

# Analisis Klasifikasi Kredit Menggunakan Regresi Logistik Biner Dan *Radial Basis Function Network* di Bank 'X' Cabang Kediri

Sistya Rosi Diaprina dan Suhartono

Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,

Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

E-mail: suhartono@statistika.its.ac.id, sistyarosidiaprina@gmail.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. Penelitian tugas akhir ini bertujuan untuk menganalisis klasifikasi kredit guna mengurangi resiko terjadinya kredit macet di Bank X Cabang Kediri. Metode statistik yang digunakan adalah Regresi logistik Biner dan *Radial Basis Function Network*. Tahapan yang digunakan dalam penelitian ini adalah membagi data menjadi dua bagian yaitu data *training* dan data *testing*. Data *training* digunakan untuk pembentukan model, sedangkan data *testing* digunakan untuk menguji seberapa besar ketepatan model yang dibentuk. Hasil analisis menunjukkan bahwa rata-rata ketepatan klasifikasi dengan menggunakan metode Regresi Logistik Biner lebih besar dibandingkan dengan menggunakan metode *Radial Basis Function Network*. Sehingga dapat disimpulkan bahwa metode Regresi Logistik Biner memiliki ketepatan klasifikasi yang lebih baik dari pada *Radial Basis Function Network* untuk kasus klasifikasi kredit di Bank X Cabang Kediri.

**Kata Kunci**—*Credit Scoring Radial Basis Function Network*, Regresi Logistik Biner, Klasifikasi.

## I. PENDAHULUAN

**P**EMBERIAN kredit bagi masyarakat adalah salah satu tugas industri di sektor perbankan. Kemudahan layanan pemberian kredit bagi masyarakat tanpa proses penyeleksian yang baik mampu menyebabkan resiko kredit bagi perbankan. Resiko kredit terjadi akibat nasabah atau debitur yang tidak memenuhi kewajiban membayar kredit pada tanggal jatuh tempo yang telah ditentukan [1].

Dalam menghadapi masalah resiko kredit 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. Apabila pengklasifikasian resiko kredit mengalami kesalahan, maka salah satu dampak yang ditimbulkan adalah kredit macet. Kredit macet dapat menyebabkan kebangkrutan pada bank.

Dalam statistik, metode yang biasanya digunakan untuk memetakan nasabah disebut metode klasifikasi. Ada beberapa metode klasifikasi yang biasanya digunakan untuk memetakan nasabah kedalam kelompok mulai dari yang klasik sampai dengan modern. Pada metode klasik biasanya sering

digunakan metode regresi logistik dan analisis diskriminan, sedangkan untuk metode modern yang telah dikembangkan saat ini adalah *Neural Network*, *Genetic Algorithm* dan *Support Vector Machine*. Pengklasifikasian seperti ini sering disebut dengan istilah Data Mining atau *Knowledge Discovery in Database* (KDD). Keluaran dari Data Mining ini bisa digunakan untuk memperbaiki pengambilan keputusan di masa depan [2].

Salah satu penerapan Data Mining di dunia perbankan saat ini adalah *Credit Scoring*. *Credit Scoring* merupakan alat statistik yang digunakan oleh perusahaan perbankan serta lembaga keuangan lainnya, pemasaran dan perusahaan iklan untuk memperkirakan kemungkinan apakah pemohon pinjaman bisa dikategorikan sebagai potensi yang baik atau tidak. Beberapa model statistik yang dikembangkan untuk kepentingan *Credit Scoring* adalah Regresi Logistik dan *Multicriteria Decision* [3], *Discriminant Analysis* [4], *Neural Network* [5] dan *Support Vector Machine* [6]. Pengembangan metode NN untuk kasus *Credit Scoring* saat ini mulai pesat, seperti contohnya pada tahun 2006 telah dilakukan penelitian oleh Sarlija dkk. [7] dengan metode Regresi Logistik, *Survival Analysis* and *Neural Network*.

Bank X Cabang Kediri merupakan industri keuangan yang salah satu fungsinya adalah memberikan jasa pinjaman bagi nasabah. Setiap tahunnya bank ini mengalami peningkatan nasabah kredit. Untuk menghindari resiko kredit yang ditimbulkan oleh nasabah maka diperlukan analisis klasifikasi kredit guna mengurangi resiko kebangkrutan bank. Penelitian untuk klasifikasi kredit kali ini akan menggunakan Regresi Logistik Biner dan *Radial Basis Function Network*. Tujuan penelitian ini ingin membandingkan dari kedua metode tersebut, metode mana yang menghasilkan klasifikasi dengan akurasi lebih baik untuk kasus klasifikasi kredit pada Bank X Cabang Kediri.

## II. TINJAUAN PUSTAKA

### A. Regresi Logistik Biner

Regresi logistik biner adalah salah satu metode statistik yang biasanya digunakan untuk menjelaskan hubungan antara variabel respon ( $y$ ) yang bersifat biner atau dikotomis dengan satu atau lebih variabel prediktor ( $x$ ) yang bersifat metrik atau non-metrik [8]. Variabel respon  $y$  terdiri dari dua kategori yaitu “sukses” dan “gagal” yang dinotasikan dengan  $y = 1$

(sukses) dan  $y=1$  (gagal). Sehingga variabel  $y$  akan mengikuti distribusi Bernoulli untuk setiap observasi tunggal.

Fungsi probabilitas untuk setiap observasi adalah

$$f(y) = \pi^y (1 - \pi)^{1-y}; \text{ dengan } y = 0, 1 \quad (1)$$

dan  $\pi$  adalah probabilitas sukses. Jika  $y = 0$  maka

$$f(y) = (1 - \pi), \text{ dan jika } y = 1 \text{ maka } f(y) = \pi.$$

Fungsi regresi logistik dapat dituliskan sebagai berikut

$$f(z) = \frac{1}{1 + e^{-z}} \quad (2)$$

dengan

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \cdots + \beta_k x_k.$$

Nilai  $z$  antara  $-\infty$  dan  $+\infty$  sehingga nilai  $f(z)$  terletak antara 0 dan 1. Hal ini menunjukkan bahwa model logistik menggambarkan probabilitas atau resiko dari suatu objek. Secara umum, model regresi logistik ditulis dalam bentuk:

$$\pi(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 \cdots + \beta_k x_k)}} \quad (3)$$

Pendugaan parameter model regresi logistik dapat diuraikan dengan menggunakan transformasi logit dari  $\pi(x)$  yaitu

$$\ln\left(\frac{\pi(x)}{1 - \pi(x)}\right).$$

Karena

$$\frac{\pi(x)}{1 - \pi(x)} = e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 \cdots + \beta_k x_k)} \quad (4)$$

maka

$$\ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \cdots + \beta_k x_k. \quad (5)$$

Dasar dari fungsi likelihood yang akan digunakan sebagai estimasi dari nilai  $\beta$  yang dimaksimumkan adalah sebagai berikut:

$$L(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (6)$$

dengan

$n$  : banyaknya observasi

$x_i$  : nilai variabel independen untuk pengamatan ke- $i$

$y_i$  : nilai variabel respon untuk pengamatan ke- $i$

Secara matematika, rumus diatas lebih mudah dikerjakan dalam bentuk log yang ditulis sebagai berikut:

$$L(\beta) = \ln[L(\beta)] = \ln\left[\prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i}\right] \quad (7)$$

Untuk mendapatkan komponen dari  $\beta$  maka dilakukan differential pada log dari fungsi likelihood yang ditulis sebagai berikut [9].

$$\frac{\partial \ln L(\beta)}{\partial \beta_j} = 0, \quad j = 1, 2, \dots, k. \quad (8)$$

### B. Neural Network

Neural Network atau jaringan saraf tiruan ditemukan pertama kali pada tahun 1943 oleh McCulloch dan Pitts. Jaringan Saraf Tiruan (Neural Network) merupakan pemroses suatu informasi yang memiliki karakteristik mirip dengan jaringan saraf biologi. Jaringan saraf Tiruan dibentuk

sebagai generalisasi dari model matematika dan syaraf biologi. Ada beberapa tipe jaringan syaraf tiruan, namun hampir semuanya mempunyai komponen-komponen yang sama. Seperti halnya otak manusia, jaringan syaraf tiruan juga terdiri dari beberapa neuron dan terdapat hubungan antara neuron tersebut. Neuron tersebut akan mentransformasikan informasi yang diterima melalui sambungan keluarnya menuju ke neuron yang lain. Pada jaringan syaraf tiruan, hubungan ini disebut dengan nama bobot [10].

### C. Radial Basis Function Network

Pada dasarnya RBFN dibentuk dari 3 lapisan (layer) dengan peran yang berbeda. Input layer dibuat dari beberapa simpul (node) yang menghubungkan ke dalam jaringan. Lapisan kedua dinamakan hidden layer, yaitu lapisan yang berisi transformasi secara nonlinier dari input layer ke hidden layer. Lapisan selanjutnya disebut output layer yang bekerja secara linier yang menghasilkan respon yang berasal dari fungsi aktivasi pada hidden layer. Diketahui sekelompok data berjumlah  $N$  yang dinotasikan  $x_1, x_2, \dots, x_n$  yang dibagi ke dalam dua kelas yang berbeda. Misalkan  $x$  merupakan sebuah vektor pada ruang berdimensi  $m_0$ . Kemudian vektor  $\phi(x)$  akan memetakan titik pada ruang berdimensi baru yaitu  $m_1$ . Dimana  $\phi(x)$  adalah fungsi aktivasi yang memegang peranan penting pada hidden layer [11].

Radial Basis Function Network (RBFN) didesain untuk membentuk pemetaan secara non linear dari variabel input ke unit hidden layer dan pemetaan linear dari hidden layer ke output. Sehingga secara keseluruhan jaringan ini mempresentasikan sebuah pemetaan dari ruang dimensi  $m_0$  ke ruang dimensi tunggal yang ditulis sebagai berikut:

$$s : \mathcal{R}^{m_0} \rightarrow \mathcal{R}^1.$$

Dalam teori interpolasi jika diberikan  $N$  buah titik yang berbeda  $\{x_i \in \mathcal{R}^{m_0} | i = 1, 2, \dots, N\}$  dan memiliki hubungan dengan  $N$  buah bilangan sebenarnya  $\{d_i \in \mathcal{R}^1 | i = 1, 2, \dots, N\}$  maka didapatkan fungsi  $F : \mathcal{R}^N \rightarrow \mathcal{R}^1$  adalah fungsi yang memenuhi kondisi:

$$F(x) = d_i \quad i = 1, 2, \dots, N \quad (9)$$

Dalam teknik RBFN dilakukan pemilihan fungsi  $F$  sebagai berikut:

$$F(x) = \sum_{i=1}^N w_i \phi(\|x - x_i\|) \quad (10)$$

Dimana  $\{\phi(\|x - x_i\|) | i = 1, 2, \dots, N\}$  adalah himpunan fungsi nonlinier yang disebut fungsi radial basis dan  $\|\cdot\|$  adalah norm jarak Euclidean.

Dengan mensubstitusi persamaan pada (9) ke dalam persamaan (10) kita dapatkan bobot  $\{w_i\}$  sebagai berikut:

$$\begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1N} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{N1} & \phi_{N2} & \cdots & \phi_{NN} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix}$$

Dimana  $\phi_{ji} = \phi(\|x_j - x_i\|), (j, i) = 1, 2, \dots, N$

dengan

$$d = [d_1, d_2, \dots, d_N]^T$$

$$w = [w_1, w_2, \dots, w_N]^T$$

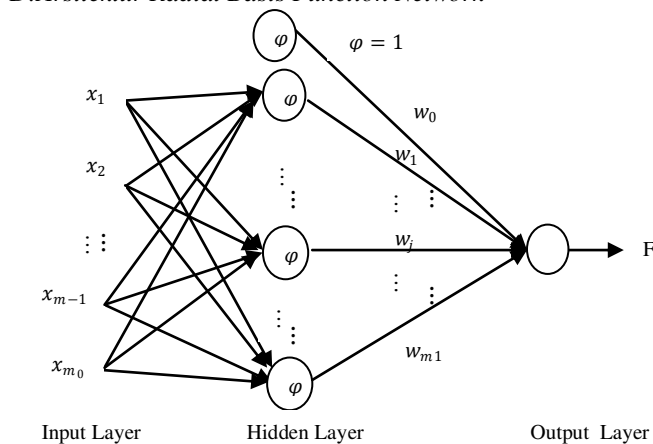
Fungsi radial basis yang digunakan adalah fungsi Gaussian. Menurut teori dari Micchelli tahun 1986 yang diambil pada buku [11] didapatkan fungsi *gaussian* sebagai berikut:

$$\phi(r) = \exp\left(-\frac{r^2}{2\sigma^2}\right) \quad (13)$$

Pada proses pemetaan secara linier dari *hidden layer* ke *output*, untuk mendapatkan nilai bobot  $w$  dilakukan dengan metode *ordinary least square*. Setelah mendapatkan matrik *Gaussian* yang dinotasikan dengan  $G$ , maka nilai bobot  $w$  dapat diketahui dari persamaan berikut:

$$w = (G^T G)^{-1} G^T d \quad (14)$$

#### D. Arsitektur Radial Basis Function Network



Gambar 1. Arsitektur Radial Basis Function Network

Dari gambar arsitektur jaringan *radial basis function* pada Gambar 1 memiliki penjelasan untuk setiap bagian sebagai berikut. Pada lapisan pertama atau yang disebut dengan *input layer* berisi titik (node) yang disusun dari variabel independen yang berjumlah  $m_0$ . Pada lapisan kedua yaitu *hidden layer* berisi unit non linier yang menghubungkan secara langsung semua ke titik (node) pada input layer.

#### E. Credit Scoring

*Credit Scoring* adalah metode numerik yang digunakan untuk menganalisa berdasarkan file kredit seseorang. Di dunia perbankan *Credit Scoring* seringkali digunakan. Bank menggunakan *Credit Scoring* untuk menentukan siapa yang memenuhi syarat untuk mendapatkan pinjaman, suku bunga, dan batas kredit.

Penerapan *Credit Scoring* berbeda di setiap negara. Di Amerika menggunakan beberapa metode seperti *Equifax* dan *Trans Union* yang juga diterapkan di negara Kanada. Selain itu, beberapa lembaga menggunakan metode *FICO* yang dikembangkan oleh *Fair Isaac Corporation*. Di Australia menggunakan metode *MARS*, *CART*, *CHAID* dan *Random Forest*. Sedangkan di Inggris menggunakan metode *Regresi Logistik* [12].

### III. METODOLOGI PENELITIAN

#### A. Sumber Data

Pada penelitian ini unit analisis yang diambil adalah nasabah. Jenis nasabah yang diteliti merupakan nasabah yang

diberikan kredit untuk kepentingan usaha. Penelitian ini dimaksudkan untuk memprediksi klasifikasi risiko kredit nasabah kredit Bank X Cabang Kediri melalui metode Regresi Logistik Biner dan *Radial Basis Function Network*. Data yang digunakan dalam penelitian ini adalah data sekunder nasabah kredit di Bank X Cabang Kediri per 31 Desember 2013 sebanyak 802 nasabah.

#### B. Variabel Penelitian

Variabel yang digunakan pada penelitian ini adalah variabel prediktor dan variabel respon pada data nasabah kredit di Bank X Cabang Kediri. Variabel respon pada data nasabah kredit Bank X Cabang Kediri bersifat biner yang terbagi dalam kategori 0 untuk kredit lancar dan 1 untuk kredit macet. Rincian variabel prediktor dan variabel respon dapat dilihat pada Tabel 1.

Tabel 1.  
Variabel Yang Digunakan Dalam Penelitian

No	Variabel	Keterangan	Type
1	Nama Cabang	$X_1$	Nominal
2	Produk	$X_2$	Nominal
3	Peruntukan Kredit	$X_3$	Nominal
4	Maksimum Kredit (juta)	$X_4$	Rasio
5	Tunggakan (juta)	$X_5$	Rasio
6	Suku Bunga (%)	$X_6$	Rasio
7	Jangka Waktu (tahun)	$X_7$	Rasio
8	Lama Kredit (tahun)	$X_8$	Rasio
9	Lama Usaha (tahun)	$X_9$	Rasio
10	Sektor	$X_{10}$	Nominal
11	Tipe Perusahaan	$X_{11}$	Nominal
12	Jenis Kredit	$Y$	Nominal

#### C. Metode Analisis

Metode Analisis yang digunakan pada penelitian ini adalah:

1. Mempersiapkan data kredit yang digunakan dalam metode ini dengan membagi data sebanyak 70% sebagai data *training* dan 30% sebagai data *testing*.
2. Melakukan *Pre-processing* data dengan menganalisis *missing value* terhadap data kredit di Bank X Cabang Kediri.
3. Melakukan analisis secara deskriptif untuk melihat karakteristik nasabah yang mengajukan aplikasi kredit di Bank X cabang Kediri.
4. Melakukan pemodelan klasifikasi menggunakan metode Regresi Logistik Biner, dan *Radial Basis Function Network*. Langkah analisis untuk Regresi Logistik Biner adalah sebagai berikut:
  - a. Mempersiapkan data training.
  - b. Meregresikan variabel respon dengan variabel prediktor secara individu.
  - c. Melakukan pengujian parameter secara parsial.
  - d. Meregresikan variabel respon dengan variabel prediktor secara serentak.
  - e. Melakukan pengujian parameter secara serentak.
  - f. Melakukan pemodelan regresi logistik biner dan membentuk model logitnya.

g. Membentuk tabel ketepatan klasifikasi.

h. Menghitung misklasifikasi dan ketepatan klasifikasi dari model yang dibentuk.

Langkah analisis untuk *Radial Basis Function Network* adalah sebagai berikut:

a. Mempersiapkan data training.

b. Menentukan nilai *centroid* dan *width* dengan metode cluster untuk tiap kelompok data.

c. Menyusun matrik Gaussian.

d. Menghitung bobot dengan metode *Ordinary Least Square*.

e. Membentuk tabel ketepatan klasifikasi.

f. Menghitung misklasifikasi dan ketepatan klasifikasi dari model yang dibentuk.

5. Melakukan pemilihan model terbaik.

6. Menarik Kesimpulan.

#### IV. ANALISIS DAN PEMBAHASAN

##### A. Pemodelan Jenis Kredit Menggunakan Metode Regresi Logistik Biner Dan Radial Basis Function Network

Sebelum dilakukan analisis menggunakan regresi logistik biner dan RBFN, tahapan pertama adalah membagi data menjadi dua kelompok yaitu data *training* dan data *testing*. Adapun pembagian data *training* dan *testing* adalah sebagai berikut:

$$N_{\text{training}} = \frac{70}{100} \times 802 \approx 561$$

$$N_{\text{testing}} = \frac{30}{100} \times 802 \approx 241$$

##### B. Pembentukan Regresi Logistik Biner Secara Individu

Berdasarkan model Regresi Logistik Biner secara individu apabila digunakan tingkat signifikansi  $\alpha$  sebesar 20%, didapatkan tiga variabel prediktor yang mempunyai pengaruh signifikan terhadap variabel respon yaitu variabel tunggakan, sektor dan tipe perusahaan. Hal ini menjelaskan bahwa variabel jumlah tunggakan, sektor dan tipe perusahaan secara individu berpengaruh signifikan terhadap variabel jenis kredit usaha di Bank X Cabang Kediri.

##### C. Pembentukan Regresi Logistik Biner Secara Serentak

Pengujian parameter secara serentak bertujuan untuk mengetahui apakah variabel prediktor berpengaruh signifikan yang secara serentak dengan menggunakan uji *likelihood* atau *likelihood ratio test*.

Tabel 2.  
Model Regresi Logistik Biner Secara Serentak

Variabel	Estimasi Parameter ( $\beta$ )	SE	Nilai Wald	P-value	Exp (B)
X1			0,001	1,000	
X1(1)	-30,944	1080	0,001	0,977	0,000
X1(2)	-32,437	2728	0,000	0,991	0,000
X5	1,054	0,553	3,631	0,057	2,868
Konstanta	-5,427	1,025	28,01	0,000	0,004

Berdasarkan Tabel 2 yaitu model regresi logistik biner secara serentak dengan memasukkan kesebelas variable

didapatkan variabel yang signifikan adalah  $X_5$  yaitu tunggakan. Variabel  $X_1$  atau nama cabang masuk ke dalam model karena menurunkan nilai misklasifikasi. Sehingga variabel  $X_1$  di masukan ke dalam model karena memberikan pengaruh terhadap nilai misklasifikasi dan ketepatan akurasi.

Model peluang regresi logistik yang diperoleh berdasarkan tabel 2 di atas adalah:

$$\hat{\pi}(x) = \frac{1}{1 + \exp(-5,427 - 30,944 X_{1(1)} - 32,437 X_{1(2)} + 1,054 X_5)}$$

Dengan menggunakan model peluang tersebut maka dapat dihitung berapa peluang seseorang dikategorikan kredit macet dan kredit lancar sesuai dengan variabel yang digunakan. Model peluang diatas adalah model yang non linier sehingga agar model menjadi linier dilakukan transformasi logit. Adapun transformasi logit sebagai berikut:

$$\hat{g}(x) = -5,427 - 30,944 X_{1(1)} - 32,437 X_{1(2)} + 1,054 X_5$$

##### D. Ketepatan Klasifikasi

Ketepatan klasifikasi pada data training pada penelitian ini adalah sebagai berikut:

Tabel 4.  
Tabel klasifikasi untuk model regresi logistik pada data training

Kategori	Prediksi Y		Prosentase Benar
	Kredit lancar	Kredit macet	
Y Kredit Lancar	544	0	100
Kredit Macet	1	16	94,1
Total	545	16	99,8

Berdasarkan Tabel 4 diatas didapatkan penjelasan bahwa jumlah nasabah yang diklasifikasikan ke dalam kategori kredit lancar secara benar adalah sebanyak 544 nasabah. Sedangkan jumlah nasabah yang diklasifikasikan ke dalam kategori kredit macet secara benar adalah sebanyak 16 nasabah. Sisanya sebanyak 1 kategori kredit macet yang diklasifikasikan sebagai kredit lancar. Dengan total ketepatan klasifikasi yang diberikan pada data training sebesar 99,8%. Selain prosentase ketepatan klasifikasi dari tabel klasifikasi diatas dapat dihitung nilai misklasifikasi sebagai berikut:

$$\text{Misklasifikasi} = \frac{0 + 1}{544 + 16} = 0,00178$$

Ketepatan klasifikasi pada data testing pada penelitian ini adalah sebagai berikut:

Tabel 5.  
Tabel klasifikasi untuk model regresi logistik pada data testing

Kategori	Prediksi Y		Prosentase Benar
	Kredit lancar	Kredit macet	
Y Kredit lancar	240	0	100
Kredit macet	0	1	100
Total	240	1	100

Berdasarkan tabel 5 diatas didapatkan penjelasan bahwa jumlah nasabah yang diklasifikasikan ke dalam kategori kredit lancar secara benar adalah sebanyak 240 nasabah. Sedangkan jumlah nasabah yang diklasifikasikan ke dalam kategori kredit macet secara benar adalah sebanyak 1 nasabah. Dengan total

ketepatan klasifikasi yang diberikan pada data testing adalah sebesar 100%.

#### E. Pembentukan Model Jaringan Radial Basis (RBFN)

Pada model *Radial Basis Function Network* terdiri dari 3 lapisan yaitu *input layer*, *hidden layer* dan *output layer*. Penentuan banyaknya *neuron* yang digunakan pada *hidden layer* didasarkan pada nilai misklasifikasi yang terkecil. Adapun pemilihan jumlah *neuron* diawali dengan memasukkan input sebanyak 2 yaitu  $X_1$  (nama cabang) dan  $X_5$  (tunggakan). Kemudian melakukan analisis RBFN dari jumlah neuron paling kecil sampai jumlah paling besar. Dalam kasus ini jumlah neuron hanya dibatasi sampai dengan 10 saja. Adapun nilai misklasifikasi yang dihasilkan pada masing-masing *neuron* adalah sebagai berikut:

Tabel 6.  
Tabel Misklasifikasi Untuk Masing-masing *Neuron*

Neuron	Misklasifikasi	
	Training	Testing
1	3%	0,4%
2	3%	0,4%
3	2,5%	1,2%
4	1,4%	1,2%
5	1,2%	0,8%
6	1,2%	0,4%
7	1,1%	0,4%
8	0,5%	0,4%
9	0,4%	0%
10	0,9%	0%

Berdasarkan Tabel 6 diatas nilai misklasifikasi yang terendah yaitu untuk jumlah neuron 9. Sehingga dalam pemodelan dengan metode *Radial Basis Function Network* kali ini menggunakan neuron sebanyak 9. Dengan memasukkan neuron sebanyak 9, didapatkan nilai centroid dan standar deviasi untuk masing-masing variabel prediktor pada setiap hidden unit sebagai berikut:

Tabel 7.  
Nilai Centroid dan Standar Deviasi

Hidden Layer	Centroid				Standar Deviasi
	[X1=0]	[X1=1]	[X1=2]	X5	
H <sub>1</sub>	0	1	0	-0,092	0,003
H <sub>2</sub>	0	1	0	0,524	-0,040
H <sub>3</sub>	0	0	1	-0,089	0,009
H <sub>4</sub>	1	0	0	-0,092	0,001
H <sub>5</sub>	1	0	0	-0,0002	0,037
H <sub>6</sub>	1	0	0	0,466	0,039
H <sub>7</sub>	0	1	0	0,305	0,001
H <sub>8</sub>	0	0	1	12,368	0,001
H <sub>9</sub>	1	0	0	9,758	0,001

Berdasarkan Tabel 7 didapatkan nilai *centroid* dan standar deviasi yang digunakan dalam perhitungan pada fungsi Gaussian. Misalkan diketahui data ke 802, nasabah kredit bank X berasal dari SKC Kediri ( $X_1=0$ ) dan memiliki tunggakan sebesar 122 juta ( $X_5=122$ ) maka fungsi *Hidden layer* yang dibentuk adalah sebagai berikut:

$$H_1 = \exp \left[ -\frac{1}{2} \left( \left( \frac{0-0}{0,003} \right)^2 + \left( \frac{122+0,092}{0,003} \right)^2 \right) \right]$$

$$H_2 = \exp \left[ -\frac{1}{2} \left( \left( \frac{0-0}{-0,040} \right)^2 + \left( \frac{122-0,524}{-0,040} \right)^2 \right) \right]$$

$$\vdots$$

$$H_9 = \exp \left[ -\frac{1}{2} \left( \left( \frac{0-1}{0,001} \right)^2 + \left( \frac{122-9,758}{0,001} \right)^2 \right) \right]$$

Setelah didapatkan perhitungan fungsi Gaussian maka langkah selanjutnya adalah mendapatkan nilai bobot dengan menggunakan metode *ordinary least square*. Nilai bobot yang dihasilkan adalah sebagai berikut:

Tabel 8.  
Bobot dari *hidden layer* ke *output layer*

Hidden Layer	Output Layer	
	Y=0	Y=1
konstanta	0,933	0,067
H <sub>1</sub>	0,067	-0,067
H <sub>2</sub>	-1,906	1,906
H <sub>3</sub>	0,071	-0,071
H <sub>4</sub>	0,164	-0,164
H <sub>5</sub>	-2,147	2,147
H <sub>6</sub>	-1,635	1,635
H <sub>7</sub>	-0,933	0,933
H <sub>8</sub>	-0,933	0,933
H <sub>9</sub>	-0,933	0,933

Berdasarkan Tabel 8 didapatkan hasil prediksi yang dibentuk berdasarkan persamaan 2.1 adalah sebagai berikut:

$$\hat{y}_0 = 0,933 + 0,067H_1 - 1,906H_2 + 0,071H_3 + 0,164H_4 - 2,147H_5 - 1,635H_6 - 0,933H_7 - 0,933H_8 - 0,933H_9$$

$$\hat{y}_1 = 0,067 - 0,067H_1 + 1,906H_2 - 0,071H_3 - 0,164H_4 + 2,147H_5 + 1,635H_6 + 0,933H_7 + 0,933H_8 + 0,933H_9$$

#### F. Ketepatan Klasifikasi Dengan Metode RBFN

Sama halnya dengan metode regresi logistik, untuk melihat akurasi dari metode RBFN dapat dilihat dari tabel ketepatan klasifikasinya.

Tabel 9.  
Tabel Klasifikasi Untuk Model RBFN Pada Data Training

Kategori		Prediksi		Prosentase Benar
		Y		
		Kredit lancar	Kredit macet	
Y	Kredit Lancar	544	0	100
	Kredit Macet	2	15	88,2
	Total	546	15	99,6

Berdasarkan Tabel 9 diatas didapatkan penjelasan bahwa jumlah nasabah yang diklasifikasikan ke dalam kategori kredit lancar secara benar adalah sebanyak 544 nasabah. Sedangkan jumlah nasabah yang diklasifikasikan ke dalam kategori kredit macet secara benar adalah sebanyak 15 nasabah. Sisanya

sebanyak 2 kategori kredit macet yang diklasifikasikan sebagai kredit lancar. Dengan total ketepatan klasifikasi yang diberikan pada data training sebesar 99,8%. Selain prosentase ketepatan klasifikasi dari tabel klasifikasi diatas dapat dihitung nilai misklasifikasi sebagai berikut:

$$\text{Misklasifikasi} = \frac{0 + 2}{544 + 15} = 0,004$$

Tabel 10.  
Tabel Klasifikasi Untuk Model RBFN Pada Data Testing

Kategori		Prediksi		Prosentase Benar
		Y		
		Kredit lancar	Kredit macet	
Y	Kredit lancar	240	0	100
	Kredit macet	0	1	100
	Total	240	1	100

Berdasarkan Tabel 10 diatas didapatkan penjelasan bahwa jumlah nasabah yang diklasifikasikan ke dalam kategori kredit lancar secara benar adalah sebanyak 240 nasabah. Sedangkan jumlah nasabah yang diklasifikasikan ke dalam kategori kredit macet secara benar adalah sebanyak 1 nasabah. Dengan total ketepatan klasifikasi yang diberikan pada data testing adalah sebesar 100%.

#### H. Pemilihan Model Terbaik

Pemilihan model terbaik pada penelitian ini menggunakan data *testing*. Dilakukan percobaan sebanyak 10 kali untuk masing-masing jumlah kredit macet yang berbeda. Untuk setiap jumlah kredit macet diambil secara random sebanyak jumlah data testing. Pada penelitian ini dilakukan sebanyak dua komposisi yaitu jumlah kredit macet 6 dengan kredit lancar berjumlah 235 dan jumlah kredit macet 8 dengan kredit lancar sebanyak 233. Didapatkan hasil ketepatan klasifikasi sebagai berikut:

Tabel 11.  
Ketepatan Klasifikasi Untuk Masing-masing Komposisi n(y=1) yang berbeda

No	[n(y=1)]=6		[n(y=1)]=8	
	Regresi logistik biner	RBFN	Regresi logistik biner	RBFN
1	97,9	98,8	99,6	98,8
2	99,1	98,8	100	99,2
3	100	99,2	100	98,3
4	99,6	98,8	99,6	97,9
5	99,2	99,2	99,6	98,8
6	99,6	99,6	99,6	98,3
7	99,2	97,9	99,6	98,8
8	99,6	98,8	100	98,3
9	100	99,6	100	98,8
10	99,6	99,6	99,6	97,9
Rata-rata	99,38	99,03	99,76	98,51

Berdasarkan Tabel 11 didapatkan rata-rata ketepatan klasifikasi dengan menggunakan metode regresi logistik biner adalah sebesar 99,38% untuk data dengan jumlah kredit macet 6 dan 99,76% untuk data dengan jumlah kredit macet 8. Rata-rata ketepatan klasifikasi untuk metode *Radial Basis*

*Function Network* adalah sebesar 99,03% untuk data dengan jumlah kredit macet 6 dan 98,51% untuk data dengan komposisi jumlah kredit macet 8. Pada data dengan jumlah kredit macet 6 didapatkan hasil metode regresi logistik biner lebih baik dibandingkan dengan metode *radial basis function network* sebanyak 6 kali, tidak lebih baik 1 kali dan memiliki nilai yang sama sebanyak 3 kali. Pada data dengan jumlah kredit macet 8 didapatkan metode regresi logistik biner lebih baik sebanyak 10 kali dibandingkan dengan *radial basis function network*.

## V. KESIMPULAN DAN SARAN

Pada tahapan pemilihan model terbaik didapatkan nilai rata-rata ketepatan klasifikasi dengan menggunakan metode regresi logistik biner adalah sebesar 99,38% untuk data dengan jumlah kredit macet sebanyak 6 dan 99,76% untuk data dengan jumlah kredit macet sebanyak 8. Rata-rata ketepatan klasifikasi untuk metode *Radial Basis Function Network* adalah sebesar 99,03% untuk data dengan jumlah kredit macet sebanyak 6 dan 98,51% untuk data dengan jumlah kredit macet sebanyak 8. Sehingga dapat disimpulkan bahwa metode yang lebih baik digunakan untuk kasus klasifikasi adalah dengan menggunakan regresi logistik biner karena memiliki nilai ketepatan klasifikasi yang lebih besar dibandingkan dengan menggunakan *radial basis function network*.

Berdasarkan hasil analisis yang telah dilakukan untuk keberlanjutan penelitian selanjutnya, saran yang diberikan adalah perlu memperhatikan aspek dari keseimbangan data.

## DAFTAR PUSTAKA

- [1] L. Misdiati, "Analisis Klasifikasi Kredit Menggunakan Metode Newton Truncated-Kernel Logistic Regression (NTR-KLR), (Studi kasus : Data Kredit Bank "X")." Tugas Akhir Mahasiswa Jurusan Statistika, FMIPA, ITS, Surabaya (2013).
- [2] B. Santosa, "Data Mining, Teknik Pemanfaatan Data Untuk Keperluan Bisnis" Graha Ilmu (2007).
- [3] N. Sarlija, K. Soric, S. Vlah, & V.V. Rosenzweig, "Logistic Regression and Multicriteria Decision Making in Credit Scoring," *Proceedings of the 10th International Symposium on Operational Research SOR '09*, (2006).
- [4] G. Mircea, M. Pirtea, M. Neamtu, and S. Bazavan, "Discriminant Analysis in Credit Scoring Model", Paper of Faculty of Economics and Business Administration West University of Timisoara, Romania (2011).
- [5] M. B. Yobas, J. N. Crook, and P. Ross, "Credit Scoring Using Neural and Evolutionary Techniques," *IMA Journal of Mathematics Applied in Business and Industry*, Vol. 11 (2000) 111-125.
- [6] C. L. Huang, M. C. Chen, and C. J. Wang, "Credit scoring with a data mining approach based on support vector machines," *Expert System with Application*, Vol. 33 (2007) 847-856.
- [7] N. Sarlija, M. Bencic, and M. Z. Susac, (2006). "Modeling Customer Revolving Credit Scoring Using Logistic Regression, Survival Analysis and Neural Networks," *Proceedings of the 7th WSEAS International Conference on Neural Networks*, (pp 164-169). Cavtat: Croatia (2006).
- [8] D. W. Hosmer, and S. Lemeshow, *Applied Logistic Regression*. New York: Inc, John Wiley and Sons. (2000).
- [9] D. G. Kleinbaum, and M. Klein, *Logistic Regression: A Self-Learning Text*. London : Springer Science + Business Media LLC (2010).
- [10] S. Kusumadewi, *Membangun Jaringan Syaraf Tiruan Menggunakan Matlab & Excel Link*. Graha Ilmu (2004).
- [11] S. Haykin, *Neural Network: A Comprehensive Foundation*. Inc, Prentice-Hall (1999).
- [12] H. Yapputro, *Banking Solution: A Collection of Articles authored by Hendrix Yapputro*. Jakarta: HSEY Publishing (2013).