

Pemodelan Faktor-Faktor yang Mempengaruhi Penggolongan Kredit di PT Bank X (Persero) Tbk dengan Menggunakan Metode *Hybrid Genetic Algorithm - Logistic Regression*

Ni Putu Budi Setianingsih dan Irhamah
Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Institut Teknologi Sepuluh Nopember (ITS)
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia
e-mail: irhamah@statistika.its.ac.id

Abstrak—PT Bank X (Persero) Tbk menunjukkan kinerja baik dalam perkreditan sampai pada tahun 2010. Namun sampai pada akhir kuartal III tahun 2013, PT Bank X (Persero) Tbk menjadi salah satu bank persero di Indonesia yang mengalami peningkatan rasio kredit bermasalah atau *non performing loan* (NPL). Terjadinya kredit bermasalah akan memberi dampak bagi kreditur maupun debitur. Untuk itu, penelitian ini melakukan pemodelan faktor-faktor yang mempengaruhi penggolongan kredit di PT Bank X (Persero) Tbk guna memprediksi risiko kredit dari calon debitur. Pemodelan tersebut dilakukan dengan menggunakan metode regresi logistik dan *hybrid genetic algorithm - logistic regression* terhadap data debitur di PT Bank X (Persero) Tbk. Fungsi *fitness* yang di-gunakan adalah ukuran kesesuaian model regresi logistik, yaitu *pseudo R²* atau R_L^2 dan MSE. Metode *hybrid genetic algorithm - logistic regression* memberikan hasil yang lebih baik karena R_L^2 yang bernilai lebih tinggi dan MSE yang bernilai lebih rendah dibandingkan dengan hasil estimasi parameter model regresi logistik menggunakan metode *maximum likelihood estimation* (MLE).

Kata Kunci—Algoritma genetika, MSE, penggolongan kredit, *pseudo R²*, regresi logistik.

I. PENDAHULUAN

MASYARAKAT Indonesia telah mengenal kredit secara luas, yang ditunjukkan oleh semakin banyaknya Bank Perkreditan Rakyat (BPR) dan Bank Syariah [1]. Statistik Perbankan Indonesia (SPI) mencatat bahwa bank persero di Indonesia telah menyalurkan kredit yang bernilai lebih dari seribu triliun rupiah sampai pada November 2013 untuk berbagai sektor lapangan usaha dan bukan lapangan usaha. PT Bank X (Persero) Tbk adalah satu di antara beberapa bank persero penyalur kredit. Dalam menyalurkan kredit, PT Bank X (Persero) Tbk menghasilkan profil yang baik hingga tahun 2010, dengan turunnya jumlah kasus kredit bermasalah secara signifikan [2]. Hal itu ditunjukkan oleh rasio kredit bermasalah atau rasio *Non Performing Loan* (NPL) yang turun dari 15,34% pada tahun 2005 menjadi 0,62% pada tahun 2010. Demikian pula yang berjalan hingga tahun 2013, yang mana penyaluran kredit PT Bank X (Persero) Tbk meningkat di seluruh segmen bisnis.

Namun pemberian kredit memiliki risiko pada terjadinya kredit bermasalah. SPI mencatat bahwa NPL dari bank persero di Indonesia pada November 2013 meningkat dari bulan dan

tahun sebelumnya. NPL dari bank persero di Indonesia mencapai sekitar 23 triliun rupiah, dengan PT Bank X (Persero) Tbk juga menjadi salah satu bank persero yang mengalami kenaikan NPL hingga akhir kuartal III tahun 2013 [3].

Dampak yang ditimbulkan kredit bermasalah akan berlanjut pada pertumbuhan ekonomi nasional. Untuk itu, kreditur melakukan beberapa upaya untuk bisa mengurangi risiko kredit bermasalah. Salah satu upaya tersebut adalah meminimalkan terjadinya kredit bermasalah dengan memprediksi risiko kredit dari calon debitur secara tepat. Ada penelitian sebelumnya [4] mengenai faktor-faktor yang mempengaruhi penggolongan kredit baik dan kredit buruk dengan *Newton Truncated - Kernel Logistic Regression* (NTR-KLR), namun dengan sensitivitas yang bernilai rendah. Berdasarkan penelitian tersebut, dalam penelitian ini akan dilakukan pemodelan faktor-faktor yang mempengaruhi penggolongan kredit pada data debitur PT Bank X (Persero) Tbk dengan metode lain, yaitu dengan *Hybrid Genetic Algorithm - Logistic Regression*. Variabel dependen dalam penelitian ini adalah penggolongan kredit debitur di PT Bank X (Persero) Tbk yang terdiri dari dua kategori, yaitu kredit baik dan kredit buruk, sehingga pemodelan faktor-faktor yang mempengaruhi penggolongan kredit di PT Bank (X) Persero Tbk dilakukan dengan regresi logistik.

Hybrid terhadap metode lain dapat dilakukan untuk meningkatkan ukuran kesesuaian model. Salah satu metode tersebut adalah algoritma genetika atau *Genetic Algorithm* (GA). Algoritma genetika, yaitu bagian dari *Evolutionary Algorithm* (EA) untuk membantu optimasi fungsi dengan mensimulasikan evolusi alam ke dalam suatu populasi yang berisi solusi-solusi untuk fungsi tersebut [5]. Algoritma genetika juga dapat diaplikasikan dengan mudah pada metode yang telah ada [6]. Dalam penelitian ini, *hybrid* algoritma genetika dilakukan pada regresi logistik biner. Terdapat beberapa kelebihan algoritma genetika untuk *hybrid*, di antaranya adalah pencarian dari populasi solusi yang luas, bukan berupa solusi tunggal [7]. Selain itu, algoritma genetika juga mampu memperoleh solusi yang global optimum, bukan hanya lokal optimum [8]. Dengan melakukan *hybrid* algoritma genetika pada metode regresi logistik biner, algoritma genetika membantu estimasi parameter untuk faktor-faktor yang berpengaruh terhadap penggolongan kredit pada data debitur di PT Bank X (Persero) Tbk berdasarkan nilai *pseudo R²* atau R_L^2 dan MSE.

II. TINJAUAN PUSTAKA

A. Uji Independensi

Salah satu uji independensi adalah uji independensi *Chi-square*, dengan hipotesis berikut.

H_0 : Kedua variabel adalah independen

H_1 : Kedua variabel adalah dependen.

Untuk menguji independensi dari tabel kontingensi berukuran $I \times J$, berikut adalah statistik Pearson pada persamaan (1) dan *likelihood-ratio* pada persamaan (2).

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} \quad (1)$$

$$G^2 = 2 \sum_{i=1}^I \sum_{j=1}^J n_{ij} \log \left(\frac{n_{ij}}{\hat{\mu}_{ij}} \right) \quad (2)$$

Notasi n_{ij} menyatakan banyaknya data dengan kategori A_i yang termasuk dalam kategori B_j dan $\hat{\mu}_{ij}$ adalah estimasi nilai harapan dengan perumusan $\hat{\mu}_{ij} = \frac{n_{i+}n_{+j}}{n}$. Kedua statistik uji tersebut mengikuti distribusi χ^2 dengan derajat bebas $df = (I-1)(J-1)$.

B. Regresi Logistik (*Logistic Regression*)

Tujuan dalam regresi logistik adalah untuk melakukan identifikasi variabel independen yang mempengaruhi pengelompokan variabel dependen dan membuat sistem klasifikasi yang berbasis pada model logistik untuk pengelompokan [9]. Berikut adalah model regresi logistik.

$$\pi(\mathbf{x}) = \frac{e^{g(\mathbf{x})}}{1 + e^{g(\mathbf{x})}} \quad (3)$$

Transformasi dari $\pi(\mathbf{x})$ adalah transformasi *logit*. Untuk variabel independen sebanyak p dengan vektor $\mathbf{x}' = (x_1, x_2, \dots, x_p)$, didefinisikan *logit* sebagai berikut.

$$g(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (4)$$

Metode yang pada umumnya digunakan dalam mengestimasi koefisien parameter model regresi logistik adalah *maximum likelihood estimation* (MLE). Untuk menguji pengaruh setiap β_j secara individu, berikut adalah hipotesis yang digunakan.

H_0 : $\beta_j = 0$

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

Statistik uji Wald adalah sebagai berikut.

$$W_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (5)$$

Untuk statistik uji pada persamaan (5), daerah penolakan dalam pengujian adalah H_0 ditolak jika $|W_j| > Z_{\alpha/2}$.

Sedangkan interpretasi untuk variabel independen dalam model regresi logistik adalah berbeda, yang bergantung pada jenis variabel berupa metrik atau non-metrik. Untuk parameter yang berupa variabel non-metrik, berlaku *odds ratio* $OR = e^\beta$, untuk setiap variabel independen. *Odds ratio* adalah ukuran asosiasi yang memperkirakan kecenderungan variabel respon akan sama dengan $x = 1$ daripada $x = 0$. Sedangkan interpretasi untuk variabel independen berupa variabel kontinyu adalah bergantung pada unit dari setiap variabel.

Kesesuaian model menunjukkan bagaimana model yang terbentuk dapat mendeskripsikan variabel respon [10]. Salah satu ukuran kesesuaian model regresi logistik adalah *pseudo R²* yang serupa dengan R^2 pada analisis regresi. Jika L_0 adalah notasi untuk *log-likelihood* model yang hanya mengandung *intercept*, dan L_p adalah notasi untuk *log-likelihood* model lengkap yang mengandung *intercept* dan p *covariate*, maka *pseudo R²* atau R_L^2 dapat dirumuskan sebagai berikut.

$$R_L^2 = \frac{L_0 - L_p}{L_0} = 1 - \frac{L_p}{L_0} \quad (6)$$

R_L^2 memiliki rentang nilai dari 0 sampai dengan 1. Selain R_L^2 , tabel klasifikasi dapat menjadi cara lain dalam mengevaluasi model regresi logistik, namun tabel klasifikasi paling tepat digunakan ketika analisis bertujuan untuk klasifikasi.

C. Algoritma Genetika (*Genetic Algorithm*)

John Holland memperkenalkan algoritma genetika atau *Genetic Algorithm* (GA) pada awal tahun 1970. Algoritma ini dimulai dengan sebuah populasi awal yang beranggotakan sekumpulan individu atau kromosom [7]. Kromosom terdiri dari n gen dengan alel (*allele*) sebagai nilai untuk setiap gen dan dapat direpresentasikan dengan bilangan riil, bilangan bulat, alfabet, maupun beberapa simbol. Setiap kromosom dibangkitkan secara acak dan akan melalui iterasi yang disebut dengan generasi. Pada setiap generasi, setiap kromosom akan dievaluasi dengan menggunakan ukuran *fitness*, kemudian terbentuk populasi untuk generasi berikutnya melalui beberapa operator genetika. Prosedur ini terus berlangsung hingga mencapai suatu kondisi. Berikut adalah algoritma genetika secara umum [11].

1. Memisalkan $m = 0$.
2. Menentukan generasi awal C_m .
3. Selama solusi belum konvergen, maka:
 - a. Mengevaluasi nilai *fitness* dari setiap kromosom $\vec{C}_{m,n} \in C_m$.
 - b. Menentukan $m = m + 1$.
 - c. Memilih orang tua (*parents*) dari C_{m-1} .
 - d. Melakukan rekombinasi terhadap orang tua (*parents*) yang terpilih dengan operator *crossover* untuk menghasilkan keturunan (*offspring*) O_m .
 - e. Melakukan mutasi pada keturunan (*offspring*) O_m .
 - f. Memilih generasi baru C_m dari generasi sebelumnya C_{m-1} dan keturunan (*offspring*) O_m .

Beberapa kriteria bahwa kondisi konvergen telah terpenuhi adalah apabila banyaknya generasi maksimum telah terlewati, individu dengan *fitness* terbaik telah diperoleh, atau rata-rata nilai *fitness* atau nilai *fitness* maksimum tidak berubah secara signifikan selama m generasi. Sedangkan berikut adalah penjelasan tentang operator genetika dan komponen lain dalam algoritma genetika.

A. Seleksi atau Reproduksi

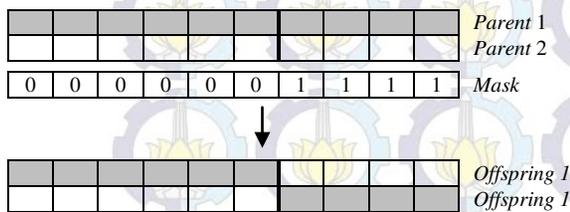
Konsep utama dalam seleksi yaitu individu dengan *fitness* yang tinggi memiliki peluang yang lebih tinggi untuk bereproduksi lebih banyak daripada individu lain. Seleksi akan menghasilkan keturunan (*offspring*) pada generasi berikutnya. Salah satu jenis seleksi adalah seleksi rolet (*roulette selection*). Langkah awal dalam seleksi rolet yaitu menghitung nilai

fitness $f_i (\geq 0)$, dengan $i = 1, \dots, N$ untuk setiap kromosom i dan jumlahan $\sum_{i=1}^N f_i$ sehingga diperoleh probabilitas kromosom akan terpilih atau bertahan yaitu p_i , $p_i = \frac{f_i}{\sum_{i=1}^N f_i}$.

Setiap kromosom merupakan potongan dari lingkaran rolet, dengan ukuran yang proporsional terhadap nilai *fitness* kromosom. Selanjutnya lingkaran rolet diputar sebanyak N kali. Pada setiap putaran, kromosom yang terpilih oleh penunjuk lingkaran rolet menjadi orang tua (*parents*) pada generasi berikutnya.

B. Crossover

Crossover adalah proses rekombinasi secara acak terhadap kromosom. Berdasarkan probabilitas, *crossover* menunjukkan pertukaran karakter secara parsial di antara dua kromosom. Salah satu operator *crossover* yang telah dikembangkan dalam komputasi adalah *crossover* satu titik (*one-point*) [11]. Prinsip umum dalam *crossover* satu titik adalah memilih satu potongan dalam kromosom *parent* secara acak untuk kemudian saling ditukarkan seperti gambar berikut.



Gambar 1. Operasi *One-point Crossover*

C. Mutasi

Perubahan acak yang terjadi pada potongan kromosom disebut sebagai mutasi. Operator mutasi membantu pengenalan material genetika baru pada kromosom yang ada, sehingga menambah perbedaan karakteristik genetika dalam populasi. Operasi mutasi juga ditentukan dengan probabilitas p_m yang biasanya bernilai kecil, antara 0,001 sampai dengan 0,01 untuk memastikan agar solusi terbaik tidak menyimpang. Ada beberapa jenis mutasi yang telah berkembang, salah satunya adalah mutasi acak yang terjadi ketika beberapa posisi potongan terpilih secara acak dan nilai potongan tersebut akan menjadi negasinya.

D. Pengertian dan Analisis Kredit

Kredit berasal dari kata *credere* yang berarti kepercayaan. Penerima kredit atau debitur harus memenuhi kewajibannya sesuai perjanjian tertulis kepada pemberi kredit atau kreditur. Beberapa macam jalannya kredit menurut Bank Indonesia [1] adalah sebagai berikut, dengan penjelasan definisi dari setiap golongan [12].

1. Golongan 1, yaitu lancar atau *pas* jika kredit tidak menimbulkan masalah.
2. Golongan 2, yaitu dalam perhatian khusus atau *special mention* jika kredit mulai bermasalah sehingga memerlukan perhatian khusus.
3. Golongan 3, yaitu kurang lancar atau *substandard* jika dibayar secara tersendat, namun masih mampu dibayar.
4. Golongan 4, yaitu diragukan atau *doubtful* jika kemampuan debitur untuk membayar semakin tidak dapat dipastikan.

5. Golongan 5, yaitu macet atau *loss* jika debitur tidak mampu lagi membayar pinjamannya.

Berdasarkan penggolongan kredit tersebut, golongan 1 disebut sebagai *performing loan* (PL), sedangkan golongan 2 sampai dengan golongan 5 termasuk dalam kategori *non performing loan* (NPL) atau kredit bermasalah. Bank pada umumnya menggunakan konsep 5C berikut dalam menilai kelayakan nasabah untuk menerima fasilitas kredit [1].

1. *Character* atau sifat pribadi.
2. *Capacity* atau kemampuan usaha, baik dari segi pengusaha maupun tempat usahanya.
3. *Capital* atau modal.
4. *Condition of economy* atau kondisi ekonomi pada masa ini dan masa mendatang.
5. *Collateral* atau agunan.

III. METODOLOGI PENELITIAN

A. Data

Data dalam penelitian adalah data sekunder berupa data penggolongan kredit dan karakteristik dari debitur di PT Bank X (Persero) Tbk pada tahun 2012. PT Bank X (Persero) Tbk memberikan 1.000 data debitur dengan penggolongan kredit yang telah ditetapkan dalam dua kategori, yakni kredit baik atau *performing loan* (PL) dan kredit buruk atau *non performing loan* (NPL). Variabel dependen (Y) dan variabel independen (X) yang digunakan dalam penelitian dijelaskan pada Tabel 1 berikut. Penentuan variabel penelitian ini mengacu pada rujukan [4] mengenai klasifikasi kredit pada studi kasus data kredit Bank "X".

Tabel 1.
Variabel Penelitian dan Skala Pengukuran

No.	Variabel	Skala Pengukuran
1	Penggolongan kredit (Y)	Nominal $Y(0)$ untuk Y = kredit baik $Y(1)$ untuk Y = kredit buruk
2	Usia (X_1)	Rasio (dalam satuan tahun)
3	Jenis kelamin (X_2)	Nominal $X_2(0)$ untuk X_2 = laki-laki $X_2(1)$ untuk X_2 = perempuan
4	Status pernikahan (X_3)	Nominal $X_3(0)$ untuk X_3 = belum menikah $X_3(1)$ untuk X_3 = menikah $X_3(2)$ untuk X_3 = duda atau janda
5	Pendidikan (X_4)	Nominal $X_4(0)$ untuk X_4 = tamat SD dan sederajat \vdots $X_4(3)$ untuk X_4 = diploma, sarjana, atau selebihnya
6	Kepemilikan rumah (X_5)	Nominal $X_5(0)$ untuk X_5 = milik sendiri $X_5(1)$ untuk X_5 = milik orang tua \vdots $X_5(5)$ untuk X_5 = lainnya
7	Pekerjaan (X_6)	Nominal $X_6(0)$ untuk X_6 = PNS) $X_6(1)$ untuk X_6 = Pegawai BUMN atau BUMD \vdots $X_6(8)$ untuk X_6 = Lainnya

Tabel 1.
Variabel Penelitian dan Skala Pengukuran (lanjutan)

No.	Variabel	Skala Pengukuran
8	Pendapatan bersih (X_7)	Rasio (dalam satuan rupiah)
9	Pinjaman lain (X_8)	Nominal $X_8(0)$ untuk $X_8 =$ tidak ada $X_8(1)$ untuk $X_8 =$ ada
10	Tenor (X_9) atau jangka waktu pelunasan kredit	Rasio (dalam satuan bulan)
11	Jenis usaha (X_{10})	Nominal $X_{10}(0)$ untuk $X_{10} =$ Industri Pengolahan $X_{10}(1)$ untuk $X_{10} =$ Pertanian, Kehutanan, dan Perikanan \vdots $X_{10}(14)$ untuk $X_{10} =$ Kebudayaan, Hiburan, dan Rekreasi $X_{10}(15)$ untuk $X_{10} =$ Lainnya
12	Lama hubungan baik dengan bank (X_{11})	Rasio (dalam satuan tahun)
13	Nominal Account (X_{12}) atau banyaknya tabungan yang dimiliki debitur	Rasio (dalam satuan rupiah)

B. Langkah Analisis

Langkah-langkah dalam menganalisis faktor yang mempengaruhi penggolongan kredit pada data kredit debitur di PT Bank X (Persero) Tbk menggunakan metode *Hybrid Genetic Algorithm - Logistic Regression* dijelaskan sebagai berikut.

1. Menghitung dan menyajikan statistik deskriptif dari setiap variabel penelitian.
2. Melakukan uji independensi antara variabel dependen (Y) dengan variabel independen (X).
3. Melakukan analisis regresi logistik dengan menggunakan metode *stepwise* dengan seleksi *forward* yang diikuti oleh eliminasi *backward*, sehingga mendapatkan model terbaik untuk faktor-faktor yang mempengaruhi penggolongan kredit pada data kredit debitur di PT Bank X (Persero) Tbk. Setelah itu, menjelaskan interpretasi dan menghitung ukuran kesesuaian model regresi logistik terbaik.
4. Menggunakan *hybrid* algoritma genetika dengan regresi logistik untuk mendapatkan nilai estimasi parameter model regresi logistik dengan ukuran kesesuaian model berupa R_L^2 yang bernilai lebih tinggi atau MSE yang bernilai lebih rendah dibandingkan ukuran kesesuaian model dari langkah 3. *Hybrid* algoritma genetika dengan regresi logistik dilakukan pada penentuan populasi awal, yaitu dengan menyisipkan satu kromosom yang terdiri dari nilai estimasi parameter model regresi logistik terbaik yang diperoleh pada langkah 3.
5. Membandingkan hasil pemodelan pada langkah 4 dan 5 berdasarkan ukuran kesesuaian model.

IV. HASIL DAN PEMBAHASAN

Dari 1.000 data debitur di PT Bank X (Persero) Tbk, terdapat 64 data debitur yang tidak lengkap, sehingga analisis yang dilakukan dalam penelitian ini hanya menggunakan data dari 936 debitur di PT Bank X (Persero) Tbk dengan hasil dan pembahasan berikut.

A. Karakteristik Debitur di PT Bank X (Persero) Tbk.

Berdasarkan data dari 936 debitur di PT Bank X (Persero) Tbk, 80,3% debitur dikategorikan sebagai debitur dengan kredit baik, dan 19,7% sisanya dikategorikan sebagai debitur dengan kredit buruk di PT Bank X (Persero) Tbk. Selanjutnya adalah Tabel 2 yang memuat statistik dari variabel independen dengan skala pengukuran rasio.

Tabel 2.
Statistik Deskriptif Debitur di PT Bank X (Persero) Tbk

Statistik	Usia (tahun)	Pendapatan bersih (rupiah)	Tenor (bulan)
Rata-rata	35,57	5.233.998	37,175
Deviasi standar	7,73	5.809.700	12,254
Median	35	3.543.952	36
Minimum	20	1.252.204	12
Maksimum	62	63.786.168	60

Statistik	Lama hubungan baik dengan bank (tahun)	Nominal account (US\$)	Rasio hutang (persen)
Rata-rata	3,61	13.686.500	31,11
Deviasi standar	2,81	112.643.440	11,90
Median	3,00	1.808.808	29,17
Minimum	0,00	18.599	7,48
Maksimum	19,00	2.880.181.084	130,72

Statistik pada Tabel 2. menunjukkan bahwa rata-rata usia debitur di PT Bank X (Persero) Tbk adalah usia muda. Namun pendapatan bersih yang dihasilkan debitur setiap bulannya sangat bervariasi dengan rata-rata penghasilan yang relatif sedang. Meskipun demikian, setengah dari debitur di PT Bank X (Persero) Tbk berpenghasilan kurang dari 3,5 juta rupiah per bulan. Selain itu, rata-rata tenor atau jangka waktu pelunasan kredit yang harus ditempuh debitur di PT Bank X (Persero) Tbk adalah sekitar tiga tahun. Statistik pada Tabel 2 juga menunjukkan bahwa terdapat debitur yang belum pernah berhubungan baik dengan bank sebelumnya. Sedangkan *nominal account* yang dimiliki oleh debitur di PT Bank X (Persero) Tbk sangat bervariasi dan rata-rata rasio hutang debitur terhadap pendapatannya adalah 31,11%. Dari statistik ini, diketahui juga bahwa terdapat debitur yang jumlah hutangnya melebihi pendapatannya.

Selain itu, sebagian besar dari debitur di PT Bank X (Persero) Tbk adalah laki-laki, dan sebagian besar debitur berstatus telah menikah. Hampir semua debitur di PT Bank X (persero) Tbk adalah debitur dengan pendidikan yang relatif cukup tinggi, yaitu minimal adalah lulusan SMA dan sederajat serta perguruan tinggi, dengan pekerjaan dari sebagian besar debitur adalah pegawai swasta. Adapun status kepemilikan rumah yang ditempati oleh debitur kebanyakan adalah milik sendiri dan milik orang tua. Untuk status adanya pinjaman lain di bank, termasuk PT Bank X (Persero) Tbk, sebagian besar

debitur tidak memiliki pinjaman lain, yang mana sebagian besar debitur juga menjalankan usaha di bidang industri pengolahan.

B. Pemodelan dengan Analisis Regresi Logistik

Bagian berikut ini adalah hasil dan pembahasan analisis regresi logistik biner untuk memodelkan faktor-faktor yang mempengaruhi penggolongan kredit terhadap debitur di PT Bank X (Persero) Tbk. Dalam melakukan analisis, ditetapkan nilai α sebesar 0,05.

1. Uji Independensi

Hasil pengujian independensi antara penggolongan kredit dengan variabel independen berupa kategori menunjukkan bahwa pekerjaan, pinjaman lain, dan jenis usaha menghasilkan P -value yang bernilai kurang dari α , yang berarti penggolongan kredit di PT Bank X (Persero) Tbk memiliki hubungan dengan pekerjaan debitur, ada atau tidaknya pinjaman lain yang dimiliki debitur, dan jenis usaha yang dijalankan debitur di PT Bank X (Persero) Tbk. Berikutnya adalah Tabel 3 yang memuat hasil pengujian independensi antara penggolongan kredit di PT Bank X (Persero) Tbk dengan variabel independen yang berupa non kategori.

Tabel 3.
Hasil Uji Independensi Variabel Non Kategori

Variabel Independen	Nilai Eta
Usia	0,212
Pendapatan bersih	0,982
Tenor	0,127
Lama hubungan baik dengan bank	0,229
Nominal account	0,801
Rasio hutang	0,945

Berdasarkan Tabel 3, variabel pendapatan bersih, *nominal account*, dan rasio hutang menghasilkan eta yang bernilai tinggi, yang berarti variabel tersebut memiliki hubungan yang kuat dengan penggolongan kredit debitur di PT Bank X (Persero) Tbk. Sedangkan variabel lainnya memiliki hubungan yang lemah terhadap penggolongan kredit di PT Bank X (Persero) Tbk.

2. Estimasi dan Uji Signifikansi Koefisien Parameter Regresi Logistik

Variabel independen yang dimodelkan adalah variabel independen berupa non kategori dan variabel independen berupa kategori yang memiliki hubungan dengan penggolongan kredit berdasarkan hasil uji independensi. Untuk mendapatkan model regresi logistik terbaik, dilakukan pemodelan dengan melakukan seleksi variabel independen yang berpengaruh secara signifikan terhadap penggolongan kredit di PT Bank X (Persero) Tbk. Seleksi variabel dilakukan dengan metode *stepwise* yang terdiri dari seleksi *forward* kemudian diikuti dengan eliminasi *backward*. Batas P -value untuk seleksi bernilai 0,05 dan batas P -value untuk mengeliminasi bernilai 0,10. Langkah akhir pada metode *stepwise* menghasilkan empat variabel independen yang memiliki pengaruh signifikan terhadap penggolongan kredit di PT Bank X (Persero) Tbk, dengan nilai estimasi, P -value, dan OR dari parameter pada Tabel 4 berikut.

Tabel 4.
Statistik Parameter Model Terbaik

Variabel Independen	Nilai Estimasi	P -value	OR
Intercept (β_0)	-0,935	0,002	0,392
Pinjaman lain (β_8)			
Ada pinjaman lain (1)	-0,614	0,004	0,541
Jenis usaha (β_{10})			
Penyediaan akomodasi dan makan minum (6)	1,104	0,001	3,016
Lama hubungan baik dengan bank (β_{11})	-0,224	0,000	0,799
Rasio hutang terhadap pendapatan (β_{13})	0,014	0,042	1,014

Berdasarkan nilai estimasi parameter dari model regresi logistik terbaik, diperoleh *logit* untuk model regresi logistik sebagai berikut.

$$g(x) = -0,935 - 0,614 x_8(1) + 1,104 x_{10}(6) - 0,223 x_{11} + 0,014 x_{13}$$

atau dapat ditulis sebagai berikut.

$$g(x) = -0,935 - 0,614 (\text{Ada pinjaman lain}) + 1,104 (\text{Jenis usaha penyediaan akomodasi dan makan minum}) - 0,224 (\text{Lama hubungan baik dengan bank}) + 0,014 (\text{Rasio hutang terhadap pendapatan})$$

Berdasarkan *logit* $g(x)$, koefisien parameter untuk adanya pinjaman lain dan lama hubungan baik dengan bank bernilai negatif, sedangkan koefisien parameter jenis usaha penyediaan akomodasi dan makan minum serta rasio hutang terhadap pendapatan bernilai positif. Hal itu berarti adanya pinjaman lain dan pertambahan lamanya hubungan baik antara debitur dengan bank menurunkan probabilitas seorang debitur tergolong sebagai debitur dengan kredit baik di PT Bank X (Persero) Tbk. Begitu pula dengan jenis usaha penyediaan akomodasi dan makan minum serta pertambahan satu persen rasio hutang debitur terhadap pendapatannya meningkatkan probabilitas seorang debitur tergolong sebagai debitur dengan kredit baik di PT Bank X (Persero) Tbk.

Berdasarkan nilai OR pada Tabel 4, kecenderungan debitur yang tidak mempunyai pinjaman lain tergolong sebagai debitur dengan kredit baik di PT Bank X (Persero) Tbk adalah sekitar dua kali daripada debitur yang mempunyai pinjaman lain. Selain itu, debitur yang menjalankan usaha di bidang penyediaan akomodasi dan makan minum cenderung tergolong sebagai debitur baik sekitar tiga kali daripada debitur yang menjalankan usaha di industri pengolahan.

3. Kesesuaian Model Regresi Logistik

Model regresi logistik terbaik menghasilkan R_L^2 yang bernilai rendah, yaitu 0,0740. Hal itu berarti model tersebut hanya dapat menjelaskan 7,40% hubungan antara variabel-variabel independen pada model dengan penggolongan kredit di PT Bank X (Persero) Tbk dengan MSE bernilai 0,1549. Meskipun menghasilkan R_L^2 yang bernilai rendah, namun ketepatan klasifikasi yang dihasilkan model tersebut relatif baik, yaitu sebesar 81,52%. Oleh sebab itu, selanjutnya dilakukan analisis dengan melakukan pemodelan terhadap faktor-faktor yang mempengaruhi penggolongan kredit di PT Bank X (Persero) Tbk menggunakan algoritma genetika dengan tujuan untuk mendapatkan model yang menghasilkan nilai R_L^2 yang lebih tinggi atau nilai MSE yang lebih rendah.

C. Pemodelan dengan Metode Hybrid Genetic Algorithm - Logistic Regression

Untuk melakukan estimasi nilai parameter model regresi logistik dengan algoritma genetika, 100 generasi dengan populasi berukuran 100 ditetapkan dengan salah satu kromosom pada populasi awal berupa nilai estimasi parameter model regresi logistik terbaik yang dimuat dalam Tabel 4. Algoritma genetika menggunakan operator seleksi jenis rolet (*roulette selection*) serta *crossover* jenis satu titik (*one-point crossover*). Probabilitas *crossover* adalah sebesar 0,8 dengan probabilitas mutasi acak sebesar 0,1. Sedangkan fungsi *fitness* yang dievaluasi adalah *pseudo* R^2 atau R_L^2 dan MSE. Algoritma genetika dijalankan sebanyak lima kali untuk mendapatkan nilai *fitness* terbaik. Rata-rata nilai *fitness* dari hasil tersebut adalah rata-rata nilai *fitness* yang dihasilkan algoritma genetika. Nilai R_L^2 dan MSE dari masing-masing percobaan algoritma genetika dimuat pada Tabel 5 berikut.

Tabel 5.
Nilai Fitness R_L^2 dan MSE dari Setiap Percobaan

Percobaan	Nilai (R_L^2)	Nilai MSE
1	0,0766	0,1454
2	0,0771	0,1459
3	0,0771	0,1453
4	0,0760	0,1454
5	0,0768	0,1459

Berdasarkan hasil percobaan algoritma genetika pada Tabel 5, diperoleh rata-rata nilai R_L^2 adalah sebesar 0,0767 dan rata-rata nilai MSE adalah 0,1459.

D. Perbandingan Model Terbaik

Estimasi parameter model regresi logistik dengan *maximum likelihood estimation* (MLE) dan *hybrid* algoritma genetika dengan regresi logistik menghasilkan nilai estimasi parameter serta nilai R_L^2 , MSE, dan ketepatan klasifikasi yang ditampilkan pada Tabel 6 berikut.

Tabel 6.
Nilai Estimasi Parameter dan Ukuran Kesesuaian

Variabel Independen	MLE	Hybrid GA (berdasarkan R_L^2)	Hybrid GA (berdasarkan MSE)
Intercept (β_0)	-0,935	-1,083	-1,062
Pinjaman lain (β_8)			
Ada pinjaman lain (1)	-0,614	-0,614	-0,614
Jenis usaha (β_{10})			
Penyediaan komodasi dan makan minum (6)	1,104	1,265	1,104
Lama hubungan baik dengan bank (β_{11})	-0,224	-0,224	-0,224
Rasio hutang terhadap pendapatan (β_{13})	0,014	0,014	0,014
	R_L^2	0,0771	0,0768
	MSE	0,1453	0,1453
Ketepatan klasifikasi	0,8152	0,8152	0,8152

Berdasarkan Tabel 6, *hybrid* algoritma genetika dengan regresi logistik menghasilkan model regresi logistik dengan nilai R_L^2 yang lebih tinggi dan nilai MSE yang lebih rendah

daripada yang dihasilkan oleh model regresi logistik dengan metode MLE, meskipun selisih nilai R_L^2 dan nilai MSE yang dihasilkan kedua metode bernilai sangat kecil. Selain itu, ketepatan klasifikasi yang diperoleh dari metode MLE maupun *hybrid* algoritma genetika dengan regresi logistik menghasilkan tabel klasifikasi dengan prediksi yang sama. Berdasarkan hasil perbandingan tersebut, *hybrid* algoritma genetika dengan regresi logistik menghasilkan model regresi logistik yang sedikit lebih baik. Selain itu, hanya nilai estimasi parameter untuk *intercept* dan variabel jenis usaha di bidang penyediaan akomodasi dan makan minum yang berbeda dengan nilai estimasi parameter berdasarkan MLE.

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan hasil penelitian, model regresi logistik menghasilkan empat faktor yang berpengaruh terhadap penggolongan kredit di PT Bank (X) Persero Tbk, yaitu adanya pinjaman lain, jenis usaha penyediaan akomodasi dan makan minum, lama hubungan debitur dengan bank, dan rasio hutang debitur terhadap pendapatannya.

B. Saran

Dalam melakukan penelitian selanjutnya mengenai penggolongan kredit, peneliti sebaiknya melengkapi data dengan variabel jenis kredit dan besarnya pinjaman dari setiap debitur. Selain itu, sebaiknya melengkapi data pada variabel lainnya agar tidak ada kategori yang memiliki nilai harapan kurang dari lima.

DAFTAR PUSTAKA

- [1] Tamin, N. (2012). Kiat Menghindari Kredit Macet. Jakarta: Dian Rakyat.
- [2] Mandiri. (2014). *Transformasi Bank Mandiri*. Diakses pada 4 Februari 2014, dari http://www.bankmandiri.co.id/corporate01/about_profile.asp
- [3] Metrotvnews.com. (2013). *Kualitas KUR Bank Mandiri Memburuk*. Diakses pada 4 Februari 2014, dari <http://www.metrotvnews.com/metronews/read/2013/11/21/2/196153/Kualitas-KUR-Bank-Mandiri-Memburuk>
- [4] Misdiati, L. (2013). *Analisis Klasifikasi Kredit Menggunakan Metode Newton Truncated – Kernel Logistic Regression (NTR-KLR) (Studi Kasus: Data Kredit Bank "X")*. Tugas akhir yang tidak dipublikasikan, Institut Teknologi Sepuluh Nopember, Surabaya.
- [5] Meyer, M. C. (2003). An Evolutionary Algorithm with Applications to Statistics. *Journal of Computational and Graphical Statistics*, 12 (2), 265-281.
- [6] Khan, M. Z. R. & Bajpai, A. K. (2013). Genetic Algorithm and Its Application in Mechanical Engineering. *International Journal of Engineering Research & Technology*, 2 (5), 677-683.
- [7] Sakawa, M. (2002). *Genetic Algorithms and Fuzzy Multiobjective Optimization*. United States of America: Kluwer Academic Publisher.
- [8] Mandiri. (2014). *Transformasi Bank Mandiri*. Diakses pada 4 Februari 2014, dari http://www.bankmandiri.co.id/corporate01/about_profile.asp
- [9] Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2009). *Multivariate Data Analysis*. United States of America: Prentice Hall.
- [10] Hosmer, J. W. & Lemeshow, S. (2000). *Applied Logistic Regression 2nd Edition*. United States of America: John Wiley & Sons, Inc.
- [11] Engelbrecht, A. P. (2002). *Computational Intelligence, An Introduction*. England: John & Wiley Sons, Ltd.
- [12] Kasmir. (2013). *Dasar-Dasar Perbankan Edisi Revisi*. Jakarta: PT Raja Grafindo Persada.
- [13] Sakawa, M. (2002). *Genetic Algorithms and Fuzzy Multiobjective Optimization*. United States of America: Kluwer Academic Publisher