kesehatan, *Maximum Likelihood Estimation*.

ABSTRACT

Nonparametric binary logistic regression with linear local estimator is a binary logistic regression method with nonparametric approach that uses a linear local estimator in analyzing the relationship between one response variable and one or more predictor variables which are categorical or continuous. This study aims to obtain an estimation of a nonparametric binary logistic regression model with a linear local estimator and to determine the factors that influence the causes of pending BPJS Health claims at Hasanuddin University Hospital. Maximum Likelihood Estimation uses to estimate logistic regression parameters using Newton Raphson iteration. The results obtained are tehe estimation of a nonparametric binary logistic regression model with a linear local estimator that correponds to the data submitted by Hasanuddin University Hospital to BPJS Kesehatan is $\hat{g}(x) = -16{,}46242 - 0{,}98638(x_1 - 71{,}9612) + 1{,}63252(x_2 - 15{,}9999)$. *Based on the model obtained, the factors that influence the cause of pending BPJS Kesehatan claims are age and claim price.*

Keywords*: Binary Logistics Nonparametric, Linear Local Estimator, Claim Pending, BPJS Kesehatan, Maximum Likelihood Estimation.*

DAFTAR ISI

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang

Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan merupakan badan hukum yang dibentuk untuk menyelenggarakan program jaminan kesehatan masyarakat. BPJS Kesehatan sama halnya dengan asuransi kesehatan yang bekerjasama dengan fasilitas kesehatan seperti dokter, klinik dan rumah sakit (Arikusnadi dkk., 2020). Pada penyelenggaraan asuransi terdapat istilah klaim, yaitu tagihan atau tuntutan atas sebuah imbalan dari hasil layanan yang diberikan.

Syarat berkas pengajuan klaim ke BPJS Kesehatan dari rumah sakit adalah menyertakan *resume* medis. Pengajuan klaim dapat dikembalikan oleh BPJS yang mengakibatkan pengajuan klaim tersebut ditunda (*pending*) sehingga pelayanan kesehatan yang dilakukan rumah sakit tidak terbayar dan akan mengakibatkan menurunnya pemasukan rumah sakit (EP, 2018). Hal tersebut mengakibatkan *cash flow* rumah sakit terganggu karena hampir 90% pasien rumah sakit adalah peserta BPJS Kesehatan (Nabila dkk., 2020). Berdasarkan permasalahan tersebut, diperlukan strategi untuk mengetahui faktor-faktor yang berpengaruh terhadap *pending* klaim yang diajukan oleh rumah sakit ke BPJS Kesehatan, salah satunya adalah dengan menggunakan analisis regresi.

Analisis regresi merupakan suatu metode statistika yang digunakan untuk menggambarkan hubungan antara variabel respon dengan satu atau lebih variabel prediktor. Salah satu jenis analisis regresi yang sering digunakan adalah regresi logistik (Hosmer dan Lemeshow, 2013). Regresi logistik merupakan suatu metode yang menjelaskan hubungan antara variabel respon pada skala kategorik dengan satu atau lebih variabel prediktor pada skala kategorik atau kontinu (Agresti, 2002). Terdapat dua sifat variabel respon pada regresi logistik yang dibagi berdasarkan jumlah kategorinya, yaitu *dichotomous* (dua kategori) dan *polychotomous* (lebih dari dua kategori) (Naim, 2020). Regresi logistik terbagi menjadi regresi logistik biner, regresi logistik multinomial dan regresi logistik ordinal. Regresi logistik biner merupakan teknik analisis data yang digunakan dalam menganalisis hubungan antara satu variabel respon yang memiliki data kualitatif bersifat *dichotomous* (dua kategori) yaitu bernilai 1 untuk ‘sukses’ dan bernilai 0 untuk ‘gagal’ dengan satu atau lebih variabel prediktor (Tampil dkk., 2017).

Regresi logistik dapat diestimasi melalui dua pendekatan yaitu pendekatan parametrik dan nonparametrik. Pendekatan nonparametrik digunakan jika bentuk hubungan antara variabel respon dengan variabel prediktor tidak mengikuti pola tertentu atau tidak ada informasi apapun tentang bentuk dari fungsi regresi. Pendekatan regresi nonparametrik memiliki fleksibilitas yang besar karena fungsi regresi tidak ditentukan dalam bentuk tertentu tetapi diasumsikan *smooth* sehingga dapat diestimasi dengan menggunakan beberapa penaksir, yaitu lokal linier, *spline least square*, *penalized spline* dan deret fourier (Eubank, 1999).

Beberapa penelitian yang telah menggunakan pendekatan regresi logistik nonparametrik diantaranya Istisadah (2017) memperoleh ketepatan klasifikasi model regresi logistik biner nonparametrik lebih unggul yaitu sebesar 70% dibandingkan dengan ketepatan klasifikasi model regresi logistik biner parametrik yaitu sebesar 66.7% pada kasus data pasien diabetes mellitus tipe 2 di RSU Haji Surabaya tahun 2016. Naim (2020) tentang model regresi logistik biner nonparametrik dengan estimator *spline truncated* pada data status gizi balita diperoleh ketepatan klasifikasi model sebesar 87.5%. Penelitian lain terkait lokal linier dilakukan oleh Nurfitriani (2021) tentang pemodelan banyaknya penambahan kasus positif covid-19 di Jawa Timur dengan pendekatan regresi binomial negatif berdasarkan estimator lokal linier diperoleh nilai *deviance* dengan pendekatan nonaparmetrik lebih unggul yaitu sebesar 59.13 dibandingkan dengan nilai *deviance* dengan pendekatan parametrik yaitu sebesar 82.50.

Pada penelitian ini menggunakan regresi logistik biner nonparametrik dengan penaksir lokal linier. Penaksir lokal linier adalah suatu penaksir yang menggunakan pembobot fungsi kernel dalam melakukan estimasi, dengan ukuran bobot sendiri ditentukan oleh parameter h atau *bandwidth* (Tilova, 2019). Kelebihan dari penaksir lokal linier adalah kemampuan dalam mengestimasi fungsi di setiap titik sehingga hasil estimasi yang diperoleh lebih mendekati pola data sebenarnya (Nottingham dan Cook, 2011).

Berdasarkan uraian yang telah dipaparkan, penulis akan mengkaji estimasi model regresi logistik biner nonparametrik dengan penaksir lokal linier. Metode diaplikasikan pada data pengajuan klaim Rumah Sakit Universitas Hasanuddin ke BPJS Kesehatan.

1.2 Rumusan Masalah

Berdasarkan latar belakang, maka rumusan masalah dari penelitian ini adalah sebagai berikut:

1. Bagaimana bentuk estimasi model regresi logistik biner nonparametrik dengan penaksir lokal liner dalam menganalisis faktor penyebab *pending* klaim BPJS kesehatan?
2. Faktor-faktor apa saja yang berpengaruh terhadap *pending* klaim BPJS kesehatan berdasarkan model regresi logistik biner nonparametrik dengan penaksir lokal liner?

1.3 Batasan Masalah

Batasan masalah dari penelitian ini adalah sebagai berikut:

1. Fungsi kernel yang digunakan, yaitu fungsi kernel *Gaussian*.
2. Data yang digunakan terdiri dari satu variabel respon yang berskala biner dan tiga variabel prediktor yaitu umur, harga klaim dan *Length of Stay* (LOS).

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan dari penelitian ini adalah sebagai berikut:

1. Mendapatkan estimasi model regresi logistik biner nonparametrik dengan penaksir lokal liner dalam menganalisis faktor penyebab *pending* klaim BPJS Kesehatan.
2. Mendapatkan faktor yang berpengaruh terhadap *pending* klaim BPJS kesehatan berdasarkan model regresi logistik biner nonparametrik dengan penaksir lokal linier.

1.5 Manfaat Penelitian

Penelitian ini diharapkan bermanfaat untuk berbagai pihak baik peneliti, mahasiswa dan instansi. Manfaat yang diharapkan yaitu:

1. Menambah wawasan dan pengetahuan mengenai estimasi model regresi logistik biner nonparametrik dengan penaksir lokal liner dalam menganalisis faktor penyebab *pending* klaim BPJS Kesehatan.

2. Sebagai bahan rujukan bagian *Case-mix* di Rumah Sakit Universitas Hasanudin dalam mengetahui faktor penyebab *pending* klaim oleh BPJS Kesehatan.

BAB II
TINJAUAN PUSTAKA

2.1 Regresi Logistik Biner

Regresi logistik merupakan metode yang dapat digunakan untuk mencari hubungan antara satu variabel dependen (respon) yang dilambangkan dengan y yang bersifat *dichotomus* atau *polychotomous* dengan satu atau lebih variabel independen (prediktor) yang dilambangkan dengan x (Sofha dkk., 2015). Terdapat beberapa jenis regresi logistik seperti regresi logistik biner, multinomial dan ordinal. Apabila variabel respon terdiri dari dua kategori maka metode regresi logistik yang dapat digunakan adalah regresi logistik biner (Hosmer dan Lemeshow, 2013).

Model dalam regresi logistik biner termasuk dalam distribusi keluarga Eksponensial, distribusi Eksponensial yang dimaksud adalah distribusi Bernoulli, yaitu distribusi peubah acak yang hanya mempunyai dua fungsi kategorik yaitu bernilai 0 dan 1 (Islamiyati, 2015). Variabel respon pada regresi logistik biner terdiri dari 2 kategori, misalkan y = 1 menyatakan hasil yang diperoleh 'sukses' dan y = 0 menyatakan hasil yang diperoleh 'gagal'. Variabel respon y mengikuti distribusi Bernoulli karena variabel respon y hanya memiliki dua kategori (Hosmer dan Lemeshow, 2013) dengan fungsi peluang distribusi Bernoulli:

$$f\left(y_i\right)=\pi\left(x_i\right)^{y_i}\left(1-\pi(x_i)\right)^{(1-y_i)} \tag{2.1}$$

jika $y_i = 0$ maka $f\ (y_i) = (1 - \pi(x_i)$ dan jika $y_i = 1$ maka $f(y_i) = \pi(x_i)$.
dengan:

$f(y_i)$ = fungsi distribusi bernoulli
y_i = variabel respon pada pengamatan ke-i
$\pi(x_i)$ = peluang kejadian sukses pada pengamatan ke-i
$1 - \pi(x_i)$ = peluang kejadian gagal pada pengamatan ke-i

Model regresi logistik dengan k variabel prediktor dinyatakan sebagai berikut:

$$\pi\left(x_i\right)=\frac{\exp\left(\beta_0+\sum_{j=1}^{k}\beta_j\left(x_{ij}\right)\right)}{1+e\exp\left(\beta_0+\sum_{j=1}^{k}\beta_j\left(x_{ij}\right)\right)} \tag{2.2}$$

dengan $i = 1,2,\dots,n$ dan $j = 1,2,\dots,k$.

Fungsi $\pi(x_i)$ pada Persamaan (2.2) merupakan fungsi non-linear sehingga transformasi logit diperlukan untuk memperoleh fungsi yang linear agar dapat dilihat hubungan antara variabel respon dan variabel prediktornya, maka diperoleh persamaan berikut:

$$g\left(x_i\right)=\ln\left[\frac{\pi\left(x_i\right)}{1-\pi\left(x_i\right)}\right]=\beta_0+\beta_1 x_1+\cdots+\beta_k x_{ik} \tag{2.3}$$

dengan:

$\pi\left(x_i\right)$	= peluang kejadian sukses pada pengamatan ke-i
$g\left(x_i\right)$	= nilai estimasi logit
β_0	= intersep
$\beta_1,\beta_2,\dots,\beta_k$	= parameter koefisien pada variabel prediktor ke-k
x_{ik}	= nilai pada variabel prediktor ke-k pada pengamatan ke-i

2.2 Estimasi Parameter

Estimasi parameter dalam regresi logistik dilakukan dengan metode *Maximum Likelihood Estimation* (MLE). Metode tersebut mengestimasi parameter $\boldsymbol{\beta}$ dengan cara memaksimumkan fungsi *likelihood* dan mensyaratkan bahwa data harus mengikuti suatu distribusi tertentu. Pada regresi logistik biner, setiap pengamatan mengikuti distribusi Bernoulli sehingga dapat ditentukan fungsi *likelihood*nya. Fungsi *likelihood* dapat dinyatakan sebagai berikut:

$$\begin{aligned} L(\beta) &= \prod_{i=1}^{n} f(y_i) \\ &= \prod_{i=1}^{n} \pi(x_i)^{y_i}\left(1-\pi(x_i)\right)^{(1-y_i)} \end{aligned} \tag{2.4}$$

Fungsi *likelihood* pada Persamaan (2.4) lebih mudah dimaksimumkan dalam bentuk logaritma natural yang kemudian disebut fungsi *log-likelihood* dan dinyatakan dengan ℓ.

$$\begin{aligned}\ell(\beta) &= \ln L(\beta)\\ &= \ln\left(\prod_{i=1}^{n}\pi(x_i)^{y_i}\left(1-\pi(x_i)\right)^{(1-y_i)}\right)\\ &= \sum_{i=1}^{n}\pi(x_i)^{y_i}\left(1-\pi(x_i)\right)^{(1-y_i)} \qquad (2.5)\end{aligned}$$

Memaksimumkan fungsi *log-likelihood* dapat diperoleh dengan menurunkan ℓ terhadap masing-masing parameter dan menyamakan dengan nol. Selanjutnya karena hasil turunan pertama fungsi *log-likelihood* terhadap masing-masing parameter tidak dapat diselesaikan secara analitik, maka pada penaksiran ini akan dilakukan pendekatan lain untuk mendapatkan nilai taksirannya. Nilai parameter β dari turunan pertama fungsi ℓ didapatkan melalui suatu prosedur iterative yang dilakukan dengan metode iterasi Newton Raphson yaitu maksimumkan fungsi *likelihood* (Agresti, 2002).

Metode Newton Raphson merupakan metode iterasi untuk menyelesaikan persamaan nonlinear. Berdasarkan iterasi Newton Raphson akan diperoleh turunan kedua dari fungsi *log-likelihood* terhadap masing-masing parameter. Formulasi iterasi Newton Raphson dengan memulai iterasi $t = 0$ sebagai berikut:

$$\hat{\boldsymbol{\beta}}_{(t+1)} = \quad \boldsymbol{\beta}_{(t)} - \left(\boldsymbol{H}_{(t)}\right)^{-1}\boldsymbol{v}_{(t)}$$

dengan:

t	=	iterasi ke-0,1,2, ... T
$\boldsymbol{\beta}_{(t)}$	=	Nilai taksiran awal iterasi ke- t
$\boldsymbol{v}_{(t)}$	=	matriks turunan pertama terhadap masing-masing parameter $\boldsymbol{\beta}$
$\boldsymbol{H}_{(t)}$	=	matriks turunan kedua terhadap masing-masing parameter $\boldsymbol{\beta}$

Proses iterasi dengan menggunakan metode Newton Raphson akan berhenti jika terpenuhi kondisi konvergen, yaitu ketika $\left\|\beta_{(t+1)} - \beta_{(t)}\right\| \leq \varepsilon$, dengan ε adalah bilangan positif yang sangat kecil (Ehlers, 2002).

2.3 Uji Signifikansi Parameter

Uji yang digunakan untuk melakukan pengujian model regresi logistik adalah uji simultan kemudian dilanjutkan dengan uji parsial.

2.3.1. Uji Simultan

Uji simultan dilakukan untuk mengetahui pengaruh dari variabel prediktor terhadap variabel respon secara bersama-sama. Hipotesis untuk uji ini adalah sebagai berikut:

H_0: $\beta_1 = \beta_2 = \cdots = \beta_k = 0$ (tidak ada pengaruh variabel prediktor secara serentak terhadap variabel respon), dengan $j = 1, 2, \ldots, k$.

H_1: paling sedikit ada $\beta_j \neq 0$ (ada pengaruh paling sedikit satu variabel prediktor secara simultan terhadap variabel respon).

dengan β_j merupakan variabel prediktor ke-j

Statistik uji yang digunakan untuk uji simultan adalah Uji G (*Likelihood Ratio Test*):

$$G \quad = -2\ln\left[\frac{L_0\left(\beta_0\right)}{L_p\left(\beta_j\right)}\right] = 2\left[\ln L_0\left(\beta_0\right) - L_p\left(\beta_j\right)\right] \tag{2.6}$$

dengan:

$L_0\left(\beta_0\right)$ = nilai fungsi *likelihood* tanpa variabel prediktor

$L_p\left(\beta_j\right)$ = nilai fungsi *likelihood* dengan variabel prediktor

Kriteria pengambilan keputusan adalah tolak H_0 jika nilai G lebih besar dari $\chi^2_{(db;\alpha)}$ atau $p - value$ kurang dari α yang berarti variabel prediktor secara bersama-sama mempengaruhi variabel respon (Hariadi dan Sulantari, 2017).

2.3.2. Uji Parsial Satu-Satu

Uji parsial satu-satu merupakan pengujian yang dilakukan untuk mengetahui pengaruh setiap variabel prediktor secara individu terhadap variabel respon dan hasil pengujian secara parsial digunakan untuk menunjukkan suatu variabel prediktor layak untuk masuk dalam model atau tidak, untuk menguji signifikansi parameter digunakan Uji *Wald*. Hipotesis yang digunakan adalah sebagai berikut:

H_0: $\beta_j = 0$, dengan $j = 1,2, \ldots k$ (tidak ada pengaruh variabel prediktor ke-j secara parsial terhadap variabel respon).

H_1: $\beta_j \neq 0$, dengan $j = 1,2, \ldots k$ (ada pengaruh variabel prediktor ke-j secara parsial terhadap variabel respon).

Statistik uji *Wald* (W):

$$W \quad = \frac{\beta_j}{SE\left(\beta_j\right)} \tag{2.7}$$

dengan:

β_j = penduga parameter β_j

$SE\left(\beta_j\right)$ = penduga standar error dari β_j

Hipotesis H_0 ditolak jika $|W_j|$ lebih besar dari $Z_{\frac{\alpha}{2}}$ atau nilai $p-value$ kurang dari α, artinya variabel prediktor mempengaruhi variabel respon (Hariadi dan Sulantari, 2019).

2.4 Regresi Nonparametrik

Regresi nonparametrik merupakan suatu metode statistika yang digunakan untuk mengetahui pola hubungan antara variabel respon dan prediktor, dengan bentuk pola hubungan antara variabel respon dan prediktor tidak diketahui atau tidak diperoleh informasi sebelumnya. Dengan asumsi bahwa y adalah variabel respon dan x adalah variabel prediktor untuk n pengamatan, maka hubungan antara variabel-variabel tersebut dapat dinyatakan sebagai berikut:

$$y_i \quad = m(x_i) + \varepsilon_i, i = 1,2,\dots,n \tag{2.8}$$

dengan ε_i disebut juga *error random* yang diasumsikan independen dengan *mean* nol dan masing-masing variansi σ^2 sedangkan $m(x_i)$ merupakan fungsi regresi yang akan diestimasi atau tidak diketahui bentuknya dan diasumsikan mulus (*smooth*) dalam artian berada di dalam suatu ruang fungsi sehingga lebih menjamin fleksibilitas dalam mengestimasi fungsi regresinya. Pendekatan nonparametrik memberikan keluasan pada data untuk menemukan sendiri pola kurva regresinya, sehingga sangat fleksibel dan objektif (Eubank, 1999). Sebagai perluasan dari Persamaan (2.8), hubungan antara variabel respon y dan variabel multiprediktor adalah sebagai berikut:

$$y_i \quad = \sum_{j=1}^{k} m_j(x_{ij}) + \varepsilon_i, i = 1,2,\dots,n$$

Menurut Hastie dan Tibshirani (1990) dengan menggunakan *Generalize Additive Models* (GLM) diperoleh secara umum bentuk model regresi logistik biner nonparametrik adalah sebagai berikut:

$$\pi\left(x_i\right) = \frac{\exp\left(\sum_{j=1}^{k} m_j\left(x_{ij}\right)\right)}{1+\exp\left(\sum_{j=1}^{k} m_j\left(x_{ij}\right)\right)} \tag{2.9}$$

2.5 Fungsi Kernel

Salah satu pembobotan yang digunakan untuk mendapatkan estimasi adalah Fungsi Kernel (Eubank, 1988). Fungsi kernel K dengan *bandwidth* (h) didefinisikan sebagai berikut:

$$K_h\left(x\right) = \frac{1}{h}K\left(\frac{x}{h}\right)$$

Estimator fungsi densitas kernel pada Persamaan (2.7) mempunyai beberapa jenis fungsi kernel univariate yang digunakan (Hardle, 1990):

1. Kernel *Uniform* = $K(x) = \frac{1}{2}I\left(|x| \le 1\right)$
2. Kernel *Triangle* = $K(x) = \left(1-|x|\right)I\left(|x| \le 1\right)$
3. Kernel *Cosinus* = $K(x) = \frac{\pi}{4}\cos\left(\frac{\pi}{2}x\right)I\left(|x| \le 1\right)$
4. Kernel *Epanechnikov* = $K(x) = \frac{3}{4}\left(1-x^2\right)I\left(|x| \le 1\right)$
5. Kernel *Gaussian* = $K(x) = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{1}{2}\left(-x^2\right)\right), -\infty < x < \infty$
6. Kernel *Kudratik* = $K(x) = \frac{15}{16}\left(1-x^2\right)^2 I\left(|x| \le 1\right)$
7. Kernel *Triweight* = $K(x) = \frac{35}{32}\left(1-x^2\right)^2 I\left(|x| \le 1\right)$

Sehingga fungsi kernel K dengan *bandwidth* (h) didefinisikan sebagai berikut:

$$K_h\left(x-x_i\right) = \prod_{j=1}^{k}\frac{1}{h_j}K\left(\frac{x_{ij}-x_{0j}}{h_j}\right)$$

dengan:

K	= fungsi kernel
h_j	= *bandwidth* pada variabel prediktor ke-j
x_{ij}	= nilai pengamatan ke-i pada variabel prediktor ke-j
x_{0j}	= nilai titik lokal pada variabel prediktor ke-j

2.6 Penaksir Lokal Linier

Penaksir lokal linier merupakan kasus khusus dari polinomial lokal yang memiliki derajat satu. Penaksir lokal linier adalah penaksir dalam fungsi nonparametrik. Penaksir lokal linier memiliki keunggulan karena fleksibilitasnya terhadap fungsi atau data disetiap titik sehingga model yang didapatkan lebih mendekati pola data yang sesungguhnya (Nottingham dan Cook 2001). Berdasarkan penaksir lokal linier, fungsi regresi $m(x_i)$ pada Persamaan (2.8) dapat didekati secara lokal dengan ekspansi Taylor pada x disekitar titik x_0 sebagai berikut (Fan dan Gijbels, 1996):

$$m(x) = \frac{f(x_0)}{0!} + \sum_{s=1}^{p} (x_i - x_0)^s \frac{f^{(s)}(x_0)}{s!} \tag{2.10}$$

dengan $s!$ Melambangkan faktorial dari s dan $f^{(s)}(x_0)$ melambangkan nilai dari turunan ke-s dari f pada titik x_0.

Jika $\beta_s = \frac{f^{(s)}(x_0)}{s!}$, maka ekspansi Taylor pada Persamaan (2.10) untuk lokal linier berderajat satu ($p = 1$) adalah sebagai berikut:

$$\begin{aligned} m(x) &= \beta_0 + \sum_{s=1}^{1} (x_i - x_0)^1 \beta_1 \\ &= \beta_0 + (x - x_0)^1 \beta_1 \\ &= \beta_0 + (x_i - x_0)\beta_1 \end{aligned} \tag{2.11}$$

Persamaan (2.11) dapat dinyatakan dalam bentuk matriks menjadi:

$$m(x) = \mathbf{x\beta}$$

dengan:

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \text{dan } \mathbf{x} = \begin{bmatrix} 1 & (x_i - x_0) \end{bmatrix}$$

Fungsi regresi tersebut hanya berlaku untuk satu variabel prediktor saja, sedangkan untuk fungsi regresi nonparametrik multiprediktor berdasarkan penaksir lokal linier diperoleh dari perluasan Persamaan (2.11) yaitu:

$$\sum_{j=1}^{k} m_j\left(x_{ij}\right) = \beta_0 + \left(x_{i1} - x_{01}\right)\beta_1 + \cdots + \left(x_{ik} - x_{0k}\right)\beta_k \quad (2.12)$$

dengan

$m_j\left(x_{ij}\right)$	= fungsi regresi logistik biner nonparametrik dengan penaksir lokal linier
β_0	= intersep
$\beta_1, \beta_{2,}, \ldots, \beta_k$	= parameter lokal linier pada variabel prediktor ke-j
x_{0j}	= titik lokal pada variabel prediktor ke-j
x_{ik}	= Nilai pengamatan ke- i pada variabel prediktor ke- k

2.7 Penaksir *Maximum Likelihood* dengan Lokal Terboboti

Salah satu metode yang dapat digunakan untuk mengestimasi fungsi regresi $m(x)$ pada titik x_0 adalah dengan penaksir *maximum likelihood* dengan lokal terboboti. Fungsi lokal *likelihood* terboboti dalam regresi nonparametrik berdasarkan penaksir lokal linier mempunyai persamaan sebagai berikut (Fan dkk., 1998):

$$L(\beta) = \prod_{i=1}^{n} f\left(y_i\right)^{K_h\left(x_i - x_0\right)} \quad (2.13)$$

Untuk bentuk fungsi lokal *likelihood* terboboti pada kasus multiprediktor yang merupakan perluasan dari Persamaan (2.13) dapat ditulis sebagai berikut:

$$\begin{aligned} L(\beta) &= \prod_{i=1}^{n} f\left(y_i\right)^{K_h\left(x_i - x_0\right)} \\ &= \prod_{i=1}^{n}\left[\pi\left(x_i\right)^{y_i}\left(1-\pi\left(x_i\right)\right)^{1-y_i}\right]^{K_h\left(x_i - x_0\right)} \quad (2.14) \end{aligned}$$

Fungsi lokal *likelihood* pada Persamaan (2.14) lebih mudah dimaksimumkan dalam bentuk logaritma natural yang kemudian disebut fungsi lokal *log-likelihood* dan dinyatakan dengan ℓ.

$$\ell(\beta) = \ln\left(\prod_{i=1}^{n}\left[\pi\left(x_i\right)^{y_i}\left(1-\pi\left(x_i\right)\right)^{1-y_i}\right]^{K_h\left(x_i - x_0\right)}\right)$$

$$= \ln\left(\prod_{i=1}^{n} \left[\pi\left(x_i\right)^{y_i} \left(1-\pi\left(x_i\right)\right)^{1-y_i} \right]^{\prod_{j=1}^{k} \frac{1}{h_j} K\left(\frac{x_j - x_{ij}}{h_j}\right)} \right) \qquad (2.15)$$

2.8 Pemilihan Titik Lokal Optimal dan *Bandwidth* Optimal

Bandwidth (h) adalah parameter pemulus (*smooothing*) yang digunakan untuk mengontrol kemulusan dari kurva yang diestimasi sedangkan titik lokal merupakan titik perpaduan bersama yang menunjukkan perubahan perilaku pada data. Model lokal linier terbaik tergantung pada titik lokal yang optimal. Proses pemilihan *bandwidth* dan titik lokal yang sesuai (parameter *smoothing*) merupakan bagian yang penting dari regresi nonparametrik (Hastie & Tibshirani, 1990). Salah satu cara menentukan *bandwidth* dan titik lokal yang optimal dengan menggunakan metode GCV (*Generalized Cross Validation*). Nilai GCV yang memberikan titik lokal optimal adalah nilai GCV yang minimum. Rumus GCV diberikan sebagai berikut: (Suparti dan Prahutama, 2016).

$$GCV(h) = \frac{MSE(h)}{\left(n^{-1} trace\left[\mathbf{I} - \mathbf{A}(h)\right]\right)^2}$$

dengan $MSE(h) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_i\right)^2$ dan $\mathbf{A}(h) = \left(\mathbf{X}^T \mathbf{K_h} \mathbf{X}\right)^{-1}$

dengan:

y_i = nilai variabel respon pada pengamatan ke-*i*

y_i = nilai prediksi variabel respon pada pengamatan ke-*i*

$\mathbf{K_h} = diag\left(K_h\left(x_i - x_0\right), \ldots, K_h\left(x_n - x_0\right)\right)$

$$\mathbf{X} = \begin{bmatrix} 1 & \left(x_{11} - x_{01}\right) & \left(x_{12} - x_{02}\right) & \left(x_{13} - x_{03}\right) \\ 1 & \left(x_{21} - x_{01}\right) & \left(x_{22} - x_{02}\right) & \left(x_{23} - x_{03}\right) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \left(x_{3621} - x_{01}\right) & \left(x_{3622} - x_{02}\right) & \left(x_{3623} - x_{03}\right) \end{bmatrix}$$

2.9 Ketepatan Klasifikasi Model

Ketepatan klasifikasi model digunakan untuk mengetahui suatu data diklasifikasikan dengan benar atau tidak benar. Salah satu cara untuk melakukan evaluasi terhadap model regresi logistik biner adalah dengan menggunakan tabel klasifikasi. Tabel klasifikasi merupakan jumlah data hasil prediksi yang diklasifikasikan benar atau tidak berdasarkan hasil uji yang sebenarnya.

Tabel 2.1 Tabel Klasifikasi

Hasil Prediksi	Hasil Aktual	
	π_0	π_1
π_0	n_{00}	n_{10}
π_1	n_{01}	n_{11}

dengan:

n_{00} = jumlah y_i dari π_0 tepat diklasifikasikan sebagai π_0

n_{01} = jumlah y_i dari π_0 salah diklasifikasikan sebagai π_1

n_{10} = jumlah y_i dari π_1 tepat diklasifikasikan sebagai π_0

n_{11} = jumlah y_i dari π_1 tepat diklasifikasikan sebagai π_1

Berdasarkan Tabel 2.1 dapat diperoleh nilai akurasi. Nilai akurasi menunjukkan tingkat keakuratan model secara keseluruhan, jika semakin tinggi nilai akurasi maka semakin tinggi pula tingkat keakuratan suatu model. Perhitungan nilai akurasi dapat dinyatakan sebagai berikut (Putri dkk., 2021):

$$Akurasi = \frac{n_{00} + n_{11}}{n_{00} + n_{01} + n_{10} + n_{11}} \tag{2.16}$$

2.10 Uji Kesesuaian Model

Uji Hosmer and Lemeshow Goodness of Fit Test adalah alat yang digunakan untuk menguji kesesuaian model dalam regresi logistik. Uji ini dilakukan untuk menguji hipotesis nol untuk mendapatkan bukti bahwa data empiris yang digunakan sesuai dengan model. Hipotesis yang digunakan adalah sebagai berikut:

H_0: Model tidak sesuai (tidak terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model)

H_1: Model sesuai (terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model)

Statistik uji:

$$\chi^2_{HL} = \sum_{s=1}^{g} \frac{(O_s - n'_s\bar{\pi}_s)^2}{n'_s\bar{\pi}_s(1 - \bar{\pi}_s)}$$

dengan:

O_s = jumlah nilai y pada grup ke-s

$\bar{\pi}_s$ = rata-rata dari $\hat{\pi}$ untuk grup ke-s

n'_s = jumlah sampel dalam grup ke-s

s = 1,2, …, 10

Statistik χ^2_{HL} mengikuti sebaran χ^2 dengan derajat bebas $g-2$. Pengambilan keputusan didasarkan pada nilai *Chi-square*, jika χ^2_{HL} lebih besar dari $\chi^2_{(g-2)}$ dan p-$value$, jika p-$value$ kurang dari 0,05 maka tolak H_0 dan jika p-$value$ lebih besar atau sama dengan 0.05 maka terima H_0 (Hosmer dan Lemeshow, 2013).

2.11 *Odds Ratio*

Odds Ratio adalah perbandingan risiko terjadinya suatu dari kejadian suatu kategori yang satu terhadap kategori yang lain. *Odds ratio* berperan dalam memudahkan interpretasi model regresi logistik yang diperoleh. Interpretasi parameter bertujuan untuk mengetahui arti dari nilai taksiran parameter pada variabel prediktor. Interpretasi koefisien dalam model regresi logistik dilakukan dalam bentuk *odds ratio* (Hosmer dan Lemeshow, 2013). *Odds ratio* berhubungan dengan transformasi logit, seperti yang diketahui agar menjadi bentuk yang linear fungsi logistic perlu ditransformasi sedemikian rupa. Transformasi logita dapat dituliskan:

$$\ln\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$$

Sehingga model *odds ratio* dituliskan pada Persamaan (2.17) sebagai berikut:

Persamaan *odds ratio* adalah sebagai berikut:

$$OR = \frac{\pi(x_i)}{1-\pi(x_i)} = \exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}\right) \qquad (2.17)$$