Penentuan Model Regresi Logistik Terbaik yang Diterapkan pada Data Nasabah Debitur di Bank A Melalui Modifikasi Delapan Kriteria Pemilihan

[1]Teguh Fayoga, [2]Abdul Kudus, [3]Siti Sunendiari
[1,2,3]*Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Jl. Tamansari No. 1 Bandung 40116*
e-mail: [1]T_prayoga@ymail.com, [2]Akudus69@yahoo.com, [3]sunen_diari@yahoo.com

Abstrak. Kredit macet merupakan salah satu faktor penyebab terjadinya kebangkrutan pada industri perbankan. Dalam dunia perbankan, diperlukan analisis yang mampu mengurangi terjadinya resiko kredit. Pada penelitian ini metode yang digunakan untuk menghindari resiko kredit macet yang ditimbulkan oleh nasabah maka diperlukan analisis klasifikasi kredit. Penelitian untuk klasifikasi kredit kali ini akan menggunakan Model Regresi Logistik. Model Regresi Logistik dikenal sebagai model pilihan kualitatif, karena model regresi ini tidak dapat ditaksir dengan mengunakan metode kuadrat terkecil biasa. Jika terdapat banyak kandidat model, maka perlu dipilih model terbaik. Dalam makalah ini akan diterapkan pemilihan model terbaik dengan delapan kriteria pemilihan yang dimodifikasi (*Modified Eight Selection Criteria*, M8SC). Makalah ini akan menggambarkan pendekatan pembentukan model. Dimana seluruh struktur pembentukan model mulai dari mengidentifikasi variabel dependen dan variabel independennya dalam rangka mendapatkan model terbaik akan digambarkan menggunakan ilustrasi numerik.

Kata kunci : Kredit Macet, Analisis Klasifikasi Kredit, Regresi Logistik, M8SC.

A. Pendahuluan

Dalam menghadapi masalah resiko kredit macet yang dialami oleh industri perbankan saat ini salah satunya dapat diatasi dengan mengidentifikasi dan memprediksi nasabah dengan baik sebelum memberikan pinjaman dengan cara memperhatikan data historis pinjaman. Oleh karena itu klasifikasi resiko kredit dalam perbankan memiliki peran yang penting.

Pada penelitian ini untuk menghindari resiko kredit macet yang ditimbulkan oleh nasabah diperlukan analisis klasifikasi kredit guna mengurangi resiko kebangkrutan bank. Untuk klasifikasi kredit kali ini akan menggunakan Model Regresi Logistik. Model Regresi Logistik dikenal sebagai model pilihan kualitatif, oleh karena itu model regresi logistik tidak dapat ditaksir dengan mengunakan metode kuadrat terkecil biasa. Jika terdapat banyak kandidat model, maka perlu dipilih model terbaik. Dalam skripsi ini akan diterapkan pemilihan model terbaik dengan delapan kriteria pemilihan yang dimodifikasi yang disebut dengan singkatan M8SC.

Berdasarkan identifikasi masalah maka tujuan yang ingin dicapai dalam penulisan skripsi ini adalah: Untuk menentukan model regresi logit biner multipel terbaik yang akan diterapkan pada data nasabah debitur di bank A dengan menggunakan M8SC.

B. Landasan Teori

1. Kredit Macet

Kredit yang diberikan oleh bank dapat didefinisikan sebagai penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam-meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam untuk melunasi hutangnya setelah jangka waktu tertentu dengan jumlah bunga, imbalan, atau pembagian hasil keuntungan

(Taswan, 2003).

Pada kenyataanya di dalam praktik selalu ada sebagian nasabah yang tidak dapat mengembalikan kredit kepada bank yang telah meminjaminya. Akibat nasabah tidak dapat membayar lunas utangnya, maka akan tergambar perjalanan kredit menjadi macet atau terhenti.

2. Regresi Logistik

Regresi logistik merupakan suatu metode analisis data yang digunakan untuk mencari hubungan antara variable respon (Y) yang bersifat biner atau dikotomus dengan variabel prediktor (X) (Hosmer dan Lemeshow, 2000). Realisasi dari variabel respon Y terdiri dari 2 kategori yaitu sukses dan gagal yang dinotasikan dengan y = 1 (sukses) dan y = 0 (gagal). Dalam keadaan demikian, variabel Y mengikuti distribusi Bernouli untuk setiap observasi tunggal. Fungsi probabilitas untuk setiap observasi adalah diberikan sebagai berikut.

$$f(y) = \pi^{y}\,(1-\pi)^{1-y}\,; \qquad y = 0,1 \tag{2.1}$$

dimana jika y = 0 maka f(y) = 1- π dan jika y = 1 maka f(y) = π. Fungsi logistiknya dapat dituliskan sebagai berikut:

$$f(x) = \frac{1}{1+e^{-\beta_0+\beta_1 x_1+\cdots+\beta_p x_p}} \quad \text{atau} \quad f(x) = \frac{e^{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p}}{1+e^{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p}} \tag{2.2}$$

dengan p = banyak variabel prediktor. Nilai f(x) terletak antara 0 dan 1. Hal tersebut menunjukkan bahwa model logistik sebenarnya mengambarkan probabilitas atau risiko dari suatu objek. Model regresi logistik adalah sebagai berikut:

$$\pi = \frac{e^{(\beta_0+\beta_1 x_1+\cdots+\beta_p x_p)}}{1+e^{(\beta_0+\beta_1 x_1+\cdots+\beta_p x_p)}}, \text{ dimana p = banyaknya variabel prediktor.} \tag{2.3}$$

Untuk mempermudah penaksiran parameter regresi maka model regresi logistik pada persamaan (2.3) dapat diuraikan dengan menggunakan transformasi logit dari π, sehingga diperoleh persamaan berikut:

$$g(x) = \ln\frac{\pi}{1-\pi} = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p \tag{2.4}$$

Model tersebut merupakan fungsi linier dari parameter-parameternya. Pada regresi logistik, variabel respon diekspresikan sebagai $y = \pi + \varepsilon$ dimana ε mempunyai salah satu dari kemungkinan dua nilai yaitu $\varepsilon = 1 - \pi$ dengan peluang π jika y = 1 dan $\varepsilon = -\pi$ dengan peluang $1 - \pi$ jika y = 0 dan mengikuti distribusi bernouli dengan rataan nol dan varians $\pi(1-\pi)$.

2.1 Penaksiran Parameter

Dalam regresi logistik penaksiran parameter model dilakukan dengan metode *maksimum likelihood*. Metode tersebut menaksir parameter β dengan cara memaksimumkan fungsi likelihood dan mensyaratkan bahwa data harus mengikuti suatu distribusi tertentu. Pada regresi logistik, setiap pengamatan mengikuti distribusi Bernouli sehingga dapat ditentukan fungsi likelihoodnya.

Jika X_i dan Y_i adalah pasangan variabel prediktor dan terikat pada pengamatan ke-i dan diasumsikan bahwa setiap pasangan pengamatan saling bebas dengan pasangan pengamatan lainnya, i = 1, 2, . . . , n, maka fungsi probabilitas untuk setiap pasangan adalah sebagai berikut:

$$f(y_i) = \pi_i^{\,y_i}\; 1 - \pi_i^{\;1-y_i} \qquad ; y_i = 0,1 \tag{2.5}$$

dengan
$$\pi_i = \frac{e^{\beta_0+\sum_{j=1}^{p}\beta_j x_j}}{1+e^{\beta_0+\sum_{j=1}^{p}\beta_j x_j}} \tag{2.6}$$

Setiap pasangan pengamatan diasumsikan saling bebas sehingga fungsi likelihoodnya merupakan gabungan dari fungsi distribusi masing-masing pasangan yaitu sebagai berikut:

$$L(\beta) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \pi_i^{yi} (1 - \pi_i)^{1-yi}$$

Nilai taksiran $\underline{\beta}$ didapatkan melalui turunan $l(\beta)$ terhadap $\underline{\beta}$ dan hasilnya sama dengan nol.

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

Sehingga, $\sum_{i=1}^{n} y_i x_{ij} - \sum_{i=1}^{n} x_{ij} \hat{\pi}_i = 0$ dengan j = 0, 1, . . . , p (2.7)

Penaksiran varians dan kovarians dikembangkan oleh (Rao, 1973 dalam Hosmer dan Lemeshow, 2000). Melalui turunan kedua L(β).

$$\frac{\partial^2 l(\beta)}{\partial \beta j \, \beta u} = \prod_{i=1}^{n} x_{ij} \, x_{iu} (1 - \pi_i) \qquad \text{dengan } j = 0, 1, \ldots, p.$$

2.2 Pengujian Parameter

Uji yang dapat digunakan untuk menguji signifikansi koefisien β dari model dapat menggunakan uji secara serentak maupun parsial.

2.2.1 Uji Serentak

Pengujian secara serentak dilakukan untuk memeriksa kemaknaan koefisien β secara keseluruhan dengan hipotesis sebagai berikut.

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

H_1 : paling tidak terdapat satu $\beta_j \neq 0$; j = 1, 2, . . . , p

Statistik uji : $G = -2\ln \dfrac{\left(\frac{n1}{n}\right)^{n1} \left(\frac{n0}{n}\right)^{n0}}{\prod_{i=1}^{n} \pi_i^{y_i} (\widehat{1 - \pi_i})^{1-y_i}}$ (2.8)

dimana n0 menyatakan banyaknya pengamatan yang nilai responnya, y = 0 dan n1 menyatakan banyaknya pengamatan yang nilai responnya, y = 1.

Dengan taraf nyata α, H0 ditolak jika G > x^2(p – 2, α) (nilai kuantil dari distribusi Chi-kuadrat dengan derajat bebas (p – 2)) atau p-value < α (Hosmer dan Lemeshow, 2000).

2.2.2 Uji Parsial

Pengujian secara parsial dilakukan untuk mengetahui signifikansi parameter terhadap variabel respon. Pengujian signifikansi parameter menggunkan uji Wald (Hosmer dan Lemeshow, 2000) dengan hipotesis sebagai berikut.

$H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0$; j = 1, 2, . . . , p

Statistik uji : $W^2 = \left(\dfrac{\hat{\beta} j}{SE\,(\hat{\beta} j)}\right)^2$ (2.9)

Statistik uji W tersebut juga disebut sebagai statistik uji Wald dengan SE ($\hat{\beta}_j$) adalah taksiran galat baku penaksir parameter. Daerah penolakan H_0 adalah jika $W^2 > x^2_{(1,\ \alpha)}$ (nilai kuatil dari distribusi Chi-kuadrat dengan derajat bebas 1).

2.3 Uji Kebermaknaan Pembuangan Variabel

Uji kebermaknaan pembuangan variabel dilakukan dengan uji rasio kemungkinan. Dimana model yang baru kemudian akan dibandingkan dengan model sebelumnya melalui uji rasio kemungkinan.

Missal L_1 adalah log likelihood dari model yang berisi variabel X1, X2, . . . , X_{j-1} dan L_2 adalah log likelihood dari model yang berisi variabel bebas

X1, X2, . . . , X_j. untuk menguji keberartian pembuangan variabel bebas X_j, maka digunakan statistik uji:

$$G = -2\{L_1 - L_2\}$$
$$= -2L_1 - (-2L_2)$$

Statistik G berdistribusi $\chi^2_{(1)}$. Dengan demikian untuk mencari p-valuenya, digunakan p($\chi^2_{(1)}$ > G) = p-value.

2.4 Model Terbaik

Model terbaik akan didapat ketika memenuhi 8 kriteria seleksi (8 *selection criteria* = 8SC). Dalam memperoleh model terbaik, Zainodin dan Khuneswari (2007; 2009) telah menjelaskan secara rinci penggunaan 8SC. Masing-masing model yang dipilih akan diperiksa berdasarkan masing-masing kriteria model seleksi versi 8SC.

8SC tersebut adalah kriteria pemilihan model berdasarkan *akaike information criterion* (AIC), *finite prediction error* (FPE), *generalized cross validation* (GCV), *Hannan and Quinn criterion* (HQ), RICE, SCHWARZ, SGMASQ, dan SHIBATA. Untuk model logistik, deviance statistik digunakan sebagai kriteria pemilihan model. Seperti yang didefinisikan oleh Kutner, dkk (2008), statistik deviance (jumlah kuadrat residual penyimpangan) adalah

$$G^2 = -2\sum_{i=1}^{n}[Y_i \ln(\hat{p}_i) + (1 - Y_i)\ln(1 - \hat{p}_i)] \qquad (2.10)$$

Vogelvang (2005) menyatakan bahwa memaksimalkan kemungkinan (*minimising deviance*) identik dengan meminimalkan jumlah kuadrat error (SSE). Oleh karena itu modifikasi disarankan pada delapan kriteria seleksi dimana SSE digantikan oleh statistik deviance dimana G^2 adalah -2 kali likelihood tersebut.

Table 2.1 Delapan Kriteria Pemilihan yang Dimodifikasi

DELAPAN KRITERIA PEMILIHAN YANG DIMODIFIKASI (M8SC)	
AIC : $\left(\frac{G^2}{n}\right)(e)^{2(p+1)/n}$	RICE : $\left(\frac{G^2}{n}\right)\left[1 - \frac{2(p+1)}{n}\right]^2$
FPE : $\left(\frac{G^2}{n}\right)\frac{n+p+1}{n-(p+1)}$	SCHWARZ : $\left(\frac{G^2}{n}\right)(n)^{2(p+1)/n}$
GCV : $\left(\frac{G^2}{n}\right)\left[1 - \frac{p+1}{n}\right]^2$	SGMASQ : $\left(\frac{G^2}{n}\right)\left[1 - \frac{p+1}{n}\right]^2$
HQ : $\left(\frac{G^2}{n}\right)(\ln n)^{2(p+1)/n}$	SHIBATA : $\left(\frac{G^2}{n}\right)\frac{n+2(p+1)}{n}$

Kriteria ini berupa jumlah sisa kuadrat penyimpangan (*statistic deviance* atau G^2 sebagaimana didefinisikan dalam persamaan (2.10)) dikalikan dengan faktor utama yang tergantung pada kerumitan model yang diukur oleh jumlah parameter (p+1) untuk diperkirakan. Ada suatu kondisi yang harus dipenuhi bila menggunakan kriteria pemilihan model ini, yaitu 2 (p+1) < n. Setelah semua kriteria dihitung, model terbaik dapat diperoleh dengan memilih model yang memiliki nilai terendah untuk sebagian besar kriteria.

2.5 Goodness-of-Fit Test

Setelah model terbaik telah diperoleh, analisis residual dilakukan untuk memeriksa kesesuaian model terbaik. Analisis residual regresi logistik lebih sulit dari pada regresi biasa karena variabel dependen Y_i menggambil hanya nilai 0 dan 1. Oleh karena itu residu u_i akan menggambil salah satu dari nilai-nilai yang sesuai dengan Y_i :

$$u_i = \begin{cases} 1 - \hat{p}_i, jika\ y_i = 1 \\ \hat{p}_i, jika\ y_i = 0 \end{cases} \quad \text{for } i = 1, 2, \ldots, n \qquad (2.11)$$

Menurut Kutner, dkk, (2008), residual tidak akan berdistribusi normal. Plot residual biasa terhadap nilai-nilai prediksi atau variabel independen umumnya tidak akan informatif. Ada dua tes yang disarankan oleh Kutner, dkk, (2008) : *pearson chi-square goodness-of-fit test* dan *residual deviance goodness-of-fit test.* Rosado dkk, (2006) juga menyarankan menggunakan *pearson chi-square goodness-of-fit test* untuk mengevaluasi penyesuaian model. *Pearson residual* seperti yang didefinisikan oleh Kutner, dkk, (2008) adalah :

$$r_{pi} = \frac{Y_i - \hat{p}_i}{\sqrt{\hat{p}_i\,(1-\hat{p}_i)}} \qquad (2.12)$$

Dimana $\sqrt{\hat{p}_i\,(1-\hat{p}_i)}$ adalah perkiraan standar error Y_i untuk i = 1, 2, . . . , n. *residual pearson* secara numerik sama dengan statistik uji *chi-square pearson.* statistic uji *chi-square pearson.*

$$x^2_{r_{pi}} = \sum_{i=1}^{n} \frac{(Yi-\hat{p}_i)^2}{\hat{p}_i\,(1-\hat{p}_i)} \qquad (2.13)$$

Residual Deviance (seperti yang didefinisikan oleh Kutner, dkk 2008) adalah

$$\text{Devi} = \text{sign}\ (Y_i - \hat{p}_i)\sqrt{-2[y_i\ \ln(\hat{p}_i) + (1-y_i)\ln(1-\hat{p}_i)]}$$

$$\text{Dimana sign} = \begin{cases} + \text{ jika } Y_{i \geq \hat{P}_i} \\ - \text{ jika } Y_{i < \hat{P}_i} \end{cases} \quad \text{untuk } i = 1,2,\ldots, n.$$

Selain dua tes ini, scatter plot residual dapat digunakan sebagai bukti pendukung untuk memeriksa kelayakan model menurut Kutner, dkk (2008), ada tiga plot residual yang umum digunakan dalam regresi logistik biner : residual biasa terhadap estimasi probabilitas, residu pearson melawan estimasi probabilitas dan residu deviance terhadap estimasi probabilitas. Plot menunjukan bahwa jika model benar, plot residual terhadap probabilitas (Pi) harus menghasilkan garis lurus dengan nol intercept.

C. Hasil Penelitian

Variabel yang digunakan pada penelitian ini adalah variabel prediktor dan variabel respon pada data nasabah debitur di bank A. Data tersebut berisi tentang nasabah yang memiliki fasilitas kredit (debitur). variabel respon yang digunakan adalah status kredit yaitu :

$$y_i = \begin{cases} 0, \text{ tidak mengalami kredit macet lebih dari 90 hari} \\ 1, \text{ mengalami kredit macet lebih dari 90 hari} \end{cases}$$

Tabel 3.1 Variabel yang digunakan dalam Penelitian

No	Nama Variabel	Simbol Variabel	Skala Pengukuran
1	Jumlah Saldo Kartu Kredit	X_1	Persentase
2	Usia Debitur	X_2	Bilangan bulat
3	Pembayaran Utang Bulanan	X_3	Persentase
4	Pendapatan Bulanan	X_4	Bilangan riil

1. Menyusun Semua Kemungkinan Model

Tabel 3.2 Semua kemungkinan model untuk 4 variabel bebas

Jumlah variabel bebas	Tunggal	interaksi			Total
		Orde-Kesatu	Orde-Kedua	Orde-Ketiga	
1	4	-	-	-	4
2	6	6	-	-	12
3	4	4	4	-	12
4	1	1	1	1	4
Total	15	11	5	1	32
Nomer Model	M_1-M_{15}	M_{16}-M_{25}	M_{26}-M_{31}	M_{32}	

2. Uji Multikolinieritas

Uji multikolieritas dilakukan untuk mengetahui ada tidaknya kebergantungan linier diantara variabel-variabel bebas.

Hipotesis:

H_0 : Tidak terdapat multikolinieritas diantara peubah bebas

H_1 : Terdapat multikolinieritas diantara peubah bebas

Hasil uji multikolinieritas dari data nasabah debitur di bank A, menggunakan software SAS adalah sebagai berikut:

Tabel 3.3 Uji Multikolinieritas

Variabel Bebas	VIF
X1	1.00008
X2	1.00146
X3	1.00083
X4	1.00230

Hasil Tabel 3.3, menunjukkan bahwa dari ke 4 variabel bebas memiliki nilai VIF < 5, berarti tidak terdapat multikolinieritas diantara peubah bebas.

3. Menaksir Model

Dilakukan penaksiran semua model yang sudah disusun pada Tabel 3.2, sebagai contoh hasil dari penaksiran model M22 diperoleh $\beta_0 = 0.9945$, $\beta_1 = -0.00017$, $\beta_2 = 0.0332$, $\beta_3 = -2.49$, $\beta_{12} = 6.806$, $\beta_{13} = -3.12$, $\beta_{23} = 4.244$.

4. Uji Global

Uji global dilakukan untuk setiap model yang sudah ditaksir, dengan tujuan untuk menyelidiki apakah semua koefisien regresinya bernilai nol. Sebagai contoh untuk model M22 hipotesisnya sebagai berikut:

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_{12} = \beta_{13} = \beta_{23} = 0$

H_1 : Paling tidak terdapat satu $\beta_j \neq 0$; j = 1, 2, 3, 12, 13, 23.

Hasil uji global pada model M22, diperoleh statistik uji G = 2087.5713 dengan p-value <.0001.

Hasil uji global, menunjukkan bahwa nilai p-value < α, dimana α=5% maka H_0 ditolak, berarti paling tidak terdapat satu β_j tidak sama dengan nol.

5. Uji Parsial dengan Uji Wald

Uji Parsial dilakukan untuk mengidentifikasi koefisien yang tidak signifikan dan membuang variabel yang tidak signifikan tersebut dari model.

Hipotesis:

$H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0$

Hasil pengujian untuk setiap koefisien model pada M22 diperoleh $\beta 0$ = 0.9945, $\beta 1$ = -0.00017, $\beta 2$ = 0.0332, **$\beta 3$ = -2.49**, $\beta 12$ = 6.806, $\beta 13$ = -3.12, $\beta 23$ = 4.244.

Hasil pengujian, menunjukkan bahwa ada koefisien dalam model yang tidak signifikan, karena nilai p-valuenya > alpha. Variabel bebas yang tidak signifikan tersebut akan dibuang dimulai dari yang p-valuenya paling besar, yaitu variabel X3, selanjutnya dilakukan uji parsial tanpa melibatkan variabel X3.

Hasil pengujian untuk setiap koefisien model pada M22.1 tanpa melibatkan X3, sebagai berikut β_0 = 0.9940, **β_1 = -0.00017**, β_2 = 0.0332, β_{12} = 6.812, β_{13} = -3.12, β_{23} = 3.772.

Hasil pengujian, menunjukkan bahwa ada koefisien dalam model yang tidak signifikan, karena nilai p-valuenya > alpha. Variabel bebas yang tidak signifikan tersebut akan dibuang dimulai dari yang p-valuenya paling besar, yaitu variabel X1, selanjutnya dilakukan uji parsial tanpa melibatkan variabel X1 hingga disebut sebagai model M22.2 dan seterusnya hingga diperoleh p-value yang < alpha. Hal ini diperoleh setelah dilakukan 5 kali pembuangan variabel.

Hasil pengujian untuk setiap koefisien model pada M22.5 tanpa melibatkan X2*X3 diperoleh β_0 = 0.9905 dan β_2 = 0.0334.

Hasil pengujian menunjukkan bahwa nilai p-value < α, dimana α=5% maka H_0 ditolak, berarti variabel dalam model signifikan.

6. Uji Kebermaknaan Pembuangan Variabel

Uji kebermaknaan pembuangan variabel dilakukan dengan uji rasio kemungkinan. Dimana model yang baru kemudian akan dibandingkan dengan model sebelumnya melalui uji rasio kemungkinan.

Untuk model 17.1 sebagai berikut:

M17.1 variabel bebas di dalam model : X3 dan X1*X3

$-2L_1 = 73590.361$

M17 variabel bebas di dalam model : X1, X3, dan X1*X3

$-2L_2 = 73587.952$

$G = -2L_1 - (-2L_2)$

$= 73590.361 - 73587.952 = 2.409$

$P(\chi^2_{(1)} > 2.409)$

p-value = 0.1206

α = 5%

jika p-value > α, maka model 17.1 dan model 17 adalah sama saja.

Artinya pembuangan variabel X1 tidak masalah.

Sehingga didapat kandidat model terpilih yaitu M3, M4, M9, M10, M14, M17.1, M20.1, M21, M22.5, M25.1 dan M29.1.

7. Model Terbaik

Menghitung nilai-nilai kriteria pemilihan model M8SC, dan kemudian memilih model terbaik. Nilai-nilai dari setiap model yang didapat dengan menggunakan M8SC ditunjukkan pada Tabel C.4.

Tabel 3.4 Nilai untuk Setiap Model M8SC

Model	G^2	p+1	AIC	FPE	GCV	HQ	RICE	SCHWARZ	SGMASQ	SHIBATA
M3	73595.342	2	0.4906	0.4906	0.4906	0.4907	0.4906	0.4908	0.4906	0.4906
M4	60390.895	2	0.5021	0.5021	0.5021	0.5022	0.5021	0.5023	0.5021	0.5021
M9	59231.564	3	0.4925	0.4925	0.4925	0.4925	0.4925	0.4928	0.4925	0.4925
M10	60379.508	3	0.5021	0.5021	0.5021	0.5021	0.5021	0.5023	0.5020	0.5021
M14	59223.845	4	0.4925	0.4925	0.4925	0.4925	0.4925	0.4928	0.4924	0.4925
M17.1	73590.361	3	0.4906	0.4906	0.4906	0.4906	0.4906	0.4909	0.4906	0.4906
M20.1	59221.108	3	0.4924	0.4924	0.4924	0.4924	0.4924	0.4927	0.4924	0.4924
M21	60295.077	4	0.4924	0.4924	0.4924	0.4925	0.4924	0.4928	0.4924	0.4924
M22.5	**71539.586**	**2**	**0.4769**	**0.4769**	**0.4769**	**0.4770**	**0.4769**	**0.4771**	**0.4769**	**0.4769**
M25.1	59114.558	6	0.4916	0.4916	0.4916	0.4916	0.4916	0.4921	0.4915	0.4916
M29.1	59100.654	7	0.4915	0.4915	0.4915	0.4915	0.4915	0.4921	0.4914	0.4915

Semua kriteria pada Tabel 3.4 menunjukkan bahwa model M22.5 sebagai model terbaik. Hasil uji parsial dengan uji wald menunjukkan bahwa seluruh koefisien dalam model terbaik M22.5 signifikan. Persamaan model terbaik M22.5 adalah sebagai berikut:

$$\ln \frac{\hat{\pi}}{1-\hat{\pi}} = 0.9905 + 0.0034X_2.$$

8. Uji Goodness of-fit

Dilakukan untuk pemeriksaan validitas dari model terbaik dengan melakukan uji kecocokan berdasarkan uji deviance dan uji pearson Chi-square, Hasil uji kecocokan model terbaik M22.5, menggunakan software SAS adalah sebagai berikut:

Tabel 3.5 Uji Kecocokan

Criterion	Df	Value	Value/df
Deviance	**15E4**	**9231.3336**	**0.0615**
Pearson Chi-Square	**15E4**	**9231.3336**	**0.0615**

Hipotesis untuk uji deviance pada model M22.5 adalah sebagai berikut:

H_0: $E\{Y\}=[1+\exp(-X\boldsymbol{\beta})]^{-1}$ (model cocok)

H_1: $E\{Y\}\neq[1+\exp(-X\boldsymbol{\beta})]^{-1}$ (model tidak cocok)

Hasil deviance, menunjukkan bahwa nilai p-value $0.0615 > \alpha$, dimana α=5% maka H_0 diterima, berarti model cocok.

Hipotesis untuk uji pearson chi-square pada model M22.5 adalah sebagai berikut:

H_0 : $E\{Y\}=[1+\exp(-X\boldsymbol{\beta})]^{-1}$ (model cocok)

H_1 :$E\{Y\}\neq[1+\exp(-X\boldsymbol{\beta})]^{-1}$ (model tidak cocok)

Hasil uji pearson, menunjukkan bahwa nilai p-value $0.0615 > \alpha$, dimana α=5% maka H_0 diterima, berarti model cocok.

Berdasarkan uji deviance dan uji pearson menunjukkan bahwa model yang hanya melibatkan variabel X_2 adalah model terbaik.

D. Kesimpulan

Berdasarkan hasil pembahasan pada bab sebelumnya maka diperoleh kesimpulan bahwa dari 4 variabel bebas diperoleh kombinasi model sebanyak 32 model yang kemudian untuk setiap model ditaksir berdasarkan uji global dan uji parsial dengan uji wald, sehingga diperoleh 11 model yang signifikan. Dari ke 11 model dihitung nilai-nilai kriteria pemilihan model M8SC dan dipilih M22.5 karena setiap nilai-nilai kriteria pada M22.5 nilainya lebih kecil dari model yang lainnya.

Maka didapat model terbaik pada data nasabah debitur di bank A adalah model M22.5 dengan persamaan:

$$\ln \frac{\hat{\pi}}{1-\hat{\pi}} = 0.9905 + 0.0034X_2.$$

Daftar Pustaka

Akaike, H. 1974. A New Look at The Statistical Model Identification. *IEEE Transactions on Automatic Control AC* 19: 716-723.

De Jong, Piet, and Gillian Z. Heller.*Generalized Linier Model for Insurance Data*. New York : Cambridge University Press, 2008.

Hajarisman, N. 2008. *Analisis Data Kategorik*.Bandung :tidakditerbitkan.

Hosmer, D.W. and Lemeshow, S. 2000. *Applied Logistic Regression*. New York : Inc. John Wiley and Sons.

Kutner, M.H., Nachtsheim, G.J., Neter, J. & Li, W. 2008. *Applied Linier Regression Models.* (4th edition). Singapore: McGraw-Hill, Inc.

Ramanathan, R. 2002. *Introductory Econometrics With Applications.* 5th edition. Ohio: South Western, Thomson Learning Ohio.

Rosado, J.F.C, Solis, C.E.M., Sanches, A.A.V., Rosado, A.J.C., Prado, B.H. and Burgos, L.A. 2006. *Prevalence and associated factors for temporomandibular disorders in a group of Mexican adolescents and yauth adults. Clin Oral Invest,* 10: 42-49.

Taswan. 2003. *Akuntansi Perbankan Transaksi Dalam Valuta Rupiah,* Edisi Revisi. Semarang: UPP AMP YKPN.

Vogelvang, B. 2005. *Econometrics*: *Theory and Applications with Eviews*. New York: Addison-Wesley.

Zainodin, H.J. and Khuneswari, G. 2007. *Justification of Omitting Independent Variables in Selecting Best Model.* Proceedings of International Applied Science and Mathematics Conference, 343-356.

---, 2009. *A Case Study on Determination of House Selling Price Model Using Multiple Regression.*, Malaysian Journal of Mathematical Sciences, 3(1):27-44.

---, 2010. *Model-Building Approach in Multiple Binary Logit Model for Coronary Heart Disease. Malaysian Journal of Mathematical Sciences*, 4(1): 107-133.