

Covid-19 Patient Mortality Risk Classification Using Bayesian Binary Logistic Regression

Klasifikasi Risiko Kematian Pasien Covid-19 Menggunakan Regresi Logistik Biner Bayesian

Muhammad Qolbi Shobri¹, Ferra Yanuar², Dodi Devianto³

Abstract

At the end of 2019 the world was shocked by a new disease caused by SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2). The disease is called Covid-19 (Coronavirus Disease). The mortality rate due to disease is increasing every day. In Indonesia as of April 2021, confirmed Covid-19 patients who died reached 42,530 patients, seeing the high mortality rate of Covid-19 patients so it needs to be studied further so that the risk of death of these Covid-19 patients can be minimized. This research utilizing binary logistic regression with Bayesian method parameter estimation. In this study, the predictor variables used were in the form of categories that each category in the predictor variables was assumed to have the same risk of death risk of Covid-19 patients. The results of this study indicate that the number of comorbids has a significant effect on the risk of death of Covid-19 patients, the more the number of comorbids suffered by the patient, the higher the risk of death of the patient. The accuracy of this method in classifying data is 84.68%.

Keywords: Binary Logistic Regression, Bayesian Method, Covid-19

Abstrak

Pada akhir tahun 2019 dunia digemparkan dengan penyakit baru yang disebabkan oleh SARS-CoV-2 (*Severe Acute Respiratory Syndrome Coronavirus 2*). Penyakit yang disebabkan oleh SARS-CoV-2 ini disebut dengan Covid-19 (*Coronavirus Disease*). Angka kematian akibat penyakit ini kian hari kian bertambah. Di Indonesia per April 2021 pasien yang terkonfirmasi Covid-19 yang meninggal dunia mencapai 42.530 pasien, melihat tingginya angka kematian pasien Covid-19 sehingga perlu dikaji lebih lanjut agar risiko kematian pasien Covid-19 ini dapat diminimalisir. Penelitian ini menggunakan regresi logistik biner dengan pendugaan parameter metode Bayes. Pada penelitian ini variabel prediktor yang digunakan berbentuk kategori sehingga setiap kategori pada variabel prediktor diasumsikan memiliki risiko yang sama terhadap risiko kematian pasien Covid-19. Hasil penelitian ini menunjukkan bahwa jumlah komorbid berpengaruh signifikan terhadap risiko kematian pasien Covid-19, semakin banyak jumlah komorbid yang diderita oleh pasien maka semakin tinggi pula risiko kematian pasien tersebut. Adapun ketepatan metode ini dalam mengklasifikasi data sebesar 84,68 %.

Kata Kunci: Regresi Logistik Biner, Metode Bayes, Covid-19

1. Pendahuluan

Coronavirus Disease 2019 (Covid-19) merupakan penyakit menular yang disebabkan oleh *Coronavirus* jenis baru yang ditemukan pada tahun 2019. Infeksi dari virus ini dapat menyebabkan penyakit dengan gejala ringan sampai berat [11]. Sebelumnya ada enam jenis *Coronavirus* yang dapat

*Program Studi Magister Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas

Email Address : 17qolbi@gmail.com¹, ferrayanuar@sci.unand.ac.id², ddevianto@sci.unand.ac.id³



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)

menginfeksi manusia yaitu *Alphacoronavirus 229E*, *Alphacoronavirus NL63*, *Betacoronavirus OC43*, *Betacoronavirus HKU1*, *Middle East Respiratory Syndrome Coronavirus* (MERS-CoV) dan *Severe Acute Respiratory Illness Coronavirus* (SARS-CoV). Ada dua jenis *Coronavirus* yang dapat menyebabkan gejala berat yaitu *Middle East Respiratory Syndrome* (MERS) dan *Severe Acute Respiratory Syndrome* (SARS). *International Committee on Taxonomy of Viruses* menetapkan SARS-CoV-2 (*Severe Acute Respiratory Syndrome Coronavirus 2*) sebagai nama *Coronavirus* yang menyebabkan Covid-19, hal ini dikarenakan *Coronavirus* ini memiliki struktur genom yang mirip dengan SARS [9].

Covid-19 dapat menyebabkan infeksi sistem pernafasan, pneumonia dan lebih lanjut menyebabkan kematian. Jumlah kasus Covid-19 kian hari semakin bertambah seiring berjalannya waktu. Menurut KPC PEN (Komite Penanganan Covid-19 dan Pemulihan Ekonomi Nasional) per April 2021, sudah ada 1.566.995 jiwa yang terinfeksi di Indonesia. Dari total tersebut terdapat 1.414.507 kasus yang dinyatakan sembuh dan total kematian mencapai 42.530 jiwa [6]. Melihat masih tingginya angka kematian pasien Covid-19 di Indonesia, hal ini perlu dilakukan kajian lebih lanjut untuk mengetahui faktor apa saja yang dapat menyebabkan risiko kematian pada pasien Covid-19. Variabel respon dalam penelitian ini adalah status pulang pasien, dimana ada dua kemungkinan yaitu pasien dinyatakan sembuh atau meninggal. Variabel ini mengikuti distribusi Bernoulli, sehingga metode yang dapat digunakan adalah Regresi Logistik Biner.

Pendugaan parameter regresi dapat dilakukan dengan beberapa cara yaitu dengan metode klasik dan metode Bayes. Penelitian yang dilakukan oleh Khairiyah [5] menyatakan bahwa pendugaan parameter dengan pendekatan Bayes adalah pendugaan parameter yang lebih baik dibandingkan metode klasik seperti metode Kuadrat Terkecil dan Maksimum *Likelihood*. Pada metode Bayes penarikan kesimpulan tidak hanya berdasarkan informasi dari data sampel (*Likelihood*) akan tetapi ada penambahan informasi subjektif mengenai peluang dari parameter yang tidak diketahui (*Distribusi Prior*). Atas dasar itulah metode yang digunakan pada penelitian ini adalah Regresi Logistik Biner dengan pendugaan parameter menggunakan metode Bayes yang selanjutnya disebut Regresi Logistik Biner Bayesian.

2. Tinjauan Pustaka

2.1 *Coronavirus Diales 2019 (Covid-19)*

Coronavirus Diales 2019 (Covid-19) merupakan penyakit yang disebabkan oleh virus SARS-CoV-2 (*Severe Acute Respiratory Syndrome Coronavirus 2*) dimana virus ini memiliki RNA strain tunggal positif, berkapsul dan tidak bersegmen. Infeksi dari Covid-19 ini dapat berupa gejala ringan, sedang atau berat. Gejala klinis utama yang muncul yaitu demam (suhu lebih dari 38°C), batuk dan kesulitan bernafas. Selain itu dapat disertai dengan sesak memberat, fatigue, mialgia, gejala gastrointestinal seperti diare dan gejala saluran pernafasan lain. Pada kasus yang berat infeksi Covid-19 ini dapat menyebabkan kritis dan kematian [11].

Kematian pasien Covid-19 tidak serta merta hanya disebabkan oleh Covid-19, ada beberapa faktor pendukung lain yang dapat menyebabkan risiko kematian pasien Covid-19 semakin besar. Berdasarkan penelitian yang dilakukan oleh [4] dan [8] ada beberapa hal yang menjadi faktor yang mempengaruhi risiko kematian pasien Covid-19, diantaranya adalah:

1. Jenis Kelamin

Faktor risiko kematian Covid-19 adalah jenis kelamin. Pasien Covid-19 yang berjenis kelamin laki-laki memiliki faktor risiko kematian lebih tinggi dibandingkan pasien yang berjenis kelamin perempuan. Hal ini dikarenakan faktor kromosom dan hormon. Perempuan terproteksi Covid-19 dikarenakan memiliki kromosom X dan hormon seks seperti progesteron yang memiliki peranan penting dalam imunitas tubuh [4].

2. Umur

Umur memiliki pengaruh yang besar terhadap risiko kematian pasien Covid-19. Pasien yang berumur lebih dari 60 tahun memiliki risiko kematian lebih besar dibandingkan dengan pasien yang berumur kurang dari 60 tahun. Hal ini dikarenakan pasien yang berumur lebih dari 60 tahun adanya proses degeneratif anatomi dan fisiologi tubuh sehingga rentan terhadap penyakit. Demikian juga pada

umur tersebut imunitas tubuh semakin menurun sehingga tubuh mudah terserang penyakit, mudah terinfeksi virus, bakteri dan berbagai patogen lainnya [4].

3. Komorbid

Komorbid merupakan penyakit penyerta. Penyakit penyerta yang memiliki risiko tinggi diantaranya hipertensi, kardiovaskuler, diabetes mellitus dan paru obstruktif kronik. Penyakit hipertensi dapat memperparah prognosis Covid-19 dikarenakan adanya konsumsi obat ACE inhibitor dan ARB sebagai intervensi obat hipertensi. Sementara itu, pasien yang menderita penyakit kardiovaskuler biasanya diberi obat ACE 2 dan ARB untuk memproteksi paru-paru. Sedangkan SARS-CoV-2 menggunakan protein ACE 2 untuk memasuki sel, sehingga hal tersebut dapat memperparah keadaan pasien.

Pasien yang memiliki beberapa komorbid biasanya akan memiliki imunitas yang rendah dibandingkan dengan pasien yang tidak memiliki komorbid. Hal inilah yang menjadi salah satu faktor penyebab kematian pada pasien Covid-19. Semakin banyak jumlah komorbid yang dimiliki pasien maka risiko kematian pasien akan semakin tinggi [4].

2.2 Analisis Regresi Logistik

Analisis regresi logistik merupakan salah satu analisis statistika yang digunakan untuk melihat keterkaitan antar variabel. Berdasarkan variabel prediktor yang digunakan, regresi logistik terbagi menjadi dua yaitu regresi logistik sederhana (dengan satu variabel prediktor) dan regresi logistik berganda (dengan dua atau lebih variabel prediktor). Sementara itu, berdasarkan variabel respon yang digunakan regresi logistik terbagi menjadi dua yaitu regresi logistik biner (variabel respon dengan dua kategori) dan regresi logistik multinomial dan ordinal (variabel respon dengan lebih dari dua kategori) [1].

Analisis regresi logistik biner adalah analisis regresi logistik dengan variabel respon yang bersifat biner atau dikotomus. Model regresi logistik biner digunakan jika variabel responnya menghasilkan dua kategori bernilai $Y=1$ (sukses) dan $Y=0$ (gagal), sehingga mengikuti distribusi Bernoulli untuk setiap observasi tunggal [3].

Secara umum model regresi logistik adalah sebagai berikut :

$$p(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)} \quad (2.1)$$

dimana $0 < p(x) < 1$. Hal ini menunjukkan bahwa regresi logistik menggambarkan suatu peluang.

2.3 Analisis Regresi Logistik Biner Bayesian (RLBB)

Analisis Regresi Logistik Biner Bayesian merupakan analisis regresi logistik biner dengan menggunakan metode Bayes.

a. Pendugaan Parameter dengan Metode Bayes

Metode Bayes adalah metode pendugaan parameter dimana parameter diasumsikan sebagai variabel acak yang mempunyai distribusi tertentu. Metode Bayes menggabungkan fungsi *likelihood* dan distribusi prior dari parameter untuk mendapatkan distribusi posterior yang selanjutnya menjadi dasar dalam pendugaan parameter.

Definisi 2.1 [2]

*Fungsi kepadatan peluang bersama dari n peubah acak Y_1, Y_2, \dots, Y_n yang diberi nilai y_1, y_2, \dots, y_n dinotasikan dengan $f(y_1, y_2, \dots, y_n | p)$ yang merupakan fungsi *likelihood*. Untuk nilai y_1, y_2, \dots, y_n tertentu, fungsi *likelihood*-nya merupakan fungsi dari parameter p . Jika Y_1, Y_2, \dots, Y_n merupakan sampel acak yang saling bebas maka;*

$$\begin{aligned} L(\mathbf{p}) &= f(y_1, p)f(y_2, p) \dots f(y_n, p) \\ &= \prod_{i=1}^n f(y_i, p) \end{aligned} \quad (2.2)$$

Distribusi Prior

Distribusi prior merupakan distribusi awal yang memberikan informasi mengenai parameter. Pada dasarnya, distribusi prior adalah representasi subjektif peneliti terhadap suatu nilai parameter yang diduga.

Distribusi prior yang berkaitan dengan bentuk distribusi hasil identifikasi pola datanya dikelompokkan menjadi:

a) Distribusi prior konjugat

Distribusi prior konjugat mengacu pada acuan analisis model terutama dalam pembentukan fungsi *likelihood*-nya sehingga dalam penentuan prior konjugat selalu dipikirkan mengenai pembentukan pola distribusi prior yang mempunyai fungsi kepekatan peluang pembangun *likelihood*-nya.

b) Distribusi prior tidak konjugat

Apabila pemberian distribusi prior pada suatu model tidak mempertimbangkan pola pembentuk fungsi *likelihood* nya.

Berdasarkan penentu parameter pada pola distribusinya, distribusi prior dikelompokkan menjadi:

a) Distribusi prior informatif

Distribusi prior informatif mengacu pada pemberian parameter dari distribusi prior yang telah dipilih, baik distribusi prior konjugat atau tidak konjugat. Pemberian nilai parameter pada distribusi prior ini akan sangat mempengaruhi bentuk distribusi posterior yang akan didapat pada informasi data yang diperoleh.

b) Distribusi prior non-informatif

Pada distribusi prior non-informatif, informasi mengenai parameter tidak tersedia, pemilihannya tidak didasarkan pada data yang ada atau distribusi priornya tidak mengandung informasi tentang parameter[2].

Distribusi Posterior

Distribusi posterior adalah distribusi yang menggabungkan informasi dari data sampel dengan informasi distribusi prior melalui teorema Bayes. Secara umum distribusi posterior dapat dinyatakan sebagai fungsi kepekatan peluang bersyarat dari β jika diketahui nilai observasi \mathbf{Y} , dengan memisalkan $\mathbf{Y} = (y_1, y_2, \dots, y_n)$.

Definisi 2.2 [2]

Fungsi Kepekatan peluang posterior dari β jika diketahui contoh pengamatan $\mathbf{Y} = (y_1, y_2, \dots, y_n)$ adalah

$$P(\beta_j | \mathbf{Y}, \mathbf{X}) = \frac{L(\mathbf{Y}, \mathbf{X}; \beta) \times P(\beta_j)}{\int_{-\infty}^{\infty} L(\mathbf{Y}, \mathbf{X}; \beta) \times P(\beta_j) d\beta_j} \quad (2.3)$$

Mean atau nilai harapan distribusi posterior digunakan untuk menentukan nilai penduga titik bayes bagi parameter β yang tidak diketahui.

Markov Chain Monte Carlo (MCMC)

Markov Chain Monte Carlo (MCMC) adalah suatu metode untuk menentukan nilai parameter dari suatu integrasi analitik yang sulit. Terdapat tiga macam metode MCMC yang dapat digunakan yaitu metode *Metropolis*, metode *Metropolis-Hasting* dan metode *Gibbs Sampler*. Metode *Gibbs Sampler* merupakan teknik yang sering dipakai oleh pengguna metode Bayes, termasuk dalam penelitian inii. Metode *Gibbs Sampler* yang diaplikasikan pada penelitian ini menggunakan *software* WinBUGS [7]. Algoritma *Gibbs Sampler* adalah sebagai berikut:

1. Tentukan nilai awal $\beta^{(0)} = (\beta_0^{(0)}, \beta_1^{(0)}, \dots, \beta_k^{(0)})$
2. Tentukan banyak iterasi (T) untuk membangkitkan sampel masing-masing parameter yang akan diduga. Untuk iterasi $t = 1, 2, \dots, T$ lakukan langkah-langkah berikut:
 - Atur $\beta^t = \beta^{(t-1)}$

Jurnal Matematika, Statistika & Komputasi

Muhammad Qolbi Shobri, Ferra Yanuar, Dodi Devianto

- Untuk $j = 1, 2, \dots, k$, perbarui nilai β_j^t
- Atur $\beta^t = \beta$, yang digunakan untuk membangkitkan iterasi ke- $(t + 1)$
- Proses sampling untuk mendapatkan nilai β_j^t dari distribusi *full conditional posterior* dapat ditulis sebagai berikut.

Bangkitkan $\beta_1^{(t)}$ dari $f(\beta_1 | \beta_2^{(t-1)}, \beta_3^{(t-1)}, \dots, \beta_k^{(t-1)}, \mathbf{Y})$

Bangkitkan $\beta_2^{(t)}$ dari $f(\beta_2 | \beta_1^{(t-1)}, \beta_3^{(t-1)}, \dots, \beta_k^{(t-1)}, \mathbf{Y})$

⋮

Bangkitkan $\beta_k^{(t)}$ dari $f(\beta_k | \beta_1^{(t-1)}, \beta_2^{(t-1)}, \dots, \beta_{k-1}^{(t-1)}, \mathbf{Y})$

3. Setelah iterasi T tersebut periksa apakah data sampel sudah konvergen, bila belum konvergen tambahkan iterasinya sampai konvergen.
4. Lakukan proses burn in yaitu membuang sebanyak B iterasi pertama untuk menghilangkan pengaruh penggunaan nilai awal dalam proses MCMC sampai kondisi stasioner tercapai. Dengan demikian dalam proses MCMC diperoleh sampel sebanyak $L = T - B$ untuk setiap parameter yang diduga.
5. Gunakan $\{\beta^1, \beta^2, \dots, \beta^L\}$ sebagai sampel dalam analisis posterior. Kemudian buat plot dari distribusi posterior setiap parameter yang diperoleh.
6. Tentukan *mean* dan *varians* dari distribusi posterior, yaitu:

Mean Posterior

$$\widehat{G(\beta)} = \frac{1}{L} \sum_{t=1}^L G(\beta^{(t)}) \quad (2.4)$$

dengan $G(\beta)$ adalah fungsi dari parameter β .

Variansi Posterior

$$Var(G(\beta)) = \frac{1}{L-1} \sum_{t=1}^L [G(\beta^{(t)}) - \widehat{G(\beta)}]^2 \quad (2.5)$$

b. Uji Kekonvergenan Parameter

Pada metode Bayes terdapat beberapa cara untuk melihat kekonvergenan parameter model diantaranya *Trace plot*, *Density plot* atau dengan membandingkan nilai *Monte Carlo Error* (MC Error) dengan standar deviasi parameter model. Nilai parameter model dikatakan konvergen jika sebaran plot tersebut stabil menyebar diantara dua garis horizontal yang paralel atau stasioner dan gambar *Density plot* yang dihasilkan menyerupai kurva distribusi normal [7].

c. Uji Signifikansi Parameter

Uji kesignifikanan parameter pada metode bayes ini adalah Uji *Wald*. Uji ini digunakan untuk menguji kelayakan model dengan menguji kesignifikanan tiap-tiap variabel bebas yang digunakan dalam model. Uji ini bertujuan untuk mengetahui apakah variabel prediktor memiliki pengaruh terhadap variabel respon [3].

Hipotesis:

$H_0 : \beta_j = 0, j = 0, 1, 2, \dots, k$ (variabel prediktor ke- j tidak signifikan mempengaruhi variabel respon)

$H_1 : \beta_j \neq 0, j = 0, 1, 2, \dots, k$ (variabel prediktor ke- j signifikan mempengaruhi variabel respon)

Statistik uji:

$$W_j = \left(\frac{\widehat{\beta}_i}{SE(\widehat{\beta}_i)} \right)^2 \quad (2.6)$$

Kesimpulan:

Jika $W_j > \chi^2_{\alpha, 1}$ maka tolak H_0 , artinya variabel prediktor ke- j memiliki pengaruh terhadap variabel respon. Sebaliknya tidak tolak H_0 jika $W_j \leq \chi^2_{\alpha, 1}$, artinya variabel prediktor ke- j tidak memiliki pengaruh terhadap variabel respon.

3. Metodologi Penelitian

3.1 Data Penelitian

Penelitian ini menggunakan data sekunder terkait pasien positif Covid-19 per Desember 2020 yang diperoleh dari Rumah Sakit Universitas Andalas dan Rumah Sakit Dr. M. Djamil Padang. Jumlah data yang digunakan pada penelitian ini sebanyak 457 data.

3.2 Metode Analisis Data

Adapun langkah-langkah analisis data adalah sebagai berikut:

1. Pengumpulan dan pengambilan data dari Rumah Sakit
2. Dalam metode RLBB, variabel prediktor dengan data kategori harus ditransformasi menjadi $(v - 1)$ variabel *dummy* dengan v merupakan banyak kategori pada suatu variabel prediktor. Variabel prediktor X_1, X_2 dan X_3 merupakan data kategori, sehingga variabel ini harus ditransformasi menjadi variabel *dummy*. Adapun variabel *dummy* untuk setiap variabel prediktor dapat dilihat pada Tabel 3.1 berikut.

Tabel 3.1. Variabel Penelitian

Variabel Prediktor	Variabel Dummy	Variabel Pemanding
Jenis Kelamin (X_1)	Perempuan (X_{1D1})	Laki-laki
	≤ 20 Tahun (X_{2D1})	
Umur (X_2)	21 – 40 Tahun (X_{2D2})	> 60 Tahun
	41 – 60 Tahun (X_{2D3})	
	Tidak Memiliki Komorbid (X_{3D1})	
Jumlah Komorbid (X_3)	1 – 3 Komorbid (X_{3D2})	> 3 Komorbid

3. Pendugaan Parameter dengan metode Bayes

- a. Menentukan formula untuk fungsi *likelihood*

Regresi logistik biner mengikuti distribusi bernoulli yang saling bebas, sehingga fungsi *likelihood* dari regresi logistik biner adalah sebagai berikut:

$$L(\mathbf{p}) = \prod_{i=1}^n (p^{y_i} (1-p)^{1-y_i})$$

dengan

$$p(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}$$

maka,

$$\begin{aligned} L(\beta) &= \prod_{i=1}^n (p^{y_i} (1-p)^{1-y_i}) \\ &= \prod_{i=1}^n \left(\left(\frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)} \right)^{y_i} \times \left(1 - \left(\frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)} \right) \right)^{1-y_i} \right) \end{aligned} \quad (3.1)$$

- b. Menetapkan distribusi prior

Prior yang digunakan dalam penelitian ini adalah prior non-informatif yaitu distribusi normal. adapun fungsi kepekatan peluang dari distribusi normal adalah sebagai berikut:

$$P(\beta_j) = \frac{1}{\sqrt{2\pi\sigma_{\beta_j}}} \exp\left(-\frac{1}{2} \left(\frac{\beta_j - \mu_{\beta_j}}{\sigma_{\beta_j}} \right)^2\right) ; j = 1, 2, \dots, k \quad (3.2)$$

- c. Menentukan distribusi posterior

Distribusi Posterior dari regresi logistik biner adalah:

$$P(\beta_j | \mathbf{Y}, \mathbf{X}) = \frac{L(\beta) \times P(\beta_j)}{\int_{-\infty}^{\infty} L(\beta) \times P(\beta_j) d\beta_j}$$

- Distribusi Gabungannya,

$$\begin{aligned} L(\boldsymbol{\beta}) \times P(\beta_j) &= L(\beta) \times P(\beta_0) \times P(\beta_1) \times \dots \times P(\beta_k) \\ &= \left\{ \prod_{i=1}^n \left(\left(\frac{\exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})} \right)^{y_i} \times \left(1 - \left(\frac{\exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})} \right) \right)^{1-y_i} \right) \right\} \\ &\quad \times \left\{ \prod_{j=0}^k \frac{1}{\sqrt{2\pi\sigma_{\beta_j}}} \exp \left(-\frac{1}{2} \left(\frac{\beta_j - \mu_{\beta_j}}{\sigma_{\beta_j}} \right)^2 \right) \right\} \end{aligned}$$

(3.3)

- Distribusi Marginalnya, distribusi posterior marginal untuk setiap parameter didapat dengan cara mengintegalkannya terhadap parameter lain.

$$\begin{aligned} &\int_{-\infty}^{\infty} L(\mathbf{Y}, \mathbf{X}; \beta) \times P(\beta_j) d\beta_j \\ \mathbf{G}(\beta_0 | \beta_1, \beta_2, \dots, \beta_k; \mathbf{Y}, \mathbf{X}) &\propto \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \prod_{i=1}^n \left(\left(\frac{\exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})} \right)^{y_i} \right. \right. \\ &\quad \times \left. \left. \left(1 - \left(\frac{\exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})} \right) \right)^{1-y_i} \right) \right\} \\ &\quad \times \left\{ \frac{1}{\sqrt{2\pi\sigma_{\beta_0}}} \exp \left(-\frac{1}{2} \left(\frac{\beta_0 - \mu_{\beta_0}}{\sigma_{\beta_0}} \right)^2 \right) \right\} d\beta_1 \cdot d\beta_2 \dots d\beta_k \\ \mathbf{G}(\beta_1 | \beta_0, \beta_2, \dots, \beta_k; \mathbf{Y}, \mathbf{X}) &\propto \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \prod_{i=1}^n \left(\left(\frac{\exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})} \right)^{y_i} \right. \right. \\ &\quad \times \left. \left. \left(1 - \left(\frac{\exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})} \right) \right)^{1-y_i} \right) \right\} \\ &\quad \times \left\{ \frac{1}{\sqrt{2\pi\sigma_{\beta_1}}} \exp \left(-\frac{1}{2} \left(\frac{\beta_1 - \mu_{\beta_1}}{\sigma_{\beta_1}} \right)^2 \right) \right\} d\beta_0 \cdot d\beta_2 \dots d\beta_k \\ &\quad \vdots \\ \mathbf{G}(\beta_k | \beta_0, \beta_1, \dots, \beta_{k-1}; \mathbf{Y}, \mathbf{X}) &\propto \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \prod_{i=1}^n \left(\left(\frac{\exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})} \right)^{y_i} \right. \right. \\ &\quad \times \left. \left. \left(1 - \left(\frac{\exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})} \right) \right)^{1-y_i} \right) \right\} \\ &\quad \times \left\{ \frac{1}{\sqrt{2\pi\sigma_{\beta_k}}} \exp \left(-\frac{1}{2} \left(\frac{\beta_k - \mu_{\beta_k}}{\sigma_{\beta_k}} \right)^2 \right) \right\} d\beta_0 \cdot d\beta_1 \dots d\beta_{k-1} \end{aligned}$$

(3.4)

Untuk mendapatkan distribusi posterior dari persamaan di atas tidak mudah karena memerlukan proses integral dengan dimensi yang cukup tinggi, sehingga untuk mengatasi hal tersebut dapat

dilakukan dengan menggunakan *Markov Chain Monte Carlo* (MCMC) dengan algoritma *gibb sampler*, yaitu pendekatan numerik untuk mendapatkan distribusi posterior khususnya jika distribusi posterior dari parameter yang diduga memiliki bentuk yang rumit. Pada penelitian ini dibantu dengan menggunakan *Software WinBugs* dan dilakukan iterasi sebanyak 25.000 kali untuk menghasilkan pendugaan yang lebih teliti, konvergen dalam suatu selang nilai dan mendekati nilai parameter yang sebenarnya.

4. Melakukan Uji kesignifikanan parameter dengan menggunakan uji *wald*.
5. Parameter yang tidak signifikan selanjutnya akan dikeluarkan dalam model. Ulangi langkah awal dengan variabel parameter yang signifikan.
6. Membentuk model logistik biner.
7. Melakukan analisis dan interpretasi dari setiap variabel dengan menggunakan *odds ratio*.
8. Menghitung keakuratan dari model dengan menghitung nilai akurasi dan *press's Q*.
9. Menarik kesimpulan berdasarkan hasil analisis yang diperoleh.

4. Hasil dan Pembahasan

4.1 Pendugaan Parameter Model

Pendugaan Parameter dilakukan menggunakan metode bayes dengan bantuan *software WinBUGS*. Parameter model regresi yang diperoleh selanjutnya akan diuji signifikansinya dengan uji *wald*. Parameter yang signifikan menunjukkan bahwa variabel prediktor yang bersesuaian dengan parameter memiliki pengaruh yang signifikan terhadap variabel respon dan parameter yang tidak signifikan menyatakan variabel prediktor tidak berpengaruh terhadap variabel respon. Adapun nilai pendugaan parameter dapat dilihat pada Tabel 4.1 di bawah ini.

Tabel 4.1. Hasil Pendugaan Parameter model dan Pengujian Kesignifikanan Parameter (dengan $\chi_{0,05;1}^2 = 3,841$)

Variabel Prediktor	Parameter	Mean	Standar Deviasi	Nilai Statistik Wald	Kesimpulan
Konstanta	β_0	1,683	0,5273	10,187	Signifikan
Jenis Kelamin (X_1)					
Jenis Kelamin Perempuan (X_{1D1})	β_{1D1}	-0,172	0,2897	0,352	Tidak Signifikan
Umur (X_2)					
Umur ≤ 20 Tahun (X_{2D1})	β_{2D1}	0,091	0,6330	0,021	Tidak Signifikan
Umur $21 - 40$ Tahun (X_{2D2})	β_{2D2}	-2,312	0,6279	13,557	Signifikan
Umur $41 - 60$ Tahun (X_{2D3})	β_{2D3}	0,120	0,3154	0,145	Tidak Signifikan
Jumlah Komorbid (X_3)					
Tidak Memiliki Komorbid (X_{3D1})	β_{3D1}	-3,739	0,5746	42,343	Signifikan
$1 - 3$ Komorbid (X_{3D2})	β_{3D2}	-2,879	0,5006	33,075	Signifikan

Pada Tabel 4.1. Terlihat nilai wald untuk parameter β_{1D1} , β_{2D1} dan β_{2D3} lebih kecil dari nilai $\chi_{0,05;1}^2$ sehingga dapat disimpulkan bahwa parameter β_{1D1} , β_{2D1} dan β_{2D3} tidak signifikan dalam model. Artinya variabel prediktor yang bersesuaian yaitu jenis kelamin perempuan (X_{1D1}), umur

≤ 20 tahun (X_{2D1}) dan umur 41 – 60 tahun (X_{2D3}) tidak memiliki pengaruh nyata terhadap risiko kematian pasien Covid-19. Sehingga variabel prediktor tersebut dikeluarkan dalam model.

Variabel prediktor yang signifikan memiliki pengaruh terhadap risiko kematian pasien Covid-19 adalah umur 21 – 40 tahun (X_{2D2}), tidak memiliki komorbid (X_{3D1}) dan 1 – 3 komorbid (X_{3D2}). Namun, dikarenakan variabel prediktor umur 21 – 40 tahun (X_{2D2}) merupakan variabel *dummy* dari kategori variabel prediktor umur (X_2) dimana kategori yang lainnya tidak berpengaruh sehingga variabel prediktor umur 21 – 40 tahun (X_{2D2}) juga dikeluarkan dalam model. Setelah diperoleh variabel prediktor yang signifikan selanjutnya akan diduga kembali parameter model regresi dengan metode Bayes menggunakan *Software* WinBUGS. Berikut adalah hasil dugaan parameter model regresinya.

Tabel 4.2. Hasil Pendugaan Parameter model Regresi

Parameter	Mean	Standar Deviasi	MC Error	Batas Bawah (2.5%)	Batas Atas (97.5%)
β_0	1,346	0,4389	0,0121	1,332	2,243
Jumlah Komorbid (X_3)					
β_{3D1}	-3,998	0,5332	0,0130	-3,988	-2,991
β_{3D2}	-2,835	0,4687	0,0127	-2,823	-1,961

Dari Tabel 4. di atas diperoleh model logistik untuk risiko kematian pasien Covid-19 adalah sebagai berikut.

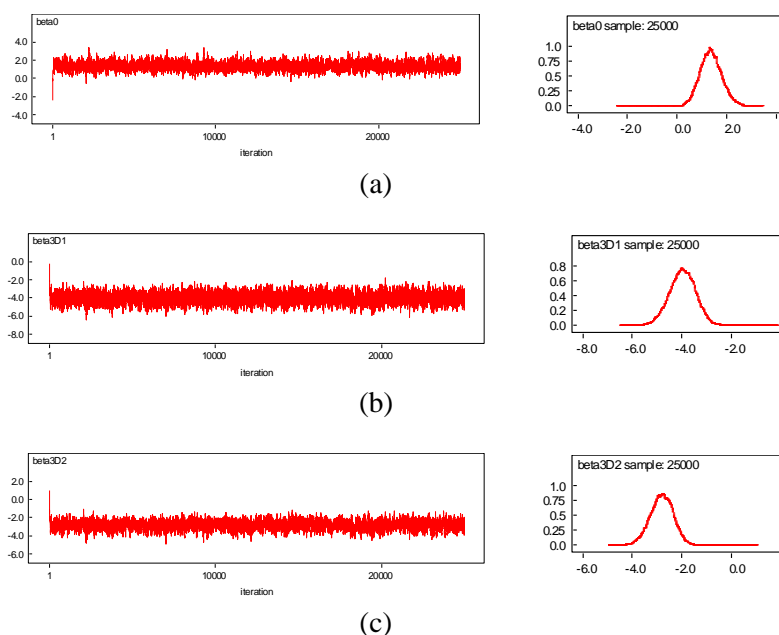
$$\hat{p}(x) = \frac{\exp(1,346 - 3,998 X_{3D1} - 2,835 X_{3D2})}{1 + \exp(1,346 - 3,998 X_{3D1} - 2,835 X_{3D2})} \quad (4.1)$$

Persamaan (4.1) dapat dibentuk menjadi model logit sebagai berikut:

$$\text{logit } \hat{p}(x) = 1,346 - 3,998 X_{3D1} - 2,835 X_{3D2} \quad (4.2)$$

4.2 Uji Kekonvergenan Parameter Model

Setelah dugaan parameter model diperoleh selanjutnya dilakukan uji kekonvergenan dengan mengamati hasil *trace plot* dan *density plot*. Berikut adalah hasil *trace plot* dan *density plot* dengan menggunakan aplikasi *Software* WinBUGS.



Gambar 4.1. Hasil History Trace Plot dan Density Plot untuk Parameter (a). β_0 (b). β_{3D1} (c). β_{3D2}

Berdasarkan gambar diatas, dapat disimpulkan bahwa asumsi kekonvergenan terpenuhi karena sebaran data iterasi sudah stabil dan terletak diantara dua garis horizontal yang paralel serta *density*

plot yang dihasilkan cukup baik karena memiliki pola yang cenderung halus berbentuk kurva distribusi normal.

4.3 Interpretasi Model

Untuk mengetahui seberapa besar variabel prediktor berpengaruh terhadap variabel respon adalah dengan melihat nilai *odds ratio* dari parameter model. Nilai *odds ratio* parameter model dapat dilihat pada tabel berikut ini.

Tabel 4.3. Nilai *Odds Ratio* Parameter Model

Variabel Prediktor	Parameter	Mean ($\hat{\beta}$)	Odd Ratio ($\exp(\hat{\beta})$)
Tidak Memiliki Komorbid (X_{3D1})	β_{3D1}	-3,998	0,018
1 – 3 Komorbid (X_{3D2})	β_{3D2}	-2,835	0,058

Berdasarkan Tabel 4.3 nilai *odds ratio* untuk variabel prediktor *dummy* tidak memiliki komorbid (X_{3D1}) adalah sebesar 0,018 yang berarti pasien Covid-19 yang tidak memiliki komorbid memiliki risiko kematian sebesar 0,018 kali dibandingkan dengan pasien Covid-19 yang memiliki lebih dari tiga komorbid dan mengasumsikan variabel prediktor lainnya konstan. Sementara itu, nilai *odds ratio* untuk variabel prediktor *dummy* 1 – 3 komorbid (X_{3D2}) adalah sebesar 0,058 yang berarti pasien Covid-19 yang memiliki jumlah 1 – 3 komorbid berisiko menyebabkan kematian sebesar 0,058 kali dibandingkan dengan pasien Covid-19 yang memiliki lebih dari tiga komorbid dengan mengasumsikan variabel prediktor lainnya konstan.

Hal ini dapat mengindikasikan bahwa pasien Covid-19 yang memiliki lebih dari tiga komorbid memiliki risiko kematian yang lebih tinggi dibandingkan pasien Covid-19 dengan jumlah komorbid lainnya. Dengan demikian dapat disimpulkan bahwa penyebab kematian pasien Covid-19 di Sumatera Barat dipengaruhi oleh jumlah komorbid.

4.4 Evaluasi Klasifikasi Model

Model regresi logistik dikatakan baik dan stabil dalam klasifikasi jika nilai akurasi mendekati 1 dan nilai *press's Q* lebih besar dari nilai *chi-square* tabel. Nilai akurasi dan *press's Q* diperoleh dengan cara melihat ketepatan model dalam pengklasifikasian atau pengelompokan. Pengelompokan didasarkan pada nilai peluang yang diperoleh dari model regresi logistik dengan menggunakan persamaan (4.1).

Jika nilai peluang $\hat{p}(x)$ lebih besar atau sama dengan 0,5 maka status pulang pasien dikelompokkan ke kategori meninggal, sedangkan jika peluang $\hat{p}(x)$ kecil dari 0,5 maka status pulang pasien dikelompokkan ke kategori sembuh. Terdapat 71 kesalahan dalam pendugaan klasifikasi yaitu 65 dan 6 yang dikelompokkan pada kategori yang salah dari total 457 responden. Sehingga nilai akurasi yang diperoleh dari pengelompokkan tersebut adalah.

$$\begin{aligned}
 \text{Akurasi} &= \frac{\text{banyak data yang diklasifikasi dengan benar}}{\text{total banyak pengamatan}} \\
 &= \frac{387}{457} \\
 &= 84,68 \text{ atau } 84,68 \%
 \end{aligned}$$

$$\begin{aligned}
 \text{Press's } Q &= \frac{[457 - (387(2))]^2}{457(2 - 1)} \\
 &= \frac{100.489}{457} \\
 &= 219,88
 \end{aligned}$$

Berdasarkan nilai akurasi dan *press's Q* di atas, model dapat mengklasifikasi data dengan benar sebesar 84,68 % dan nilai *press's Q* yang lebih besar dari nilai $\chi^2_{0,05;1}$ tabel ($\chi^2_{0,05;1} = 3,841$) yaitu $219,88 > 3,841$. Artinya model sudah akurat dan stabil secara statistik dalam mengklasifikasi risiko kematian pasien Covid-19.

5. Kesimpulan

Berdasarkan pembahasan yang telah dipaparkan, maka dapat disimpulkan bahwa jumlah komorbid memiliki pengaruh yang signifikan terhadap risiko kematian pasien Covid-19. Model regresi logistik yang diperoleh untuk risiko kematian pasien Covid-19 adalah:

$$\hat{p}(x) = \frac{\exp(1,346 - 3,998 X_{3D1} - 2,835 X_{3D2})}{1 + \exp(1,346 - 3,998 X_{3D1} - 2,835 X_{3D2})}$$

dengan ketepatan klasifikasi model tersebut sebesar 84,68%. Dengan demikian dapat dikatakan bahwa model tersebut sudah layak mengklasifikasi risiko kematian pasien Covid-19.

Daftar Pustaka

- [1] Bergerud, W. A., 1996. *Introduction to Logistic Regression Models with Worked Forestry Examples Biometrics Information Handbook No. 7*. Victoria British Columbia: Ministry of Forests.
- [2] Box, G.E.P. dan Tiao, G.C., 1973. *Bayesian Inference in Statistical Analysis*. Philippines: Addison Wesley Company. Inc.
- [3] Harlan, J., 2018. *Analisis Regresi Logistik*. Jakarta : Gunadarma.
- [4] Hindayani, W. R., 2020. Faktor-faktor Risiko yang Berhubungan dengan Covid-19: Literatur Review. *Jurnal Untuk Masyarakat Sehat (JUKMAS)*, Vol.4, No. 2, pp.120-134.
- [5] Khairiyah, R., Maiyastri dan Diana, R., 2016. Perbandingan Metode Kuadrat Terkecil dan Metode Bayes Pada Model Regresi Linier dengan Galat yang Autokorelasi. *Jurnal Matematika UNAND*, Vol.7, No. 1, pp. 115-124.
- [6] NN. 2020. *Data Penyebaran COVID-19*. <https://covid19.go.id/> (diakses pada 14 April 2021).
- [7] Ntzoufras, I., 1973. *Bayesian Modelling Using WinBUGS*. Greece: Departement of Statistics Athens University of Economics and Bussines.
- [8] Siagian, T. H., 2020. Mencari Kelompok Berisiko Tinggi Terinfeksi Virus Corona dengan Discourse Network Analysis. *Jurnal Kebijakan Kesehatan Indonesia (JKKI)*, Vol. 9, No. 2, pp. 98-106.
- [9] Susilo, A., Rumende, C.M., Pitoyo, C.W., Santoso, W.D., Yulianti, M., Herikurniawan, Sinto, R., Singh, G., Nainggolan, L., Nelwan, E.J., Chen, L.K., Widhani, A., Wijaya, E., Wicaksana, B., Maksum, M., Annisa, F., Jasirwan, C.O. dan Yuniastuti, E., 2020. Coronavirus Diases 2019: Tinjauan Literatur Terkini. *Jurnal Penyakit Dalam Indonesia*, Vol. 7, No. 1, pp. 45-67.
- [10] Walpole, R.E., Myers, S.L. dan Ye, K., 2011. *Probability & Statistics for Engineers & Scientists Ninth Edition*. United State of America: Pearson Education, Inc.
- [11] Yuliana, 2020. Coronavirus Diases (Covid-19) : Sebuah Tinjauan Literatur. *Wellness and Healty Magazine*, Vol. 2 No. 1, pp. 187-192.