

## MODEL PENILAIAN KREDIT MENGGUNAKAN ANALISIS DISKRIMINAN DENGAN VARIABEL BEBAS CAMPURAN BINER DAN KONTINU

Moch. Abdul Mukid<sup>1</sup>, Tatik Widiharih<sup>2</sup>

<sup>1,2</sup> Staf Pengajar Departemen Statistika Universitas Diponegoro Semarang

e-mail: <sup>1</sup> mamukid@yahoo.com; <sup>2</sup> widiharih@gmail.com

DOI: 10.14710/medstat.9.2.107-118

### Abstract

Credit scoring models is an important tools in the credit granting process. These models measure the credit risk of a prospective client. This study aims to applied a discriminant model with mixed predictor variables (binary and continuous) for credit assesment. Implementation of the model use debitur characteristics data from a bank in Lampung Province which the used binary variables involve sex and marital status. Whereas, the continuous variables that was considered appropriate in the model are age, net income, and length of work. By using the data training, it was known that the misclassification of the model is 0.1970 and the misclassification of the testing data reach to 0.3753.

**Keywords:** *discriminant analysis, mixed variables, credit scoring*

### 1. PENDAHULUAN

Kredit adalah sebuah katalis penting untuk pertumbuhan ekonomi dan merupakan aktivitas inti dari bank-bank di seluruh dunia. Menurut The Hong Kong Institute of Bankers (2012), sukses atau gagal sebuah bank dan industri keuangan secara umum bergantung pada sistem yang digunakan untuk mengatur kredit dan seberapa baik resiko kredit tersebut ditangani. Ketersediaan kredit memungkinkan rumah tangga untuk melakukan konsumsi yang lebih baik dan memungkinkan perusahaan untuk melakukan investasi yang tidak bisa dilakukan dengan dana sendiri. Namun adanya permasalahan *moral hazard* dan *adverse selection*, bank memainkan peran penting dalam mengalokasikan kapital dan melakukan pemantauan untuk memastikan bahwa dana masyarakat disalurkan pada kegiatan yang memberikan keuntungan optimal (Utari dkk, 2012). Salah satu cara agar alokasi kapital mampu tepat sasaran adalah dengan membuat prediksi mengenai kemampuan bayar calon konsumen di waktu yang akan datang. Oleh karena itu diperlukan sebuah model prediksi.

Model-model penilaian kredit telah memainkan sebuah peran penting dalam praktek manajemen resiko perbankan kontemporer. Mereka berkontribusi terhadap kunci pokok yaitu pada proses persetujuan pinjaman yang secara akurat dan efisien mengkuantifikasi level resiko kredit seorang calon peminjam. Model-model penilaian kredit ini bertujuan untuk memprediksi perilaku masa depan dalam hal resiko kredit yang didasarkan atas pengalaman masa lalu dari nasabah-nasabah yang memiliki karakteristik mirip. Level dari resiko kredit seorang peminjam dihubungkan dengan peluang bahwa ia

akan gagal membayar pinjaman yang disetujui pada waktu yang telah ditentukan. Tugas utama dari model penilaian kredit adalah menyediakan pemisah antara mereka yang gagal dan mereka yang tidak gagal dalam hal pembayaran kredit. Kemampuan memisahkan adalah indikator kunci dari kesuksesan sebuah model (Nicolic *et al.*, 2013).

Sampai dengan saat ini, terdapat tiga teknik dasar yang digunakan untuk membuat model penilaian kredit, yaitu model penskoran oleh ahli, model-model statistik dan model-model kecerdasan buatan (Li dan Zhong, 2012). Model penskoran oleh ahli adalah pendekatan pertama yang diterapkan untuk menyelesaikan permasalahan penilaian kredit. Ahli (analisis kredit) menyetujui pemberian kredit atau tidak berdasarkan pada karakteristik dari peminjam. Mereka membuat analisis dengan menskor faktor-faktor utama dari kredit seperti kualitas moral, kemampuan membayar, adanya jaminan dari peminjam, tujuan meminjam dan batas pengembalian dari pinjaman. Sayangnya, metode ini sangat bergantung pada pengalaman dan pengetahuan dari analis-analisis kredit yang membuat hal itu menjadi sebuah tugas yang menyita waktu dan menimbulkan keraguan serta kesalahan klasifikasi.

Model-model statistik yang digunakan untuk penilaian kredit dalam rumpun keilmuan statistika dikategorikan sebagai model-model klasifikasi (Hand, 1981; Hsieh, 2004; Lee *et al.*, 2006). Analisis Diskriminan Linear (ADL) dan Analisis Regresi Logistik (ARL) adalah dua buah model klasifikasi parametrik yang sering digunakan untuk penilaian kredit (Crook *et al.*, 2007; Lee *et al.*, 2006). Dalam analisis diskriminan, amatan-amatan pada variabel bebas yang digunakan bertipe numerik. Dalam beberapa kasus kondisi ini sangat membatasi. Krazanowski (1975) mengembangkan sebuah model diskriminan yang mampu melibatkan variabel bebas bertipe kategorik maupun numerik.

Penelitian ini bertujuan menerapkan model diskriminan dengan variabel bebas campuran untuk analisis penilaian kredit. Dengan model ini diharapkan makin banyak model alternatif yang dapat digunakan untuk penilaian kredit.

## 2. TINJAUAN PUSTAKA

Paralel dengan pertumbuhan volume kredit, banyak model penilaian kredit yang telah dikembangkan oleh pihak bank maupun para peneliti dalam rangka untuk mengevaluasi pelamar-pelamar kredit termasuk didalamnya Analisis Diskriminan Linear (ADL), Analisis Regresi Logistik (ARL), *Multiplicative Adaptive Regression Spline* (MARS), *Classification and Regression Trees* (CART), *Artificial Neural Network* (ANN), *Support Vector Machine* (SVM) dan algoritma genetika (Abdou *et al.*, 2008).

Beberapa penelitian *banchmarking* telah pula dilakukan untuk membandingkan secara empiris performa dari teknik-teknik tersebut dalam memperkirakan skor kredit, seperti yang dilakukan Baesens *et al.* (2003). Mereka membandingkan tujuh belas model dengan menggunakan delapan himpunan data real dan diketahui bahwa teknik-teknik yang lebih kompleks cenderung menghasilkan performa yang lebih baik berdasarkan kriteria *Area Under Curve* (AUC). De-La-Vega *et al.* (2013) membandingkan model analisis diskriminan linear, analisis diskriminan kuadratik, regresi logistik, *multilayer perceptron*, SVM, pohon-pohon klasifikasi dan metode-metode gabungan pada data dari sebuah lembaga keuangan mikro dan menyimpulkan bahwa penggunaan *multilayer perceptron* lebih baik dibandingkan dengan model-model yang lainnya. Banyaknya penelitian-penelitian *banchmarking* pada model penilaian kredit sering kali memunculkan kesimpulan-kesimpulan yang terkadang saling bertentangan. Yobas, *et al* (2000)

menemukan bahwa ADL lebih baik daripada ANN sedangkan Desai, *et al* (1996) melaporkan bahwa ANN lebih baik secara signifikan dibandingkan dengan ADL. Sampai dengan saat ini tidak diketahui dengan jelas literatur mana yang menyatakan bahwa ada model yang tepat untuk penilaian kredit (Brown dan Mues, 2012).

## 2.1. Model Diskriminan dengan Variabel Bebas Campuran

Misalkan  $\mathbf{x}$  adalah vektor variabel-variabel biner dengan ukuran  $qx1$  dan  $\mathbf{y}$  adalah vektor variabel-variabel kontinu dengan ukuran  $px1$ . Variabel-variabel biner dapat dinyatakan sebagai variabel multinomial  $\mathbf{z}' = (z_1, z_2, \dots, z_k)$  dengan  $k = 2^q$ . Oleh karena itu setiap nilai  $\mathbf{x}$  yang berbeda mendefinisikan sebuah sel multinomial secara tunggal, dengan  $\mathbf{x}' = (x_1, x_2, \dots, x_q)$  berada di sel  $c = 1 + \sum_{i=1}^q x_i 2^{(i-1)}$ . Selanjutnya dengan mengikuti Olkin dan Tate (1961), untuk permasalahan klasifikasi dari sebuah amatan  $\mathbf{w}' = (\mathbf{x}', \mathbf{y}')$  ke salah satu dari dua populasi,  $\pi_1$  dan  $\pi_2$ , model ini dapat digeneralisasi dengan mengasumsikan bahwa  $\mathbf{y}$  berdistribusi normal multivariat dengan vektor rata-rata di sel  $m$  adalah  $\boldsymbol{\mu}_i^{(m)}$  dan berada di populasi  $\pi_i$  dimana  $i = 1, 2$  dengan  $m = 1, 2, \dots, k$ . Matriks varian kovarian  $\boldsymbol{\Sigma}$  nilainya sama di semua sel  $m$  pada kedua populasi, yaitu

$$(\mathbf{y} | z_m = 1, z_j = 0, j \neq m = 1, 2, \dots, k) \sim N(\boldsymbol{\mu}_i^{(m)}, \boldsymbol{\Sigma})$$

Model diskriminan disetiap sel  $m$  dinyatakan dengan

$$M_m = \left( \boldsymbol{\mu}_1^{(m)} - \boldsymbol{\mu}_2^{(m)} \right)' t \boldsymbol{\Sigma}^{-1} \left\{ \mathbf{y} - \frac{1}{2} \left( \boldsymbol{\mu}_1^{(m)} + \boldsymbol{\mu}_2^{(m)} \right) \right\}$$

## 2.2. Aturan Lokasi

Pada bagian ini diasumsikan bahwa semua parameter populasi diketahui sehingga aturan alokasi dapat diturunkan dari teori umum tentang klasifikasi. Jika  $p_i(\mathbf{w})$  adalah peluang dari  $\mathbf{w}$  berada di  $\pi_i$  ( $i = 1, 2$ ), dapat ditunjukkan bahwa aturan optimalnya adalah menempatkan  $\mathbf{w}$  ke populasi  $\pi_1$  jika  $\frac{p_1(\mathbf{w})}{p_2(\mathbf{w})} \geq 1$ , jika tidak maka  $\mathbf{w}$  ditempatkan ke populasi  $\pi_2$ . Karena  $p_i(\mathbf{w}) = p_i(\mathbf{x}, \mathbf{y}) = p_i(\mathbf{x}) \cdot p_i(\mathbf{y} | \mathbf{x}) = p_{im} \cdot p_i(\mathbf{y} | z_m)$  untuk  $i = 1, 2$ , dan karena  $\mathbf{y}$  berdistribusi Normal multivariat di sel  $m$  maka alokasikan  $\mathbf{w}' = (\mathbf{x}', \mathbf{y}')$  ke populasi  $\pi_1$  jika

$$\left( \boldsymbol{\mu}_1^{(m)} - \boldsymbol{\mu}_2^{(m)} \right)' t \boldsymbol{\Sigma}^{-1} \left\{ \mathbf{y} - \frac{1}{2} \left( \boldsymbol{\mu}_1^{(m)} + \boldsymbol{\mu}_2^{(m)} \right) \right\} \geq \ln \frac{p_{2m}}{p_{1m}}$$

$$\left( \boldsymbol{\mu}_1^{(m)} - \boldsymbol{\mu}_2^{(m)} \right)' t \boldsymbol{\Sigma}^{-1} \left\{ \mathbf{y} - \frac{1}{2} \left( \boldsymbol{\mu}_1^{(m)} + \boldsymbol{\mu}_2^{(m)} \right) \right\} \geq \ln \frac{p_{2m}}{p_{1m}} \tag{1}$$

dan alokasikan  $\mathbf{w}' = (\mathbf{x}', \mathbf{y}')$  ke populasi  $\pi_2$  jika

$$\left( \boldsymbol{\mu}_1^{(m)} - \boldsymbol{\mu}_2^{(m)} \right)' t \boldsymbol{\Sigma}^{-1} \left\{ \mathbf{y} - \frac{1}{2} \left( \boldsymbol{\mu}_1^{(m)} + \boldsymbol{\mu}_2^{(m)} \right) \right\} < \ln \frac{p_{2m}}{p_{1m}} \tag{2}$$

Aturan optimal yang diturunkan dari model lokasi mengarahkan secara efektif pada sebuah diskriminan linear yang berbeda untuk setiap sel multinomial dengan titik-titik *cut-off* ditentukan di masing-masing kasus oleh komponen-komponen diskrit dari model. Kasus dengan variabel biner tunggal dan  $p$  variabel kontinu telah dibahas oleh Chang dan Afifi (1974).

Peluang kesalahan klasifikasi dari  $\pi_1$  dan  $\pi_2$  masing-masing adalah  $P(2|1)$  dan  $P(1|2)$  yang didefinisikan sebagai jumlah dari peluang-peluang kesalahan klasifikasi di setiap sel multinomial dari populasi  $\pi_i$  yang terboboti dengan peluang terjadinya kesalahan klasifikasi di sel tersebut, yaitu

$$P(2|1) = \sum_{m=1}^k p_{1m} \Phi \left\{ \frac{\left( \ln \frac{p_{2m}}{p_{1m}} - \frac{1}{2} D_m^2 \right)}{D_m} \right\} \quad (3)$$

$$P(1|2) = \sum_{m=1}^k p_{2m} \Phi \left\{ \frac{\left( \ln \frac{p_{1m}}{p_{2m}} - \frac{1}{2} D_m^2 \right)}{D_m} \right\} \quad (4)$$

dengan  $\Phi(\cdot)$  adalah fungsi distribusi kumulatif normal standard dan  $D_m^2 = (\boldsymbol{\mu}_1^{(m)} - \boldsymbol{\mu}_2^{(m)})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1^{(m)} - \boldsymbol{\mu}_2^{(m)})$

Dalam praktek, parameter-parameter populasi biasanya tidak diketahui sedemikian sehingga aturan alokasi seperti pada bagian sebelumnya tidak bisa digunakan. Oleh karena itu parameter-parameter tersebut perlu diduga agar aturan alokasi dapat digunakan.

### 2.3. Pendugaan Parameter

Misalkan  $n_{1m}$  dan  $n_{2m}$  adalah banyaknya amatan yang berada di sel  $m$  dari populasi  $\pi_1$  dan  $\pi_2$ , dan misalkan  $\mathbf{y}_{ji}^{(m)}$  adalah vektor variabel-variabel kontinu terkait dengan

amatan ke  $j$  di sel  $m$  dari sampel populasi  $\pi_i$ . Kemudian jika  $\bar{\mathbf{y}}_i^{(m)} = \frac{\sum_{j=1}^{n_{im}} \mathbf{y}_{ji}^{(m)}}{n_{im}}$  maka penduga-penduga kemungkinan maksimum (*maximum likelihood*) dari parameter-parameter populasi  $p_{im}$ ,  $\boldsymbol{\mu}_i^{(m)}$  dan  $\boldsymbol{\Sigma}$  diberikan oleh

$$\hat{p}_{im} = \frac{n_{im}}{n_i} \quad (5)$$

$$\hat{\boldsymbol{\mu}}_i^{(m)} = \bar{\mathbf{y}}_i^{(m)} \quad (6)$$

$$\hat{\boldsymbol{\Sigma}} = \mathbf{V} = \frac{1}{(n_1 + n_2 - 2k)} \sum_{i=1}^2 \sum_{m=1}^k (\mathbf{y}_{ji}^{(m)} - \bar{\mathbf{y}}_i^{(m)}) (\mathbf{y}_{ji}^{(m)} - \bar{\mathbf{y}}_i^{(m)})' \quad (7)$$

dengan  $i = 1, 2$  dan  $m = 1, 2, \dots, k$

Dalam praktek jika  $n_1$  dan  $n_2$  tidak relatif besar terhadap  $k$ , sangat mungkin bahwa beberapa sel akan memiliki amatan yang sedikit atau bahkan tidak ada amatan sama sekali. Dari hal ini, penduga-penduga parameter yang bersesuaian menjadi lemah. Oleh sebab itu diperlukan suatu aproksimasi yang mampu menghasilkan penduga-penduga parameter di semua sel yang mungkin. Krzanowski (1975) mengusulkan untuk menggunakan model log linear, yaitu

$$\log \eta_{im} = \sum_{j=1}^s a_{imj} \theta_j$$

dengan  $\theta_1, \theta_2, \dots, \theta_s$  adalah himpunan konstanta yang belum diketahui dan  $a_{imj}$  adalah koefisien-koefisien yang telah diketahui. Dalam model ini diasumsikan bahwa frekuensi yang teramati  $n_{im}$  di sel  $m$  dari populasi  $\pi_i$  adalah sebuah realisasi dari suatu variat yang memiliki rata-rata  $\eta_{im}$ . Pendugaan dengan metode kemungkinan maksimum (*maximum likelihood*) terhadap parameter-parameter model ini dapat dilakukan dengan menggunakan sebuah prosedur iterative sederhana. Selanjutnya dugaan bagi parameter  $p_{im}$  diberikan oleh

$$\tilde{p}_{im} = \frac{\hat{\eta}_{im}}{n_i}$$

Kemudian parameter-parameter  $\mu_i^{(m)}$  dan  $\Sigma$  yang berkaitan dengan variabel-variabel kontinu diduga dengan menggunakan bantuan model aditiv dengan komponen-komponen yang dapat diinterpretasikan sebagai efek-efek utama dari variabel biner dan efek-efek interaksi antara variabel-variabel biner. Oleh karena itu vektor rata-rata variabel-variabel kontinu di populasi  $\pi_i$  dapat ditulis dengan

$$\mu_i = \mathbf{v}_i + \sum_{j=1}^q \alpha_{j,i} x_j + \sum_{j < k} \beta_{jk,i} x_j x_k + \sum_{j < k < l} \gamma_{jkl,i} x_j x_k x_l + \dots + \delta_{12\dots q} x_1 x_2 \dots x_q$$

dan vektor rata-rata bersyarat di sel  $m$   $\mu_i^{(m)}$ , diperoleh dengan cara mensubstitusikan nilai variabel-variabel biner di sel  $m$  ke ruas kanan persamaan di atas. Dalam prakteknya, sebuah aproksimasi diperoleh dengan memperlakukan semua suku order tinggi sebagai residual dan menduga parameter-parameter lainnya dengan metode regresi multivariate. Untuk model order kedua, model aditivnya menjadi

$$\mu_i = \mathbf{v}_i + \sum_{j=1}^q \alpha_{j,i} x_j + \sum_{j < k} \beta_{jk,i} x_j x_k$$

### 3. METODE PENELITIAN

Penelitian ini merupakan penelitian terapan statistika khususnya data mining pada bidang keuangan. Data yang digunakan merupakan data sekunder dari sebuah Bank swasta di Propinsi Lampung. Debitur yang dianalisis terdiri atas 1000 debitur yang tercatat pada tahun 2011. Variabel-variabel yang digunakan dalam penelitian ini disajikan pada Tabel 1.

Terdapat beberapa tahapan dalam analisis data penelitian ini. Tahap pertama adalah melakukan eksplorasi data, yaitu mengidentifikasi karakteristik-karakteristik debitur yang terpilih menjadi anggota sampel berdasarkan variabel-variabel yang diperhatikan.

Selanjutnya menyusun tabel kontingensi dengan mempertimbangkan variabel-variabel biner yang digunakan dalam penelitian. Kemudian mengidentifikasi sel-sel yang tidak ada amatannya, guna menetapkan penduga bagi parameter-parameter model. Setelah itu membagi data menjadi dua dengan komposisi 80% untuk data training dan 20% untuk data testing. Kemudian membangun model klasifikasi dengan menggunakan analisis

diskriminan dengan variabel bebas campuran biner dan kontinu. Langkah yang terakhir adalah melakukan validasi model.

**Tabel 1.** Variabel-Variabel dalam Penelitian

<b>Nama Variabel</b>	<b>Jenis Variabel</b>	<b>Satuan</b>	<b>Keterangan</b>
Usia	Kuantitatif	Tahun	-
Status Pernikahan	Kualitatif	-	1= menikah, 2 = tidak menikah /duda / janda
Jenis Kelamin	Kualitatif	-	1= laki-laki, 2 = perempuan
Masa Kerja	Kuantitatif	Tahun	-
Pendapatan Bersih	Kuantitatif	Rupiah	-
Status Kredit	Kualitatif	-	1= macet, 2 = lancar

#### 4. HASIL DAN PEMBAHASAN

Dalam penelitian ini, 1000 debitur di sebuah Bank di Kota Lampung dianalisis potensi macet dan lancarnya dalam pembayaran kredit berdasarkan variabel-variabel jenis kelamin, status pernikahan, usia, pendapatan bersih dan masa kerja. 800 debitur dikategorikan lancar dalam membayar kredit dan 200 yang lainnya dikategorikan macet.

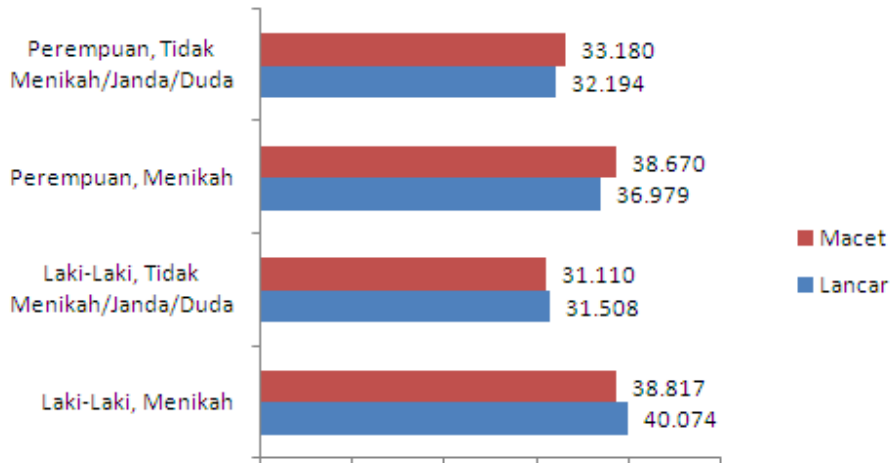
Variabel bebas biner yang digunakan dalam penelitian ini terdiri atas jenis kelamin dan status pernikahan, sehingga sel yang terbentuk sebanyak 4. Sel pertama adalah jenis kelamin laki-laki dan menikah, sel yang kedua jenis kelamin laki-laki dan tidak menikah/janda/duda, sel yang ketiga jenis kelamin perempuan dan menikah, dan sel yang keempat adalah jenis kelamin perempuan dan tidak menikah/janda/duda. Tabel 2 berikut ini adalah proporsi debitur menurut jenis kelamin, status pernikahan dan status kredit dari para debitur.

**Tabel 2.** Proporsi Debitur yang Lancar dan Macet Berdasarkan Jenis Kelamin dan Status Pernikahan

<b>Sel</b>	<b>Status Kredit</b>	
	<b>Lancar</b>	<b>Macet</b>
Laki-Laki, Menikah	0.516	0.115
Laki-Laki, Tidak Menikah/Janda/Duda	0.120	0.035
Perempuan, Menikah	0.097	0.033
Perempuan, Tidak Menikah/Janda/Duda	0.067	0.017

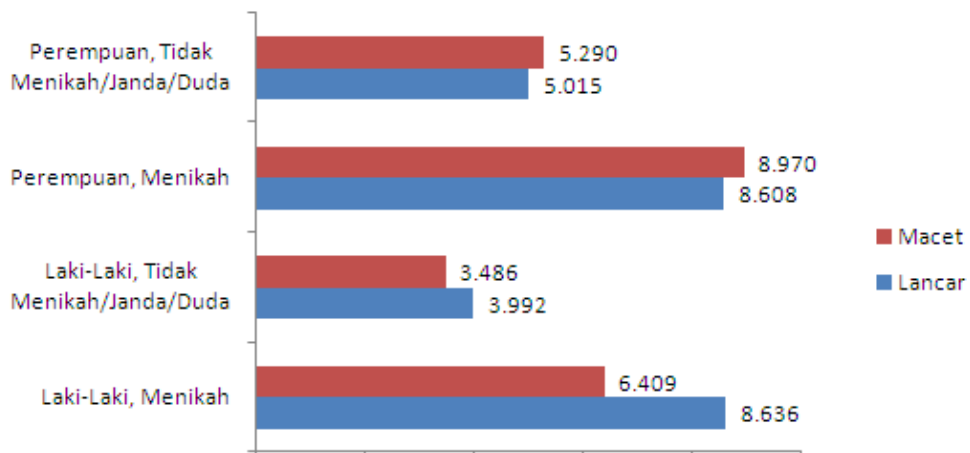
Dari Tabel 2, diketahui bahwa tabel kontingensi antara jenis kelamin dan status pernikahan menghasilkan sel-sel yang semuanya memiliki amatan, sehingga prosedur pendugaan peluang disetiap sel dan vektor rata-rata disetiap sel dapat dilakukan dengan mudah. Selanjutnya diketahui bahwa 80% dari data debitur yang digunakan dalam penelitian ini tercatat sebagai debitur yang lancar dalam pembayaran pinjamannya sedangkan 20% sisanya diketahui sebagai debitur yang macet. Disamping itu 76.1%

debitur-debitur tersebut berstatus menikah dan hanya 23.9% sisanya berstatus tidak menikah, duda maupun janda. Dari sisi jenis kelamin 78.6% dari keseluruhan debitur adalah laki-laki sedangkan sisanya berjenis kelamin perempuan.



**Gambar 1.** Grafik Rata-Rata Usia untuk Kombinasi antara Jenis Kelamin dan Status Pernikahan

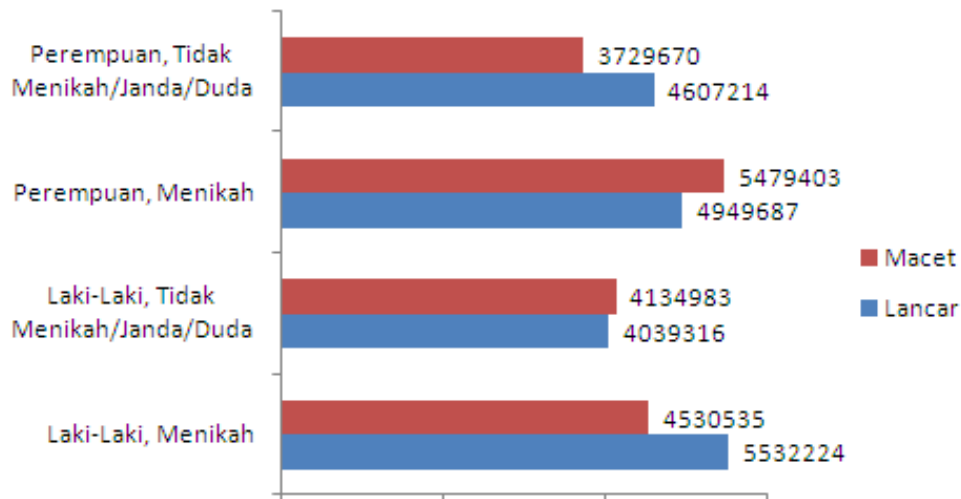
Dari Gambar 1, dapat diketahui bahwa debitur dengan jenis kelamin laki-laki dengan status kredit lancar cenderung memiliki usia yang lebih tinggi daripada debitur laki-laki dengan status kredit macet. Sebaliknya debitur yang berjenis kelamin perempuan dengan status kredit lancar cenderung memiliki usia yang lebih rendah daripada debitur perempuan yang status kreditnya macet. Namun demikian selisih usia mereka tidak berbeda nyata pada taraf signifikansi 5%.



**Gambar 2.** Grafik Rata-Rata Lama bekerja untuk Kombinasi antara Jenis Kelamin dan Status Pernikahan

Dari Gambar 2, dapat diketahui bahwa debitur yang berjenis kelamin laki-laki dengan status kredit lancar, pada saat pengajuan kredit cenderung usia kerjanya lebih lama

daripada debitur laki-laki dengan status kredit macet. Sebaliknya debitur yang berjenis kelamin perempuan dengan status kredit lancar, pada saat pengajuan kredit cenderung usia kerjanya lebih pendek daripada debitur perempuan yang status kreditnya macet. Namun demikian selisih usia mereka tidak berbeda nyata pada taraf signifikansi 5%, kecuali untuk kategori debitur laki-laki dan sudah menikah.



**Gambar 3.** Grafik Rata-Rata Penghasilan Bersih untuk Kombinasi antara Jenis Kelamin dan Status Pernikahan

Dari Grafik 3, diketahui bahwa debitur yang status kreditnya lancar dengan karakteristik jenis kelamin laki-laki dan menikah serta debitur berjenis kelamin perempuan dan tidak menikah/janda/duda cenderung memiliki penghasilan bersih yang lebih tinggi daripada mereka mereka yang macet. Sebaliknya debitur yang status kreditnya lancar dengan karakteristik jenis kelamin laki-laki dan tidak menikah/janda/duda serta debitur berjenis kelamin perempuan dan menikah cenderung memiliki penghasilan bersih yang lebih rendah daripada mereka mereka yang macet. Pada taraf nyata 5%, penghasilan bersih mereka berbeda nyata, kecuali untuk kategori debitur laki-laki dan tidak menikah/janda/duda. Pada kategori tersebut penghasilan bersih mereka tidak berbeda nyata.

Dalam rangka untuk membangun model diskriminan, pada bagian selanjutnya akan ditampilkan hasil-hasil yang diperoleh berdasarkan perhitungan dengan menggunakan persamaan (1) sampai dengan persamaan (7). Vektor rata-rata dan matrik varian kovarian sampel di setiap sel secara lengkap ada di Tabel 3 dan Tabel 4.



**Tabel 3.** Vektor Rata-Rata dan Matriks Varian Kovarian di Setiap Sel untuk Data Training pada Status Kredit Lancar

Sel	Vektor Rata-Rata	Matriks Varian Kovarian (S)
Laki-Laki, Menikah	$\begin{pmatrix} 40.125 \\ 8.506 \\ 5686301 \end{pmatrix}$	$\begin{pmatrix} 46 & 22 & 1.21E+07 \\ 22 & 49 & -3.56E+06 \\ 12078361 & -3564654 & 4.80E+13 \end{pmatrix}$
Laki-Laki, Tidak Menikah,/Janda/Duda,	$\begin{pmatrix} 31.13 \\ 3.898 \\ 3918849 \end{pmatrix}$	$\begin{pmatrix} 25 & 7 & 1.98E+06 \\ 7 & 24 & 1.66E+05 \\ 1975454 & 166152 & 7.67E+12 \end{pmatrix}$
Perempuan, Menikah	$\begin{pmatrix} 37.046 \\ 8.667 \\ 4537732 \end{pmatrix}$	$\begin{pmatrix} 48 & 32 & 3.04E+06 \\ 32 & 45 & -1.51E+06 \\ 3040135 & -1509271 & 1.01E+13 \end{pmatrix}$
Perempuan, Tidak Menikah/Janda/Duda, Lancar	$\begin{pmatrix} 32.095 \\ 5.111 \\ 4647789 \end{pmatrix}$	$\begin{pmatrix} 43 & 23 & 1.16E+07 \\ 23 & 26 & 7.74E+06 \\ 11632993 & 7741003 & 2.41E+13 \end{pmatrix}$

**Tabel 4.** Vektor Rata-Rata dan Matriks Varian Kovarian di Setiap Sel untuk Data Training pada Status Kredit Macet

Sel	Vektor Rata-Rata	Matriks Varian Kovarian (S)
Laki-Laki, Menikah, Macet	$\begin{pmatrix} 39.096 \\ 6.433 \\ 4634994 \end{pmatrix}$	$\begin{pmatrix} 71 & 22 & 2.79E+06 \\ 22 & 45 & -1.88E+06 \\ 2788265 & -1878583 & 2.95E+13 \end{pmatrix}$
Laki-Laki, Tidak Menikah, Macet	$\begin{pmatrix} 30.52 \\ 3.29 \\ 4268773 \end{pmatrix}$	$\begin{pmatrix} 64 & 17 & 1.46E+07 \\ 17 & 16 & 1.07E+06 \\ 14581217 & 1070630 & 8.12E+12 \end{pmatrix}$
Perempuan, Menikah, Macet	$\begin{pmatrix} 39.1 \\ 9.1 \\ 5437520 \end{pmatrix}$	$\begin{pmatrix} 50 & 42 & -1.20E+06 \\ 42 & 69 & -1.94E+07 \\ -1204044 & -1,9E+07 & 4.03E+13 \end{pmatrix}$
Perempuan, Tidak Menikah, Macet	$\begin{pmatrix} 33.79 \\ 5.79 \\ 4066323 \end{pmatrix}$	$\begin{pmatrix} 91 & 62 & 1.48E+07 \\ 62 & 66 & 7.03E+06 \\ 14768401 & 7031768 & 4.41E+12 \end{pmatrix}$

Dengan menggunakan informasi dari matrik varian kovarian di Tabel 3 dan Tabel 4, penduga kemungkinan maksimum (*maximum likelihood*) bagi  $\Sigma$  adalah

$$\mathbf{V} = \begin{pmatrix} 47.693 & 22.254 & 8.579 \times 10^6 \\ 22.254 & 43.324 & -2.155 \times 10^6 \\ 8.579 \times 10^6 & -2.155 \times 10^6 & 3.354 \times 10^{13} \end{pmatrix}$$

Untuk memperoleh model diskriminan di setiap sel, perlu dicari nilai peluang sebuah amatan yang berasal dari populasi  $i$  dan sel ke  $m$  yang disimbolkan dengan  $p_{im}$ . Hasil selengkapnya ada di Tabel 5.

**Tabel 5.** Nilai Dugaan terhadap  $p_{im}$  Berdasarkan Data *Training*

		Laki-Laki		Perempuan	
		Menikah	Tidak menikah/ Duda/Janda	Menikah	Tidak menikah/ Duda/Janda
Populasi	Lancar	0.516	0.120	0.097	0.070
	Macet	0.116	0.034	0.033	0.014

Berdasarkan teori pada bagian sebelumnya, model diskriminan di setiap sel seperti pada Tabel 6.

**Tabel 6.** Model Diskriminan di Setiap Sel Berdasarkan Data *Training*

Sel	Model Diskriminan Berdasarkan Data <i>Training</i>	Aturan Alokasi
Laki-Laki, Menikah	$M_1 = \left[ \begin{pmatrix} 40.125 \\ 8.506 \\ 5686301 \end{pmatrix} - \begin{pmatrix} 39.096 \\ 6.433 \\ 4634994 \end{pmatrix} \right]^t \mathbf{V}^{-1} \left[ \mathbf{y} - \frac{1}{2} \left\{ \begin{pmatrix} 40.125 \\ 8.506 \\ 5686301 \end{pmatrix} + \begin{pmatrix} 39.096 \\ 6.433 \\ 4634994 \end{pmatrix} \right\} \right]$	Alokasikan ke populasi lancar jika $M_1 \geq \frac{0.116}{0.516}$
Laki-Laki, Tidak Menikah/ Duda/Janda	$M_2 = \left[ \begin{pmatrix} 31.13 \\ 3.898 \\ 3918849 \end{pmatrix} - \begin{pmatrix} 30.52 \\ 3.29 \\ 4268773 \end{pmatrix} \right]^t \mathbf{V}^{-1} \left[ \mathbf{y} - \frac{1}{2} \left\{ \begin{pmatrix} 31.13 \\ 3.898 \\ 3918849 \end{pmatrix} + \begin{pmatrix} 30.52 \\ 3.29 \\ 4268773 \end{pmatrix} \right\} \right]$	Alokasikan ke populasi lancar jika $M_2 \geq \frac{0.034}{0.120}$
Perempuan, Menikah	$M_3 = \left[ \begin{pmatrix} 37.046 \\ 8.667 \\ 4537732 \end{pmatrix} - \begin{pmatrix} 39.1 \\ 9.1 \\ 5437520 \end{pmatrix} \right]^t \mathbf{V}^{-1} \left[ \mathbf{y} - \frac{1}{2} \left\{ \begin{pmatrix} 37.046 \\ 8.667 \\ 4537732 \end{pmatrix} + \begin{pmatrix} 39.1 \\ 9.1 \\ 5437520 \end{pmatrix} \right\} \right]$	Alokasikan ke populasi lancar jika $M_3 \geq \frac{0.033}{0.1097}$
Perempuan, Tidak Menikah/ Duda/Janda	$M_4 = \left[ \begin{pmatrix} 32.095 \\ 5.111 \\ 4647789 \end{pmatrix} - \begin{pmatrix} 33.79 \\ 5.79 \\ 4066323 \end{pmatrix} \right]^t \mathbf{V}^{-1} \left[ \mathbf{y} - \frac{1}{2} \left\{ \begin{pmatrix} 32.095 \\ 5.111 \\ 4647789 \end{pmatrix} + \begin{pmatrix} 33.79 \\ 5.79 \\ 4066323 \end{pmatrix} \right\} \right]$	Alokasikan ke populasi lancar jika $M_4 \geq \frac{0.014}{0.070}$

dengan  $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$  dimana  $y_1$  adalah usia debitur,  $y_2$  adalah lama bekerja debitur dan

$y_3$  adalah penghasilan bersih debitur.

Peluang kesalahan klasifikasi untuk data training dari model di atas dihitung dengan menggunakan persamaan (3) dan (4) dimana  $P(\text{macet}|\text{lancar}) = 0$  dan  $P(\text{lancar}|\text{macet}) = 0.1970$ . Sedangkan untuk data testing dihitung dengan mensubstitusikan amatan-amatan variabel kontinu ke dalam model-model diskriminan pada sel yang bersesuaian. Hasilnya pada sel ke-1 kesalahan klasifikasi sebesar 0.3810; pada sel ke-2 kesalahan klasifikasi sebesar 0.3125; pada sel ke-3 kesalahan klasifikasi sebesar 0.3077; dan pada sel ke-4 kesalahan klasifikasi sebesar 0.5000. Secara keseluruhan kesalahan

klasifikasi untuk data testing pada model diskriminan dengan variabel campuran ini sebesar 0.3753.

#### 4. KESIMPULAN

Ide membangun model diskriminan di setiap sel yang merupakan kombinasi dari kategori variabel-variabel biner menjadikan model-model diskriminan lebih fleksibel untuk digunakan. Model diskriminan dengan variabel bebas campuran telah memberikan kesempatan yang lebih luas bagi analis kredit untuk memanfaatkan berbagai macam variabel yang bisa digunakan untuk penilaian kredit. Model ini menjadi alternatif bagi para pengambil keputusan untuk meminimalkan resiko gagal bayar untuk sebuah kredit yang diberikan.

#### DAFTAR PUSTAKA

- Abdou, H.A., Pointon, J., dan Ahmed, E.-M., 2008, Neural Nets Versus Conventional Techniques in Credit Scoring in Egyptian Banking, *Expert Systems with Applications* 35, pp. 1275–1292.
- Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., dan Vanthienen, J., 2003, Benchmarking State-of-the-art Classification Algorithms for Credit Scoring., *Journal of the Operational Research Society*, 54(6), pp. 627–635.
- Brown, I. dan Mues, C., 2012, An Experimental Comparison of Classification Algorithms for Imbalanced Credit Scoring Data Sets, *Expert Systems with Applications*, 39, pp. 3446–3453.
- Chang, P. C. dan Afifi, A. A., 1974, Classification Based on Dichotomous and Continuous Variables, *Journal of The American Statistical Association*, 69, pp. 336-339.
- Crook, J. N., Edelman, D. B., dan Thomas, L. C. 2007. Recent Developments in Consumer Credit Risk Assessment, *European Journal of Operational Research*, 183(3), pp. 1447–1465.
- Desai, V.S., Crook, J.N., dan Overstreet, G.A., 1996, A Comparison of Neural Networks and Linear Scoring Models in The Credit Union Environment, *European Journal of Operational Research*, 95, pp. 24–37.
- Hand, D.J., 1981, *Discrimination and Classification*, Wiley, New York.
- Hsieh N. C., 2004, An Integrated Data Mining and Behavioral Scoring Model for Analyzing Bank Customers, *Expert Systems with Applications*, 27, pp. 623-633.
- Krzanowski, W. J., 1975, Discrimination and Classification Using Both Binary and Continuous Variables, *Journal of The American Statistical Association*, 70, pp. 782-790.
- Lee, T. S., Chiu, C.C., Chou, Y.C., dan Lu, C.J., 2006, Mining The Customer Credit Using Classification and Regression Tree and Multivariate Adaptive Regression Splines, *Computational Statistics and Data Analysis*, 50, pp. 1113–1130.
- Li X. L., dan Zhong Y., 2012, An Overview of Personal Credit Scoring: Techniques and Future Work, *International Journal of Intelligence Science*, 2, pp. 181-189.

- Nicolic N., Joksimovic N. Z., Stojanovski, D. dan Joksimovic I., 2013, The Application of Brute Force Logistic Regression to Corporate Credit Scoring Models: Evidence from Serbian Financial Statements. *Expert Systems with Applications*, 40, pp. 5932-5944.
- Olkin, I. dan Tate, R. F., 1961, Multivariate Correlation Models with Mixed Discrete and Continuous Variables, *Annals of Mathematical Statistics*, 32, pp. 448-465.
- The Hong Kong Institute of Bankers, 2012, *Credit Risk Management*. Wiley, Hong Kong.
- Utari, G. A. D., Arimurti, T., dan Kurniati, I. N., 2012, Pertumbuhan Kredit Optimal, *Buletin Ekonomi Moneter dan Perbankan*. 10, pp. 1-34.
- Yobas, M. B., Crook, J. N., dan Ross, P., 2000, Credit Scoring Using Neural and Evolutionary Techniques. *IMA Journal of Management Mathematics*, 11(2), pp. 111-125.