

Analisis Pengobatan Pasien Menggunakan Model Regresi Logistik Biner

Krishna Prafidya Romantica

Matematika Bisnis, Institut Teknologi dan Bisnis Kalbis Jalan Pulomas Selatan kav.22, Jakarta Timur 13210 Email: krishna.romantica@kalbis.ac.id

Abstract: Researcher used patient data spread across two residential areas, namely sector 1 and sector 2. The research data consisted of four explanatory variables, namely: the age of the patient, the class of patients found in the hospital, the patient's area of residence, and the findings of the disease suffered by the patient. Class, sector, and disease variables are variables categorized into categories 0 and 1. The researcher considers the dummy variables discussed in the explanatory variable variables. Category 0 indicates that the sample does not meet the criteria in the category. Choosing, category I shows that the sample meets the criteria in the category. Next, the researcher will estimate the explanatory parameter variables and dummy variables, then do the partial test to get the parameter significance and model it using the Binary Logistic Regression Model. With the Logistic Regression Model, researcher will calculate the consideration of the patient's recovery. This probability is used as an alternative decisionmaking doctor will improve or stop hospitalization treatment for his patients.

Keywords: binary logistic regression model, dummy variables, probabilities, test the significance of parameters

Abstrak: Peneliti menggunakan data pasien yang tersebar ke dalam dua sektor wilayah tempat tinggal, yaitu sektor 1 dan sektor 2. Data penelitian terdiri atas empat variabel penjelas, yaitu : usia pasien, kelas pasien dirawat di rumah sakit, sektor wilayah pasien tinggal, dan keterangan penyakit yang diderita pasien. Variabel kelas, sektor, dan penyakit adalah variabel yang dikategorikan ke dalam kategori 0 dan 1. Peneliti mempertimbangkan variabel dummy yang dijelaskan pada ketiga variabel penjelas tersebut. Kategori 0 menunjukkan bahwa sampel tidak memenuhi kriteria pada kategori. Sebaliknya, kategori 1 menunjukkan bahwa sampel memenuhi kriteria pada kategori. Selanjutnya, peneliti akan menaksir parameter variabel penjelas dan variabel dummy, serta melakukan uji signifikansi parameter dan memodelkannya menggunakan Model Regresi Logistik Biner. Dengan Model Regresi Logistik, peneliti akan menghitung besarnya probabilitas kesembuhan pasien. Probabilitas ini digunakan sebagai alternatif pengambilan keputusan dokter apakah akan melanjutkan atau memberhentikan pengobatan rawat inap bagi pasiennya.

Kata kunci: model regresi logistik biner, probabilitas, uji signifikansi parameter, variabel dummy

I. PENDAHULUAN

Penyakit adalah suatu keadaan dimana seseorang merasakan disfungsi organ dalam tubuhnya. Keadaan ini apabila tidak diobati akan berpengaruh terhadap keadaan diri di masa mendatang. Penyakit manusia terdiri dari penyakit yang tergolong ringan dan berat. Penyakit ringan adalah suatu keadaan dimana seseorang yang terserang sakit dapat melakukan rawat jalan, sedangkan penyakit berat adalah suatu keadaan dimana seseorang memerlukana perawatan khusus layaknya rawat inap di rumah sakit.

Melihat kondiri ini, peneliti mencoba untuk menaksir probabilitas pasien yang terjangkit penyakit menggunakan data pasien yang tersebar ke dalam dua sektor wilayah tempat tinggal. Data ini akan dianalisis menggunakan Model Regresi Logistik Biner karena variabel bebas merupakan campuran antara variabel kontinyu (metric) dan kategorial (non-metrik) [1].

Pertama, peneliti mempertimbangkan variabel dummy yang dijelaskan pada variabel penjelas. Peneliti menggunakan empat variabel penjelas, yang terdiri atas: usia pasien, kelas pasien dirawat di rumah sakit, sektor wilayah pasien tinggal, dan keterangan penyakit yang diderita pasien. Variabel *dummy* akan dikategorikan ke dalam 0 dan 1. Kategori 0 menunjukkan bahwa sampel tidak memenuhi kriteria pada kategori. Berbanding terbalik dengan kategori 1 yang menunjukkan bahwa sampel memenuhi kriteria pada kategori.

Kedua, peneliti menaksir parameter variabel penjelas dan variabel *dummy*, serta melakukan uji signifikansi parameter guna menentukan variabel penjelas dan variabel *dummy* mana saja yang berpengaruh terhadap variabel respon. Setelah memenuhi kriteria, peneliti akan memodelkan data menggunakan Model Regresi Logistik Biner.

Ketiga, peneliti akan menaksir probabilitas kesembuhan pasien menggunakan Model Regresi Logistik Biner. Probabilitas ini digunakan sebagai alternatif pengambilan keputusan dokter untuk melanjutkan atau memberhentikan pengobatan rawat inap bagi pasiennya.

II. METODE PENELITIAN

Secara harfiah, kata dummy memiliki arti buatan. Variabel dummy adalah variabel yang nilai sebenarnya adalah buatan karena nilai variabel tersebut sebenarnya bukanlah skala. Jika variabel penjelas berukuran kategori atau dikotomi, maka variabel dummy dinyatakan dengan kode 0 atau 1. Kategori 0 menunjukkan ketidakhadiran ciri dan 1 menunjukkan sampel memiliki ciri tertentu. Setiap variabel dummy menyatakan satu kategori variabel dan setiap variabel dengan k kategori dapat dinyatakan dengan k-1 variabel dummy [2].

Pada Analisis Regresi Logistik, peneliti akan menentukan pengaruh sejumlah variabel penjelas $X_1, X_2, ..., X_k$. terhadap variabel respon Y yang berupa variabel kategori (binomial, multinomial, atau ordinal) atau memprediksi nilai suatu variabel respon Y (yang berupa variabel kategorik) berdasarkan nilai variabel-variabel penjelas $X_1, X_2, ..., X_k$.

Asumsi-asumsi yang digunakan dalam regresi logistik, antara lain: Tidak mengasumsikan hubungan linier antara variabel respon dan variabel penjelas; Variabel respon harus bersifat dikotomi (dua variabel); Variabel respon tidak harus memiliki keragaman yang sama antar kelompok variabel; Kategori dalam variabel penjelas harus terpisah satu sama lain atau bersifat eksklusif; dan Sampel yang diperlukan dalam jumlah relative besar, minimum yang dibutuhkan hingga 50 sampel data untuk sebuah variabel penjelas [3].

A. Regresi Logistik Biner

Regresi logistik biner merupakan regresi logistik dimana variabel responnya berupa *variabel dikotomi* atau *variabel biner*. Bila kedua nilai *variabel biner* ini dilambangkan dengan 1 = sukses dan 0 = gagal, maka distribusi Bernoulli untuk variabel biner ini adalah $P(Y=1) = \pi$ dan $P(Y=0) = 1 - \pi$ dengan nilai harapan $E(Y) = 1(\pi) + 0(1-\pi) = \pi$ [4].

Bentuk umum Regresi Logistik Biner dengan link function logit adalah

$$\log it(\pi_{j}) = \ln - - - = \beta_{0} + \beta_{1}x_{j1} + \beta_{2}x_{j2} + ... + \beta_{k}x_{jk}$$

$$\frac{\pi_{j}}{1 - \pi_{j}} = \exp(\beta_{0} + \beta_{1}x_{j1} + \beta_{2}x_{j2} + ... + \beta_{k}x_{jk}) \text{ atau}$$

$$\pi_{j} = \frac{\exp(\beta_{0} + \beta_{1}x_{j1} + \beta_{2}x_{j2} + \dots + \beta_{k}x_{jk})}{1 + \exp(\beta_{0} + \beta_{1}x_{j1} + \beta_{2}x_{j2} + \dots + \beta_{k}x_{jk})}$$

dimana:

 β_0 = konstanta, β_i = koefisien, dan x_{ji} = predictor ke-i.

 π_j = probabilitas bahwa faktor atau *covariate* ke- j mempunyai response = 1 (sukses) dari response regresi logistik biner yang mempunyai nilai 0 (gagal) dan 1 (sukses) [5].

1. Penaksiran Parameter Model Regresi Logistik

Tujuan dari regresi Logistik adalah menaksir parameter-parameter, β_k , yang berkontribusi pada link function logit. Andaikan terdapat k variabel penjelas $X_1, X_2, ..., X_k$, fungsi densitas peluang bersyarat Y terhadap β mengikuti Distribusi Bernoulli adalah

$$f(y | \beta) = \prod_{i=1}^{N} \pi_i^{y_i} (1 - \pi_i)^{1 - y_i} ; y_i = 0,1$$

Y diberi kode dengan 0 dan 1 untuk setiap pasangan (x_i, y_i) . Jika $y_i = 1$, maka konstribusi untuk fungsi Likelihood adalah $\pi(x_i)$ dan jika $y_i = 0$, maka konstribusi untuk fungsi Likelihood adalah $1 - \pi(x_i)$, dimana $\pi(x_i)$ menotasikan nilai dari $\pi(x)$ pada x_i sehingga kontribusi untuk fungsi Likelihood dari pasangan (x_i, y_i) adalah

$$L(\beta) = \prod_{i=1}^{N} \pi_i^{y_i} (1 - \pi_i)^{1 - y_i} ; \quad y_i = 0, 1$$

Parameter pada model Regresi Logistik Biner dapat ditaksir menggunakan metode $Maximum\ Likelihood\ Estimation\ (MLE)$. Taksiran maksimum $Likelihood\ adalah\ nilai-nilai\ eta\ yang$ dapat memaksimalkan fungsi $Likelihood\ Agar$ mempermudah perhitungan, perlu dilakukan

pendekatan terhadap fungsi *log likelihood* yang didefinisikan oleh:

$$l(\beta) = \sum_{i=1}^{N} \left\{ y_{i} \sum_{k=0}^{K} x_{ik} \beta_{k} - \ln \left(1 + e^{\sum_{k=0}^{K} x_{ik} \beta_{k}} \right) \right\}.$$

Titik kritis fungsi log *Likelihood* diperoleh jika nilai tutunan parsial pertama terhadap masing-masing β adalah sama dengan nol [5].

2. Uji Signifikansi Parameter Model Regresi Logistik

Uji signifikansi parameter Model Regresi Logistik yang digunakan dalam penelitian ini antara lain : uji *Likelihood Ratio*, uji Wald, uji Hosmer dan Lameshow, dan $R-kuadrat(R^2)$.

Uji Likelihood Ratio

Prinsip utama dari uji Likelihood Ratio adalah membandingkan nilai observasi dari variabel respon dengan nilai prediksi yang diperoleh dari model dengan variabel penjelas dan tanpa variabel penjelas yang ditanyakan [5]. Pada regresi Logistik perbandingan dari nilai observasi dan nilai prediksi didasarkan pada fungsi Likelihood, dinyatakan sebagai statistik D, yaitu:

$$D = -2\ln\left[\frac{likelihood\ dari\ model\ prediksi}{likelihood\ dari\ model\ observasi}\right]$$
$$= -2\sum_{i=1}^{n}\left[y_{i}\ln\left(\frac{\hat{\pi}_{i}}{y_{i}}\right) + \left(1 - y_{i}\right)\ln\left(\frac{1 - \hat{\pi}_{i}}{1 - y_{i}}\right)\right]$$

dimana $\hat{\pi}_i = \hat{\pi}(x_i)$.

Dengan maksud untuk menaksir variabel bebas yang signifikan, bandingkan nilai D dari model dengan variabel bebas dan nilai D dari model tanpa variabel bebas [5]. Perubahan D yang menyertakan variabel bebas dalam model didefinisikan sebagai:

$$G = D_1 - D_2$$

$$G = 2 \left[\sum_{i=1}^{n} y_{i} \ln \hat{\pi}_{i} + \sum_{i=1}^{n} (1 - y_{i}) \ln (1 - \hat{\pi}_{i}) - n_{1} \ln n_{1} - n_{0} \ln n_{0} + n \ln n \right]$$

dimana : $D_1 = D$ untuk model tanpa variabel $D_2 = D \text{ untuk model dengan variabel}$

Uji Wald

Uji Wald digunakan untuk melakukan uji signifikasi parameter β_i secara individu [5]. Uji Wald menggunakan pendekatan statistik Z yang mengikuti distribusi Normal Baku. Statistik Z dinyatakan oleh:

$$Z = \frac{\hat{\beta}_k}{SE\left(\hat{\beta}_k\right)}$$

dimana:

 $\hat{\beta}_k$ = penduga untuk parameter (β_1) $SE(\hat{\beta}_k)$ penduga *error* baku untuk koefisien (β_k)

Uji Hosmer dan Lameshow

Uji Hosmer and Lemeshow dikenal sebagai uji kecocokan Model Regresi Logistik terhadap data [5]. Persamaan yang digunakan adalah

$$\begin{split} \hat{C} &= \sum_{k=1}^g \frac{\left(o_k - n_k \overline{\pi}_k\right)^2}{n_k \overline{\pi}_k \left(1 - \overline{\pi}_k\right)} \\ \text{dimana:} \\ o_k &= \sum_{j=1}^{n_k} y_j \quad \text{dan} \quad \overline{\pi}_k = \sum_{j=1}^{n_k} \frac{m_j \hat{\pi}_j}{n_k}. \end{split}$$

Uji Hosmer dan Lemeshow menggunakan pendekatan distribusi Chi Kuadrat dengan derajat kebebasan (g-2).

$$R - kuadrat(R^2)$$

Nilai R^2 pada Analisis Regresi Logistik menunjukkan kuatnya hubungan antara variabel penjelas dengan variabel respon [5]. Nilai dinyatakan oleh:

$$R^2 = 1 - \exp\left[-\left(\frac{L^2}{n}\right)\right]$$

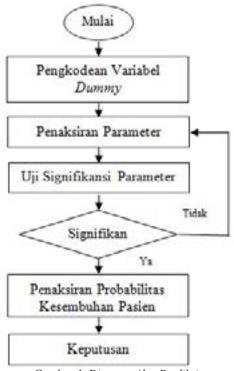
dimana

L = nilai log Likelihood dari model n = banyaknya data

B. Metode Penelitian

Data yang digunakan dalam penelitian ini adalah kumpulan data penyakit yang terdiri atas 196 pasien yang dipilih dalam sampel probabilitas dua sektor. Data ini diperoleh melalui website http://calcnet.mth. cmich.edu/org/spss, dimana terdapat 89 pasien yang dinyatakan memerlukan pengobatan rawat jalan dan 107 sisanya memerlukan pengobatan rawat inap.

Peneliti menggunakan empat variabel penjelas, yang terdiri atas: usia pasien, kelas pasien dirawat di rumah sakit, sektor wilayah pasien tinggal, dan keterangan penyakit yang diderita pasien. Dan terdapat tiga variabel penjelas yang dikategorikan, yaitu: Kelas pasien dirawat di rumah sakit terbagi menjadi kelas 1, kelas 2, dan kelas 3. Ini berarti ada dua variabel *dummy* yang digunakan untuk menyatakan hal tersebut; dan Sektor wilayah pasien tinggal terbagi menjadi dua, yaitu sektor 1 dan sektor 2. Serta, pasien yang mengidap penyakit ringan dikategorikan 0 dan pasien yang mengidap penyakit berat dikategorikan 1. Pada data sektor wilayah dan penyakit yang diderita pasien, peneliti menggunakan satu variabel *dummy* [6].



Gambar 1 Diagram Alur Prediksi Probabilitas Kesembuhan Pasien

Alur penelitian yang digunakan oleh peneliti seperti pada Gambar 1. Langkah pertama, peneliti akan melakukan pengkodean terhadap variabel dummy (kelas, sektor, dan penyakit). Kedua, peneliti menaksir parameter β dengan memaksimumkan fungsi Likelihood. Ketiga, peneliti melakukan uji signifikansi parameter-parameter dari variabel tersebut menggunakan uji Likelihood Ratio, Uji Wald, Uji Hosmer dan Lameshow, dan $R-kuadrat(R^2)$. Pada tahap ini, peneliti akan memprediksi variabel penjelas mana sajakah yang mempengaruhi variabel respon Y. Jika terdapat variabel bebas yang tidak signifikan terhadap variabel respon Y, peneliti melakukan taksiran kembali terhadap variabel-variabel bebas yang signifikan saja. Setelah diperoleh variabel-variabel bebas yang signifikan mempengaruhi variabel respon Y, peneliti akan memodelkan data menggunakan Model Regresi logistik Biner agar dapat memprediksi probabilitas kesembuhan para pasien yang terjangkit penyakit. Probabilitas ini dapat digunakan sebagai acuan pihak rumah sakit dalam mengambil keputusan apakah para pasiennya memerlukan tidakan lanjutan atau tidak. Keputusan ini juga memerlukan beberapa pertimbangan medis lainnya yang disesuaikan oleh kebutuhan pasien dalam masa penyembuhan.

III. HASIL DAN PEMBAHASAN

Pada tahap I, diperoleh hasil koefisien $-2 \cdot Log \ Likelihood = 215,357$ dan disubstitusikan ke

dalam persamaan uji *Likelihood Ratio* : G = 700,7723. Dengan menggunakan $\alpha = 0,05$, nilai statistik uji chi-square > $_{(1\ 0.05)(5)} = 1,145$, dimana $G > \chi^2_{(1-0.05)(5)}$. Berdasarkan kriteria hipotesis :

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$$

(variabel X_1, X_2, X_3, X_4, X_5 secara bersama tidak mempengaruhi $\pi(X)$) peneliti menolak H_0 artinya variabel (dummy) X_1, X_2, X_3, X_4, X_5 secara bersama mempengaruhi $\pi(X)$ dengan tingkat kepercayaan 95%.

Uji *Wald* digunakan untuk melihat variabel (dummy) yang signifikan terhadap model regresi logistik biner tersebut. Rumusan hipotesis yang digunakan adalah

 $H_0: \beta_k = 0$ (variabel X_k tidak mempunyai pengaruh terhadap $\pi(X)$)

Peneliti menolak H_0 apabila nilai $p_{value} < H_0$. Berdasarkan kriteria penolakan, terdapat empat variabel *(dummy)* yang signifikansi terhadap model regresi logistik biner, Seperti pada Tabel 1.

Tabel 1 Hasil perhitungan Ratio ${f Z}$ pada parameter variabel

Koefisien Parameter Variabel (Dummy)	Taksiran parameter B	p_{value}	Signifikansi
Age	0.036	0.000	Signifikansi
Kelasl	1.955	0.000	Signifikansi
Kelas2	0.823	0.050	Signifikansi
Sektorl	-0.775	0.032	Signifikansi
Penyakit0	0.041	0.918	Tidak Signifikansı
Constant	-1.157	0.027	Signifikansi

Pada tahap II, peneliti melakukan uji hipotesis kembali untuk menaksir parameter variabel (dummy). Hasil koefisien $-2 \cdot Log \ Likelihood = 218,904$ dan disubstitusikan ke dalam persamaan uji $Likelihood \ Ratio$ diperoleh G=707,8663. Dengan menggunakan $\alpha=0,05$, nilai statistik uji chi-square $G>\chi^2_{(1-0.05)(4)}=0,711$, dimana $G>\chi^2_{(1-0.05)(4)}$. Peneliti menolak , artinya variabel (dummy) X_1,X_2,X_3,X_4 secara bersama mempengaruhi $\pi(X)$ dengan tingkat kepercayaan 95%.

Pada uji Wald, kriteria penolakan H_0 adalah $p_{value} < H_0$. Berdasarkan kriteria penolakan, peneliti menyimpulkan bahwa keempat variabel (dummy) ini tetap signifikansi terhadap model regresi logistik biner, Seperti pada Tabel 2.

Uji *Hosmer and Lemeshow*. Dengan $\alpha = 0.05$, diperoleh $p_{value} = 0.624$, dimana $p_{value} > \alpha$. Peneliti tidak menolak H_0 , artinya bahwa tidak ada perbedaan antara nilai observasi dengan nilai prediksi

Tabel 2 Hasil perhitungan Ratio $\overline{Z}\,$ pada parameter variabel

Koefisien Parameter Variabel (Dummy)	Taksiran parameter B	P_{value}	Signifikansi
Age	0.036	0.000	Signifikansi
Kelas1	1.957	0.000	Signifikansi
Kelas2	0.825	0.049	Signifikansi
Sektor1	-0.765	0.029	Signifikansi
Constant	-1.130	0.012	Signifikansi

model. Uji R^2 . Pada analisis regresi logistik, nilai R^2 menunjukkan hubungan antara variabel penjelas dengan variabel respon. Pada penelitian ini, nilai $R^2 = 1$, artinya variabel penjelas memiliki hubungan sempurna dengan variabel respon.

Sesuai dengan hasil estimasi di atas, maka fungsi logit yang digunakan pada Model Regresi Logistik Biner adalah:

$$\log it(\pi_j) = -1,130 + 0,036x_{j1} + 1,957x_{j2} + 0,825x_{j3} - 0,765x_{j4}$$
 dimana:

 x_{i1} : variabel usia

 x_{i2} : variabel dummy kelas 1

 x_{i3} : variabel dummy kelas2

 x_{i4} : variabel dummy sektor1

i: data sampel ke-1, 2, ..., 196

Secara statistika, peneliti akan mengestimasi probabilitas sampel yang berasal dari dua sektor tersebut menggunakan fungsi logit. Hal ini dilakukan untuk pengambilan keputusan apakah dokter akan melanjutkan atau memberhentikan pengobatan rawat inap kepada pasiennya.

Ilustrasi Kasus: Pasien 17 berusia 67 tahun, Ia dirawat di kelas 1 dan ia tinggal pada wilayah sektor 2. Pasien ini didiagnosa mengidap penyakit berat.

Analisis Kasus: Berdasarkan informasi-informasi tersebut, estimasi probabilitas Pasien 17 adalah sebesar $\pi_{17}=0.891582$. Ini artinya bahwa peluang Pasien 17 sembuh adalah sebesar 89,1582%. Mengingat peluang yang diprediksi cukup besar nilainya, penulis menyarankan kepada dokter untuk menghentikan pengobatan rawat inap kepada Pasien 17. Dengan demikian, Pasien 17 dapat melakukan pengobatan rawat jalan saja.

IV. SIMPULAN

Berdasarkan analisis yang dilakukan peneliti, terdapat beberapa keseimpulan, yaitu: Variabel dummy adalah variabel penjelas yang wujudnya berskala non-metric atau kategori. Setiap variabel dengan k kategori dapat dinyatakan dengan k-1 variabel dummy. Pada penelitian ini, peneliti menggunakan satu variabel penjelas (variabel usia) dan empat variabel dummy (Kelas1, Kelas2, Sektor1, dan Penyakit0); Sesuai dengan prosedur analisis Regresi Logistik Biner, peneliti menganalisis terdapat satu variabel penjelas dan tiga variabel dummy (Kelas1, Kelas2, dan Sektor1) yang mempengaruhi variabel respon keputusan. Model Regresi Logistik yang digunakan menaksir probabilitas kesembuhan tersebut dinyatakan oleh:

$$\pi_j = \frac{\exp\left(-1,130+0,036x_{j1}+1,957x_{j2}+0,825x_{j3}-0,765x_{j4}\right)}{1+\exp\left(-1,130+0,036x_{j1}+1,957x_{j2}+0,825x_{j3}-0,765x_{j4}\right)} ; dan$$
 Secara statistika, model Regresi Logistik Biner dapat digunakan sebagai alternatif dalam pengambilan keputusan. Ketentuan yang diambil disesuaikan seperti Tabel 3.

Tabel 3 Pengambilan keputusan.

No	Probabilitas	Keputusan
1	$\pi_j \le 0.35$	Melanjutkan pengobatan rawat inap
2	$0,35 < \pi_j \le 0,65$	Melanjutkan atau menghentikan pengobatan rawat inap
3	$\pi_j \le 1,00$	Menghentikan pengobatan rawat inap

V. DAFTAR RUJUKAN

- [1] H. Akhtar, 1 April 2018. [Online]. [Accessed 20 Juni 2019].
- [2] H. Huang, 19 Desember 2018. [Online]. [Accessed 10 Juni 2019].
- [3] I. Ghozali, Aplikasi Analisis Multivariate dengan Program SPSS (Cetakan IV), Semarang: Badan Penerbit Universitas Diponegoro, 2006.
- [4] Sukono, A. Sholahuddin, M. Mamat and K. Prafidya, "Credit Scoring for Corporative of Financial Services Using Logistic Regression Estimated by Logistic Algorithm," Applied Mathematical Sciences, pp. 45 -57, 2014.
- [5] S. S. Uyanto, Pedoman Analisis Data dengan SPSS, Yogyakarta: Graha Ilmu, 2009.
- [6] W. Widhiarso, 24 November 2010. [Online]. [Accessed 27 Juni 2019].