

PERFORMANSI MODEL-MODEL *CREDIT SCORING* PADA DATA DEBITUR KREDIT PRODUKTIF PT. BANK X

Astri Afrilia

Universitas Padjadjaran

Program Pendidikan Magister Program Studi Statistika Terapan, Konsentrasi
Statistika Sosial

Email : astri_afrilia@yahoo.com

Abstrak. Penelitian ini mengkaji mengenai metode-metode *credit scoring* yang dapat diaplikasikan pada data debitur kredit produktif Bank X. Metode-metode tersebut yakni metode regresi logistik biner, CART, CHAID dan MARS. Setiap metode menghasilkan model yang kemudian dievaluasi performansinya berdasarkan nilai *error rate* serta error tipe I dan II. Tujuan analisis adalah untuk memperoleh metode *credit scoring* terbaik yang paling sesuai untuk data dimaksud. Metode terbaik ditandai dengan nilai *error rate*, error tipe I dan II yang terkecil dibandingkan metode lain.

Kata Kunci : Regresi logistik biner, *Classification and Regression Tree* (CART), *Chi Square Automatic Interaction Detection* (CHAID), *Multivariate Adaptive Regression Spline* (MARS).

1. Pendahuluan

Dalam setiap penyaluran kredit terkandung risiko yang ada di dalamnya yang dinamakan risiko kredit. Pada PT. Bank X, risiko kredit yang tengah disoroti saat ini adalah risiko kredit pada produk kredit produktif dikarenakan produk tersebut menempati urutan kedua dalam 4 (empat) pilar utama produk Bank X. Jumlah debitur kredit produktif Bank X sampai saat ini berjumlah 3.503 debitur dimana 6,9% diantaranya tergolong debitur bermasalah. Statistik berperan untuk membuat perangkat penilaian kredit dengan data historis tersebut dengan tujuan meminimalisir risiko kredit sehingga calon debitur yang memiliki risiko kredit cukup tinggi dapat tersaring lebih awal ketika ada pengajuan kredit. Perangkat penilaian kredit tersebut saat ini dikenal dengan istilah *credit scoring*.

Sabzevari *et al.* (2007) menyebutkan bahwa metode *credit scoring* pertama kali digunakan pada era tahun 1950an dan kebanyakan digunakan untuk jenis kredit konsumtif seperti kartu kredit dan kredit multiguna. Penelitian yang dilakukannya bertujuan untuk menggali performansi model *credit scoring* menggunakan Metode Regresi Logistik Biner dan teknik data mining seperti *Classification and Regression Tree* (CART), *Chi-squared Automatic Interaction Detection* (CHAID), *Neural Network*, serta *Multivariate Adaptive Regression Spline* (MARS). Data yang digunakan adalah nasabah pengguna kredit Bank Karafarin Iran dengan jumlah sampel yang terbatas. Dewasa kini, metode *Neural*

Network bisa menjadi alternatif untuk membangun model *credit scoring* dan digunakan ketika variabel dependen dan independen memiliki hubungan kompleks yang nonlinear. Meskipun seringkali metode ini diklaim sebagai metode yang lebih baik dibandingkan analisis diskriminan dan regresi logistik, namun metode *Neural Network* memiliki beberapa kelemahan antara lain proses pengerjaan yang lebih lama dibandingkan keduanya dalam mencapai topologi jaringan yang optimal, kesulitan dalam mengidentifikasi variabel input yang potensial, serta juga aplikasinya yang terbatas dalam menangani permasalahan *credit scoring*. Dalam rangka menguji efektifitas metode-metode *credit scoring*, peneliti menggunakan data dari sebuah bank kecil di Iran yang berjumlah 719 data nasabah kredit, 413 diantaranya tergolong nasabah kategori “baik” dan sisanya kategori “buruk”. Penelitian ini mencakup data finansial, spesifikasi nasabah, jenis dan jumlah pinjaman serta performansi debitur dalam melakukan pembayaran.

Sabzevari *et al.* (2007) kemudian melakukan evaluasi dari setiap model *credit scoring* yang diperoleh sehingga didapatkan performansi masing-masing model. Yang Liu (2002), dalam penelitiannya memaparkan evaluasi model merupakan alat yang dibutuhkan untuk membantu pencarian dan pemilihan solusi yang terbaik serta berguna untuk pemilihan teknik/metode yang cocok atas data dan seleksi atas model final yang dapat diaplikasikan pada data.

Berangkat dari latar belakang sebagaimana yang telah dipaparkan di atas, maka penulis bermaksud melakukan penelitian untuk menggali performansi beberapa metode *credit scoring* seperti regresi logistik biner, CART, CHAID dan MARS pada data debitur produk kredit produktif Bank X.

2. Model Regresi Logistik Biner

Model regresi logistik biner secara umum memiliki bentuk: .

$$\pi(x_i) = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi})}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi})}; i = 1, 2, \dots, N \quad (2.1)$$

Dalam pemodelan *credit scoring*, variabel respon Y_i merupakan variabel yang memiliki 2 (dua) buah kategori atau biner (*dichotomus*). $Y_i = 1$ merupakan nilai kualitatif untuk kualitas kredit debitur yang tergolong *non performing loan* dengan peluang $\pi(x_i)$, sedangkan $Y_i = 0$ menyatakan nilai kualitatif untuk kualitas kredit debitur yang tergolong *performing loan* dengan peluang $1 - \pi(x_i)$. Sedangkan $X = (X_1, X_2, \dots, X_p)$ merupakan variabel bebas yang dapat tergolong kualitatif, kuantitatif, ataupun campuran antar keduanya. (Agresti, 1996).

3. Model CART

$$\Phi(s|t) = 2P_L P_R \sum_{j=1}^{\text{jumlah kategori}} |P(j|t_L) - P(j|t_R)|, \quad (2.2)$$

dengan

t_L = calon cabang kiri dari noktah keputusan

t_R = calon cabang kanan dari noktah keputusan

P_L = peluang t_L terjadi

P_R = peluang t_R terjadi

dan

$$P_L = \frac{\text{jumlah catatan pada calon cabang kiri } t_L}{\text{jumlah catatan pada data latihan}}, \quad (2.3)$$

$$P_R = \frac{\text{jumlah catatan pada calon cabang kanan } t_R}{\text{jumlah catatan pada data latihan}}, \quad (2.4)$$

$$P(j|t_L) = \frac{\text{jumlah catatan berkategori } j \text{ pada calon cabang kiri } t_L}{\text{jumlah catatan pada noktah keputusan } t}, \quad (2.5)$$

$$P(j|t_R) = \frac{\text{jumlah catatan berkategori } j \text{ pada calon cabang kanan } t_R}{\text{jumlah catatan pada noktah keputusan } t}, \quad (2.6)$$

$$Q(s|t) = \sum_{j=1}^{\text{jumlah kategori}} |P(j|t_L) - P(j|t_R)|, \quad (2.7)$$

4. Model CHAID

CHAID digunakan untuk membentuk segmentasi yang membagi sebuah sampel menjadi dua atau lebih kelompok yang berbeda berdasarkan sebuah kriteria tertentu. Hal ini kemudian diteruskan dengan membagi kelompok-kelompok tersebut menjadi kelompok yang lebih kecil berdasarkan variabel-variabel independen yang lain. Hasil dari pengklasifikasian dalam CHAID akan ditampilkan dalam sebuah diagram pohon. (Kunto dan Hasana, 2006). Algoritma CHAID adalah sebagai berikut.

1. Bentuk tabel kontingensi dua arah untuk masing-masing variable independen dengan variabel dependennya.
2. Hitung statistik *chi-square* untuk setiap pasang kategori yang dapat dipilih untuk digabung menjadi satu, Langkah uji *chi-square* adalah sebagai berikut :
 - a. Menuliskan hipotesis•
 - $H_0 : p_{ij} = p_{i\bullet} \cdot p_{\bullet j}$ (Variabel *i* dan variabel *j* independen atau tidak terdapat hubungan antara variabel *i* dan variabel *j*)
 - $H_1 : p_{ij} \neq p_{i\bullet} \cdot p_{\bullet j}$ (Variabel *i* dan variabel *j* dependen atau terdapat hubungan antara variabel *i* dan variabel *j*)
 - b. Menentukan $\alpha = 0.05$
Mencari harga $\chi^2_{0,05;(b-1)(k-1)}$ dari tabel *chi-square*.
 - c. Menentukan daerah penolakan, yaitu $\chi^2 > \chi^2_{0,05;(b-1)(k-1)}$ atau *p-value* $< \alpha$
 - d. Mencari $\chi^2 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$
 - e. Mengambil kesimpulan, yakni bila χ^2 masuk daerah penolakan, H_0 ditolak dan sebaliknya bila χ^2 tidak masuk dalam daerah penolakan, H_0 diterima
3. Untuk masing-masing nilai *chi-square* berpasangan, hitung *p-value* berpasangan bersamaan. Diantara pasangan-pasangan yang tidak signifikan, gabungkan sebuah pasangan kategori yang paling mirip (yaitu pasangan yang mempunyai nilai *chi-square* berpasangan terkecil dan *p-value* terbesar) menjadi sebuah kategori tunggal, dan kemudian dilanjutkan ke langkah nomor 4. Jika terdapat pasangan dengan *p-*

value lebih besar dari taraf signifikansinya, maka pasangan tersebut akan digabungkan.

5. Model MARS

$$\hat{f}(x) = a_0 + \sum_{m=1}^M a_1 \prod_{k=1}^{k_m} H[S_{1m} \cdot (x_{v(1,m)} - t_{1m})] + \sum_{m=1}^M a_2 \prod_{k=1}^{k_m} H[S_{2m} \cdot (x_{v(2,m)} - t_{2m})] + \dots \quad (2.8)$$

No.	Variabel	Definisi Variabel
1.	a_0	Konstanta
2.	a_m	Koefisien dari basis fungsi ke- m
3.	M	Maksimum basis fungsi (<i>nonconstant basis function</i>)
4.	k_m	Derajat interaksi
5.	S_{km}	Indeks bernilai ± 1
6.	$x_{v(k,m)}$	Variabel prediktor/ independen
7.	t_{km}	Nilai <i>knot</i> dari variabel prediktor/independen

6. Evaluasi Model

$$\hat{P}_{err} = \frac{\text{Jumlah data yang salah diklasifikasikan}}{\text{Jumlah seluruh data}}, \quad (2.9)$$

$$= \frac{n_{12} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}},$$

dengan

\hat{P}_{err} : error rate

Tabel 6 Tabel Persentase Prediksi

Kelas Prediksi	Kelas Aktual			
	Jumlah		Persentase	
	Baik	Buruk	Baik	Buruk
Baik	n_{11}	n_{12}	$1-\alpha$	$\frac{n_{12}}{n_{11} + n_{12}} = \alpha$
Buruk	n_{21}	n_{22}	$\frac{n_{21}}{n_{21} + n_{22}} = \beta$	$1-\beta$

Maka, *Error Tipe 1* = α dan *Error Tipe 2* = β

7. Data dan Hasil Penaksiran

No.	Metode	<i>Error Rate</i>	<i>Error Tipe 1</i>	<i>Error Tipe 2</i>
1.	Regresi Logistik Biner	0,0244	0,7155	0,0000
2.	CART	0,1069	0,1494	0,1031
3.	CHAID	0,0674	1,0000	0,0000
4.	MARS	0,0018	0,0909	0,0000

8. Kesimpulan

Metode regresi logistik biner, CART, CHAID dan MARS sebagai metode statistik yang dapat digunakan sebagai metode *credit scoring*, dapat diaplikasikan pada data debitur kredit produktif Bank X. Metode MARS merupakan metode terbaik yang dapat digunakan sebagai metode *credit scoring* pada data debitur kredit produktif Bank X karena memiliki nilai *error rate*, error tipe I dan II yang terkecil, yakni *error rate* sebesar 0,0018 error tipe I sebesar 0,0909 dan error tipe II sebesar 0,0000. Model MARS yang diperoleh dapat dijadikan alternatif model *credit scoring* PT. Bank X karena metode ini memiliki resiko terkecil dalam kesalahan memprediksi kelompok nasabah baik resiko komersial (error tipe I) serta resiko kredit (error tipe II).

9. Daftar Pustaka

- Agresti, A. 1996. *An Introduction to Categorical Data Analysis*. New York: John Wiley and Sons, Inc.
- Breiman, L. 1984. *Classification and Regression Tree*. California: Pacific Grove.
- Breiman, L. 1993. *Bagging Predictor*. New York: Department of Statistic University of California.
- Boges, W. P. 1967. *Screen Test Your Credit Risks*. Harvard Business Review.
- Dobson, Annette J. 1983. *Introduction to Statistical Modelling*. London: Chapman and Hall.
- Du Toit, Steyn and Stump. 1986. *Graphical Exploratory Data Analysis*. New York: Springer.
- Friedman, J. H. 1991. *Multivariate Adaptive Regression Splines, The Annals of Statistics*. New York: Jstor and Institute of Mathematical Statistics.

- Galagher. 2000. *Risk Classification Aided by New Software Tool*. New York: Guilford Press.
- Hosmer, D. W. and Lemeshow, S. 1989. *Applied Logistic Regression*. New York: John Wiley and Sons, Inc.
- Joos, P. , and Vanhoof, K. 1998. *Credit Classification: A Comparison of Logit Models and Decision Trees*. European Conference on Machine Learning.
- Kunto, Y. S. , and Hasana, S. N. 2006. *Analisis CHAID sebagai Alat Bantu Statistika untuk Segmentasi Pasar*. Surabaya: Universitas Kristen Petra.
- Lee, Thyan Shyug. 2004. *Mining the Customer Credit Using Classification and Regression Tree and Multivariate Adaptive Regression Splines*. Taiwan:
- Lehmann, T. , and Enheler, D. 2001. *Responden Profilling with CHAID and Dependency*. New York: Guilford Press.
- Liu, Yang. 2002. *The Evaluation of Classification Models for Credit Scoring*. Gottingen: Gottingen University.
- Lohrbach, T. 1994. *Einsatz van Kunstliclen Neuronalen Netzen fur Ausgewahhe*. Gottingen: Gottingen Wirstchafts Informatik.
- Nash, M. S. , and Bradford, D. F. 2001. *Parametric and Non Parametric Logistic Regression for Prediction of Presence/Absence of an Application*. Las Vegas: Nevada.
- Otok, B. W. 2007. *M.A.R.S (Multivariate Adaptive Regression Splines)*. Yogyakarta: Universitas Gajah Mada.
- Weiss, S. M. , and Indurkhya, N. 1998. *Predictive Data Mining: A Practical Guide*. San Fransisco: Morgen Kaufmann Publishers.
- Witten, I. H. , and Frank, E. *Data Mining: Pratical Machine Learning Tools and Techniques with Java Implementations*. San Fransisco: Morgen Kaufmann Publishers.