

**OPTIMASI REGRESI LOGISTIK MENGGUNAKAN ALGORITMA GENETIKA
UNTUK PEMODELAN FAKTOR-FAKTOR YANG MEMPENGARUHI
PENGKOLONGAN KREDIT BANK
(Studi Kasus: Debitur di PT BPR Gunung Lawu Klaten Periode Tahun 2017)**

Reno Penggalih Surya Wardhani¹, Sudarno², Di Asih I Maruddani³

^{1,2,3} Departemen Statistika FSM Universitas Diponegoro

dsghani@gmail.com

ABSTRACT

Credit is the greatest asset managed by banks and also the most dominant contributor to the bank's income. But in its implementation, the provision of credit to the public is at risk for non-performing loans. For this reason, creditors try to minimize the occurrence of non-performing loans by predicting credit risk appropriately. In this study, modeling the factors that influence credit classification at PT BPR Gunung Lawu is useful for predicting the credit risk of prospective debtors. Modeling are done using logistic regression and genetic algorithms. Factors suspected of influencing credit classification include age, gender, marital status, education, home ownership, employment, net income, tenor, type of business, type of loan, type of loan interest, and loan size. Estimated model parameters obtained from logistic regression were optimized using genetic algorithms. The fitness function used is *pseudo* R^2 or R_L^2 and MSE. The best model is generated by modeling with genetic algorithms based on MSE fitness. The model produces the highest R_L^2 value of 0.1958 and the lowest MSE value of 0.1648 with classification accuracy of 75.33%.

Keywords: credit classification, logistic regression, genetic algorithms

1. PENDAHULUAN

Semakin luasnya masyarakat mengenal kredit, sebanding dengan semakin banyaknya Bank Perkreditan Rakyat (BPR) yang bermunculan. Prosedur pengajuan kredit yang dipermudah mendorong masyarakat melakukan pinjaman guna memenuhi kebutuhan finansial di segala bidang, termasuk memenuhi kebutuhan sehari-hari (konsumtif), tidak hanya kebutuhan finansial di bidang ekonomi.

Tujuan awal pemberian kredit memang sudah baik, namun pada pelaksanaannya memiliki risiko terjadinya kredit bermasalah, yang menyebabkan kredit tidak berjalan dengan semestinya atau yang disebut kredit macet, yaitu kredit yang mengalami kesulitan pelunasan akibat adanya faktor-faktor atau unsur kesengajaan atau karena kondisi di luar kemampuan debitur. 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 secara tepat.

Regresi logistik merupakan salah satu metode klasifikasi yang biasa digunakan untuk mengetahui hubungan antara dua satu variabel respon yang bersifat kategori dan beberapa variabel prediktor yang bersifat kategori maupun kontinyu (Agresti, 1990). *Hybrid* terhadap metode lain dapat dilakukan untuk meningkatkan ukuran kesesuaian model. Salah satu metode yang dapat digunakan adalah algoritma genetika (*Genetic Algorithm*). Algoritma genetika (AG) adalah suatu algoritma pencarian yang berbasis pada mekanisme seleksi alam dan genetika. Algoritma genetika merupakan salah satu algoritma yang sangat tepat digunakan dalam menyelesaikan masalah optimasi kompleks, yang sulit dilakukan oleh metode konvensional (Desiani dan Arhami, 2006). Pada penelitian ini, *hybrid* algoritma genetika dilakukan pada regresi logistik biner dengan tujuan dapat

membantu mengoptimasi estimasi parameter untuk faktor-faktor yang berpengaruh terhadap penggolongan kredit debitur berdasarkan nilai *pseudo R²* dan MSE.

2. TINJAUAN PUSTAKA

2.1. Regresi Logistik

Regresi logistik merupakan salah satu bentuk analisis data dengan menggunakan teknik regresi yang dapat digunakan untuk mengetahui hubungan antara satu variabel respon yang bersifat kategori dan beberapa variabel prediktor yang bersifat kategori maupun kontinyu (Agresti, 1990).

2.1.1. Uji Independensi

Uji Independensi ialah uji yang digunakan untuk mengetahui hubungan antara dua variabel (Agresti, 1990). Untuk menguji independensi dari tabel kontingensi $I \times J$.

Hipotesis:

H_0 : kedua variabel adalah independen

H_1 : kedua variabel adalah dependen

Tingkat Signifikansi:

α

Statistik Uji:

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

n_{ij} = banyaknya data dengan kategori A_i yang termasuk dalam kategori B_j

$\hat{\mu}_{ij} = \frac{n_{i+}n_{+j}}{n}$, estimasi nilai harapan

Kriteria Uji:

H_0 ditolak jika nilai $\chi^2 > \chi^2_{(\alpha, (I-1)(J-1))}$ atau p-value $< \alpha$

Untuk mengetahui hubungan antara dua variabel yang berskala data tidak sama.

$$\eta = \sqrt{1 - \frac{\sum Y_T^2 - (n_1)(\bar{Y}_1)^2 - (n_2)(\bar{Y}_2)^2}{\sum Y_T^2 - (n_1 + n_2)(\bar{Y}_T)^2}}$$

2.1.2. Regresi Logistik Biner

Regresi logistik biner merupakan suatu metode analisis data yang digunakan untuk menjelaskan hubungan antara beberapa variabel prediktor X terhadap variabel respon Y yang bersifat dikotomi atau biner (Hosmer dan Lemeshow, 1989).

Persamaan model regresi logistik:

$$\pi(x_i) = \frac{e^{\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}}} \quad (1)$$

dengan fungsi logit:

$$g(x_i) = \ln \left[\frac{\pi(x_i)}{1 - \pi(x_i)} \right] = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} \quad (2)$$

Persamaan (1) dan (2) dapat disederhanakan menjadi:

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

2.1.3. Regresi Logistik Bertatar

Regresi logistik bertatar (*stepwise logistic regression*) digunakan untuk menentukan variabel-variabel prediktor yang bisa membedakan respon yang diamati. Prosedur ini memilih atau menghilangkan variabel prediktor satu persatu dari model sampai ditemukan variabel prediktor yang berpengaruh nyata terhadap model (Hosmer dan Lemeshow, 1989).

2.1.4. Estimasi Parameter

Menurut Agresti (1990), untuk menentukan estimasi parameter regresi logistik biner dapat menggunakan metode *Maximum Likelihood Estimator* (MLE) yang membutuhkan turunan pertama dan kedua dari fungsi log likelihood.

Jika ditulis dalam bentuk matriks, maka turunan pertamanya:

$$\frac{\partial L(\beta)}{\partial \beta} = \mathbf{X}'(\mathbf{Y} - \boldsymbol{\pi}(x_i))$$

dengan :

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{p1} \\ 1 & x_{12} & \dots & x_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \dots & x_{pn} \end{bmatrix} \text{ dan } \mathbf{Y} - \boldsymbol{\pi}(x_i) = \begin{bmatrix} y_1 - \pi(x_1) \\ y_2 - \pi(x_2) \\ \vdots \\ y_n - \pi(x_n) \end{bmatrix}$$

Jika ditulis dalam bentuk matriks, persamaan turunan keduanya menjadi:

$$\frac{\partial^2 L(\beta)}{(\partial \beta)^2} = \mathbf{X}'\mathbf{V}\mathbf{X}$$

dengan:

$$\mathbf{V} = \begin{bmatrix} \pi(x_1)[1 - \pi(x_1)] & 0 & \dots & 0 \\ 0 & \pi(x_2)[1 - \pi(x_2)] & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \pi(x_n)[1 - \pi(x_n)] \end{bmatrix}$$

2.1.5. Pengujian Parameter

2.1.5.1. Uji Rasio Likelihood

Uji Rasio Likelihood digunakan untuk menguji model yang didapat dari estimasi parameter, apakah variabel prediktor signifikan mempengaruhi model atau tidak secara bersama-sama.

Hipotesis:

H_0 : $\beta_j = 0$, artinya tidak ada pengaruh antara variabel prediktor dengan variabel respon, $j = 1, 2, \dots, p$

H_1 : paling sedikit ada satu $\beta_j \neq 0$, $j = 1, 2, \dots, p$, artinya paling tidak ada satu variabel prediktor yang berpengaruh terhadap variabel respon

Tingkat Signifikansi:

α

Statistik Uji:

$$G = -2 \ln \left(\frac{\text{likelihood tanpa variabel bebas}}{\text{likelihood dengan variabel bebas}} \right)$$

Kriteria Uji:

H_0 ditolak jika nilai $G > \chi^2_{(\alpha, p)}$ atau p-value $< \alpha$

2.1.5.2. Uji Wald

Uji Wald merupakan uji signifikansi parameter untuk masing-masing variabel prediktor, yang gunanya untuk mengetahui apakah masing-masing variabel prediktor tersebut memiliki pengaruh terhadap model atau tidak.

Hipotesis :

H_0 : $\beta_j = 0$, artinya tidak ada pengaruh variabel prediktor ke- j terhadap variabel respon, $j = 1, 2, \dots, p$

H_1 : $\beta_j \neq 0$, artinya ada pengaruh variabel prediktor ke- j terhadap variabel respon, $j = 1, 2, \dots, p$

Statistik Uji:

$$W_j = \left\{ \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right\}^2$$

Tingkat Signifikansi:

α

Kriteria Uji:

H_0 ditolak jika nilai $W_j > \chi^2_{(\alpha, 1)}$ atau p-value $< \alpha$

2.1.6. Kesesuaian Model

Kesesuaian model menunjukkan bagaimana model yang terbentuk dapat mendeskripsikan variabel respon (Hosmer dan Lemeshow, 1989). Salah satu ukuran kesesuaian model regresi logistik adalah *pseudo R²* yang serupa dengan *R²* pada analisis regresi.

$$R_L^2 = \frac{L_0 - L_p}{L_0} = 1 - \frac{L_p}{L_0} \quad (4)$$

2.1.7. Odds Ratio

Menurut Hosmer dan Lemeshow (1989), *odds ratio* adalah ukuran yang memperkirakan berapa besar kecenderungan variabel prediktor terhadap variabel respon.

$$\hat{\theta} = \frac{\text{odds 1}}{\text{odds 2}} = \frac{\frac{\pi(x_1)}{1-\pi(x_1)}}{\frac{\pi(x_0)}{1-\pi(x_0)}} = \frac{\frac{\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}}{1}}{\frac{\frac{e^{\beta_0}}{1+e^{\beta_0}}}{1}} = \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} = \frac{e^{\beta_0} e^{\beta_1}}{e^{\beta_0}} = e^{\beta_1} \quad (5)$$

Sehingga $\hat{\theta} = e^{\beta}$

2.2. Algoritma Genetika

Algoritma Genetika (AG) merupakan suatu teknik optimasi yang didasarkan pada prinsip genetik dan seleksi alam. Dalam AG, populasi terbentuk dari banyak individu yang berkembang sesuai aturan seleksi spesifik dengan memaksimalkan nilai fitness (Haupt dan Haupt, 2004).

2.2.1. Struktur Umum Algoritma Genetika

Kusumadewi (2003) menerangkan bahwa teknik pencarian algoritma genetika dilakukan sekaligus atas sejumlah solusi yang mungkin yang dikenal dengan istilah populasi. Individu yang terdapat dalam satu populasi disebut kromosom. Populasi awal dibangun secara acak, sedangkan populasi berikutnya merupakan hasil evolusi kromosom-kromosom melalui iterasi yang disebut dengan istilah generasi. Pada setiap generasi, kromosom akan melalui proses evaluasi dengan menggunakan alat ukur yang disebut dengan fungsi *fitness*. Nilai *fitness* dari suatu kromosom akan menunjukkan kualitas kromosom dalam populasi tersebut. Generasi berikutnya dikenal dengan istilah anak (*offspring*), yang terbentuk dari gabungan dua kromosom generasi sekarang yang bertindak sebagai orangtua (*parent*) yang terpilih menggunakan operator kawin silang (*crossover*). Selain operator kawin silang, suatu kromosom dapat dimodifikasi dengan menggunakan operator mutasi.

2.2.2. Komponen-Komponen Utama dalam Algoritma Genetika

Algoritma genetika memiliki beberapa komponen utama. Pengkodean yaitu suatu teknik untuk menyatakan populasi awal sebagai calon solusi suatu masalah ke dalam suatu kromosom sebagai suatu kunci pokok persoalan ketika menggunakan algoritma genetika (Desiani dan Arhami, 2006). Gen-gen yang diinisialisasi merupakan estimasi nilai mula-mula yang mengandung informasi dalam bentuk kode-kode tertentu. Satu gen mewakili satu parameter yang akan diestimasi nilainya agar suatu fungsi optimal.

Setelah itu suatu individu dievaluasi berdasarkan suatu fungsi tertentu sebagai ukuran performansinya yang disebut nilai *fitness*. Di dalam evolusi alam, individu yang bernilai *fitness* tinggi memiliki tingkat ketahanan hidup dan tingkat reproduksi yang tinggi pula jika dibandingkan dengan individu yang bernilai *fitness* lebih rendah (Kusumadewi, 2003).

Individu-individu dalam generasi akan diseleksi untuk dipilih menjadi induk, salah satunya menggunakan *roulette wheel selection*. Setelah individu-individu terpilih dalam proses seleksi, selanjutnya pada individu-individu tersebut dilakukan kawin silang. Tujuannya adalah untuk menambah keanekaragaman dalam populasi. Selain proses kawin silang, juga dilakukan proses mutasi, yaitu proses untuk mengubah nilai dari satu atau beberapa gen dalam suatu kromosom untuk menghindari terjadinya konvergensi prematur dan tetap menjaga perbedaan (*diversity*) kromosom-kromosom dalam populasi.

Adanya proses kawin silang dan mutasi, berakibat muncul risiko hilangnya suatu individu bernilai *fitness* tertinggi, karena proses seleksi dilakukan secara random sehingga tidak ada jaminan bahwa suatu individu bernilai *fitness* tertinggi akan selalu terpilih. Kalaupun individu bernilai *fitness* tertinggi terpilih, kemungkinan individu tersebut akan rusak (nilai *fitness* menurun) karena proses kawin silang pasti ada. Untuk menjaga agar individu bernilai *fitness* tertinggi tersebut tidak hilang selama evolusi, maka perlu dibuat satu atau beberapa salinan. Prosedur ini dinamakan dengan elitisme (Suyanto, 2005). Penggantian populasi (*generational replacement*) dalam AG diartikan bahwa semua individu (N individu dalam satu populasi) dari suatu generasi digantikan sekaligus oleh N individu baru hasil kawin silang dan mutasi.

2.3. Ketepatan Klasifikasi

Tingkat kesalahan klasifikasi atau biasa disebut *Apparent Error Rate* (APER) digunakan untuk menilai kemampuan prosedur klasifikasi dalam menentukan keanggotaan kelompok. Bisa juga menggunakan sebaliknya, yaitu *correct error rate* atau tingkat ketepatan klasifikasi (Rencher, 2002).

$$APER(\%) = \frac{n_{12} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}}$$

2.4. Kredit

Pengertian kredit menurut Undang-Undang Perbankan Nomor 10 tahun 1998 adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam-meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam untuk melunasi utangnya setelah jangka waktu tertentu dengan pemberian bunga.

3. METODOLOGI PENELITIAN

3.1. Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari PT BPR Gunung Lawu, Delanggu, Klaten. Data tersebut merupakan data 150 nasabah yang mengajukan kredit di tahun 2017 dengan penggolongan kredit yang telah ditetapkan oleh PT BPR Gunung Lawu dalam dua kategori, yaitu kredit baik atau *performing loan* (PL) dan kredit bermasalah atau *non performing loan* (NPL).

3.2. Variabel Penelitian

Tabel 1. Variabel Penelitian

No.	Variabel	Skala Pengukuran
1	Penggolongan kredit (Y)	Nominal Y(0) untuk Y = kredit baik Y(1) untuk Y = kredit bermasalah
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 = tidak tamat/tamat SD atau sederajat $X_4(1)$ untuk X_4 = tamat SMP atau sederajat $X_4(2)$ untuk X_4 = tamat SMA atau sederajat $X_4(3)$ 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
7	Pekerjaan (X_6)	Nominal $X_6(0)$ untuk X_6 = PNS $X_6(1)$ untuk X_6 = karyawan swasta $X_6(2)$ untuk X_6 = wiraswasta $X_6(3)$ untuk X_6 = petani/pekebun $X_6(4)$ untuk X_6 = pedagang $X_6(5)$ untuk X_6 = sopir $X_6(6)$ untuk X_6 = buruh harian lepas
8	Pendapatan bersih (X_7)	Rasio (dalam satuan rupiah)
9	Tenor atau jangka waktu pelunasan (X_8)	Rasio (dalam satuan bulan)
10	Jenis usaha (X_9)	Nominal $X_9(0)$ untuk X_9 = industri pengolahan $X_9(1)$ untuk X_9 = pertanian, peternakan, kehutanan, atau perikanan $X_9(2)$ untuk X_9 = perdagangan $X_9(3)$ untuk X_9 = bidang jasa $X_9(4)$ untuk X_9 = tidak ada
11	Jenis kredit (X_{10})	Nominal $X_{10}(0)$ untuk X_{10} = investasi $X_{10}(1)$ untuk X_{10} = modal kerja $X_{10}(2)$ untuk X_{10} = produktif $X_{10}(3)$ untuk X_{10} = konsumtif
12	Jenis bunga kredit (X_{11})	Nominal $X_{11}(0)$ untuk X_{11} = bulanan menurun $X_{11}(1)$ untuk X_{11} = tetap/flat $X_{11}(2)$ untuk X_{11} = musiman
13	Besarnya pinjaman dari setiap debitur (X_{12})	Rasio (dalam satuan rupiah)

3.3. Metode Analisis

Metode analisis yang akan digunakan adalah penggunaan algoritma genetika pada regresi logistik biner menggunakan *software* R 3.5.1 dan SPSS 16.0. Adapun langkah-langkah analisis yang dilakukan adalah sebagai berikut.

1. Menginput data.
2. Menghitung statistika deskriptif dan melakukan tabulasi silang antara variabel respon terhadap masing-masing variabel prediktor.
3. Melakukan uji independensi.
4. Melakukan analisis regresi logistik dengan metode *stepwise* untuk mendapatkan model terbaik faktor-faktor yang mempengaruhi penggolongan kredit debitur di PT BPR Gunung Lawu.
 - a. Mendapatkan nilai estimasi parameter model regresi logistik berdasarkan *Maximum Likelihood Estimation* (MLE).
 - b. Melakukan uji signifikansi parameter model regresi logistik.
 - c. Menganalisis ukuran kesesuaian model regresi logistik.
5. Melakukan analisis algoritma genetika untuk mengoptimalkan nilai estimasi parameter model regresi logistik yang telah diperoleh.
 - a. Merepresentasikan nilai parameter dari model regresi logistik yang diperoleh ke dalam bentuk kromosom b. Kemudian menyisipkan kromosom b tersebut ke populasi awal.
 - b. Menghitung nilai *fitness* untuk setiap kromosom. Fungsi *fitness* yang digunakan dalam penelitian ini adalah *pseudo R²* atau R_L^2 dan MSE.
$$g(b) = b_1 + (b_2 \times x_2) + (b_3 \times x_6) + (b_4 \times x_6) + (b_5 \times x_9)$$
$$\pi(b) = \frac{e^{g(b)}}{1 + e^{g(b)}}$$
$$L_p(b) = \sum_{i=1}^n [y_i \times \ln(\pi(b)) + (1 - y_i) \times \ln(1 - \pi(b))]$$
$$R_L^2(b) = 1 - \frac{L_p(b)}{L_0}$$
dengan L_0 adalah *log-likelihood* untuk model regresi yang hanya mengandung *intercept*.
$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \pi(b)_i)^2$$
 - c. Menjalankan prosedur algoritma genetika. Kromosom yang memiliki nilai *fitness* R_L^2 yang lebih tinggi atau nilai *fitness* MSE yang lebih rendah memiliki peluang yang lebih besar untuk terpilih, kemudian melakukan kawin silang dan mutasi dengan peluang tertentu.
 - d. Apabila solusi belum konvergen, maka kembali ke langkah 5.c. Apabila solusi telah konvergen, maka diperoleh nilai parameter model berdasarkan algoritma genetika.
6. Membandingkan model pada langkah 4 dan 5.

4. HASIL DAN PEMBAHASAN

4.1. Karakteristik Debitur di PT BPR Gunung Lawu

Tabel 2. Penggolongan Kredit

Golongan Kredit	Jumlah	Persentase
Baik	102	68%
Bermasalah	48	32%
Total	150	100%

Berdasarkan Tabel 1 dapat diketahui bahwa terdapat 150 data yang digunakan dalam penelitian. Jumlah debitur yang digolongkan ke dalam kredit baik lebih besar dibandingkan dengan jumlah debitur yang digolongkan ke dalam kredit bermasalah.

Tabel 3. Statistik Deskriptif Variabel Non Kategori

Statistik	Usia (tahun)	Pendapatan bersih (rupiah)	Tenor (bulan)	Besar pinjaman (rupiah)
Rata-rata	51,49	5.817.613	29,82	21.916.667
Deviasi Standar	10,07	7.004.809,215	14,35	30.947.315,215
Minimum	28	1.640.000	10	1.000.000
Maksimum	77	75.000.000	60	200.000.000

4.2. Pemodelan dengan Analisis Regresi Logistik

4.2.1. Uji Independensi

Hipotesis:

H_0 : kedua variabel adalah independen

H_1 : kedua variabel adalah dependen

Tingkat Signifikansi:

$\alpha = 0,05$

Statistik Uji:

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

Kriteria Uji:

H_0 ditolak jika nilai $\chi^2 > \chi^2(\alpha; (I-1)(J-1))$ atau p-value $< \alpha$

Tabel 4. Hasil Uji Independensi Variabel Prediktor Kategori

Variabel Prediktor (X_j)	χ^2	P-value	Keputusan
Jenis kelamin (X_2)	4,102	0,043	H_0 ditolak
Status Pernikahan (X_3)	0,568	0,753	H_0 diterima
Pendidikan (X_4)	5,189	0,158	H_0 diterima
Kepemilikan rumah (X_5)	$4,15 \times 10^{-30}$	1,000	H_0 diterima
Pekerjaan (X_6)	21,660	0,001	H_0 ditolak
Jenis usaha (X_9)	18,981	0,001	H_0 ditolak
Jenis kredit (X_{10})	5,368	0,068	H_0 diterima
Jenis bunga kredit (X_{11})	0,481	0,786	H_0 diterima

Kesimpulan:

Jadi berdasarkan Tabel 3, pada taraf signifikansi 5%, karena H_0 ditolak pada variabel jenis kelamin, pekerjaan, dan jenis usaha, maka penggolongan kredit debitur di PT BPR Gunung Lawu memiliki hubungan dengan ketiga variabel prediktor tersebut.

Tabel 5. Hasil Uji Independensi Variabel Prediktor Non Kategori

Variabel Prediktor (X_j)	Nilai Eta
Usia (X_1)	0,532
Pendapatan bersih (X_7)	0,678
Tenor (X_8)	0,266
Besar pinjaman (X_{12})	0,597

Berdasarkan Tabel 4, pendapatan bersih debitur memiliki hubungan yang kuat dengan penggolongan kredit di PT BPR Gunung Lawu. Usia debitur dan besar pinjaman debitur memiliki hubungan yang sedang dengan penggolongan kredit di PT BPR Gunung Lawu. Sedangkan variabel tenor menunjukkan tingkat hubungan yang lemah dengan penggolongan kredit debitur di PT BPR Gunung Lawu.

4.2.2. Estimasi dan Uji Signifikansi Parameter Regresi Logistik

Variabel prediktor yang digunakan dalam pemodelan adalah variabel prediktor kategori dan variabel prediktor non kategori yang memiliki hubungan dengan penggolongan kredit berdasarkan hasil uji independensi. Pemodelan dengan melakukan seleksi variabel prediktor yang berpengaruh signifikan terhadap penggolongan kredit di PT BPR Gunung Lawu menggunakan metode *stepwise*. Pemodelan regresi logistik dengan metode *stepwise* terdiri dari seleksi *forward* yang diikuti oleh eliminasi *backward*.

Tabel 6. Statistik G dan *P-value* pada Setiap Langkah Seleksi *Forward*

Langkah	Variabel Prediktor (X_j)	G	df	<i>P-value</i>
Langkah 0	Pekerjaan (X_6)	27,2	6	0,000
Langkah 1	Jenis Usaha (X_9)	16,95	4	0,002
Langkah 2	Jenis Kelamin (X_2)	7,363	1	0,007

Tabel 7. Statistik G dan *P-value* dari Setiap Langkah Eliminasi *Backward*

Langkah	Variabel Prediktor (X_j)	G	<i>P-value</i>
Langkah 0	Pekerjaan (X_6)	27,2	0,000
Langkah 1	Pekerjaan (X_6)	25,62	0,000
	Jenis Usaha (X_9)	16,95	0,000
Langkah 2	Pekerjaan (X_6)	26,45	0,000
	Jenis Usaha (X_9)	18,77	0,001
	Jenis Kelamin (X_2)	7,363	0,007

Tabel 8. Nilai Estimasi Parameter Model Regresi Logistik

Variabel Prediktor (X_j)	Nilai Estimasi	<i>Odds Ratio</i>	
<i>Intercept</i> ($\hat{\beta}_0$)	-2,568	0,077	
Jenis kelamin ($\hat{\beta}_2$)	-1,601	0,202	
			Perempuan (1)
Pekerjaan ($\hat{\beta}_6$)	2,795	16,363	
			Petani/pekebun (3)
			Buruh harian lepas (6)
Jenis usaha ($\hat{\beta}_9$)	2,657	14,254	
			Tidak ada (4)

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

dengan $g(x)$:

$$g(x) = -2,568 - 1,601 X_2(1) + 2,795 X_6(3) + 2,038 X_6(6) + 2,657 X_9(4)$$

4.2.3. Kesesuaian Model Regresi Logistik

Nilai *pseudo* R^2 atau R_L^2 yang dihasilkan model regresi logistik terbaik adalah sebesar 0,1865. Selain itu, nilai MSE yang dihasilkan oleh model adalah sebesar 0,1666. Ketepatan klasifikasi yang dihasilkan model tersebut relatif baik, yaitu sebesar 75,33%. Ukuran kesesuaian model regresi logistik yang dihasilkan berdasarkan nilai R_L^2 relatif rendah. Oleh karena itu, selanjutnya dilakukan analisis dengan memodelkan faktor-faktor yang mempengaruhi penggolongan kredit di PT BPR Gunung Lawu menggunakan algoritma genetika yang bertujuan untuk mendapatkan model dengan nilai R_L^2 yang lebih tinggi atau model dengan nilai MSE yang lebih rendah.

4.3. Pemodelan dengan Algoritma Genetika

Pada penelitian ini, ditetapkan 100 generasi dengan populasi berukuran 100. Salah satu kromosom dalam populasi awal adalah kromosom yang mengandung nilai estimasi parameter model yang telah diperoleh dari analisis regresi logistik. Seleksi yang digunakan adalah seleksi Roda *Roulette* (*Roulette Wheel selection*). Sedangkan operasi kawin silang yang digunakan menggunakan *crossover* satu titik (*single point crossover*) dengan probabilitas *crossover* 0,8 dan probabilitas mutasi acak 0,1. Selain itu, juga ditetapkan elitisme yang akan mempertahankan 5% kromosom dari ukuran populasi atau sebanyak 5 buah kromosom dengan nilai *fitness* optimum pada generasi berikutnya.

Tabel 9. Nilai *Fitness* R_L^2 dan MSE dari Setiap Percobaan

Percobaan	Nilai R_L^2	Nilai MSE
1	0,1865	0,1648
2	0,1865	0,1650
3	0,1865	0,1659
4	0,1865	0,1666
5	0,1865	0,1652

Dari lima percobaan algoritma genetika yang telah dilakukan, diperoleh rata-rata nilai R_L^2 sebesar 0,1865 dan rata-rata nilai MSE sebesar 0,1655. Kelima percobaan menghasilkan nilai R_L^2 yang sama dan pada percobaan pertama menghasilkan nilai MSE paling rendah, yaitu 0,1648 sehingga nilai estimasi parameter dari algoritma genetika berdasarkan nilai *fitness* R_L^2 tertinggi dan MSE terendah, yaitu pada percobaan pertama.

Tabel 10. Nilai Estimasi Parameter Algoritma Genetika

Variabel Prediktor (X_j)	Nilai Estimasi (berdasarkan R_L^2)	Nilai Estimasi (berdasarkan MSE)
<i>Intercept</i> ($\hat{\beta}_0$)	-2,568	-2,568
Jenis kelamin ($\hat{\beta}_2$)		
Perempuan (1)	-1,601	-1,601
Pekerjaan ($\hat{\beta}_6$)		
Petani/pekebun (3)	2,795	2,795
Buruh harian lepas (6)	2,038	2,124
Jenis usaha ($\hat{\beta}_9$)		
Tidak ada (4)	2,657	3,114

Kedua model regresi logistik berdasarkan algoritma genetika dapat dituliskan sebagai berikut.

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

dengan $g(x)$ untuk model regresi logistik berdasarkan $fitness R_L^2$:

$$g(x) = -2,568 - 1,601 X_2(1) + 2,795 X_6(3) + 2,038 X_6(6) + 2,657 X_9(4)$$

dan $g(x)$ untuk model regresi logistik berdasarkan $fitness$ MSE:

$$g(x) = -2,568 - 1,601 X_2(1) + 2,795 X_6(3) + 2,124 X_6(6) + 3,114 X_9(4)$$

4.4. Perbandingan Model Terbaik

Tabel 11. Perbandingan Nilai Estimasi Parameter Model

Variabel Prediktor (X_j)	Regresi Logistik	GA (berdasarkan R_L^2)	GA (berdasarkan MSE)
<i>Intercept</i> ($\hat{\beta}_0$)	-2,568	-2,568	-2,568
Jenis kelamin ($\hat{\beta}_2$)			
Perempuan (1)	-1,601	-1,601	-1,601
Pekerjaan ($\hat{\beta}_6$)			
Petani/pekebun (3)	2,795	2,795	2,795
Buruh harian lepas (6)	2,038	2,038	2,124
Jenis usaha ($\hat{\beta}_9$)			
Tidak ada (4)	2,657	2,657	3,114
R_L^2	0,1865	0,1865	0,1958
MSE	0,1666	0,1666	0,1648
Ketepatan klasifikasi	0,7533	0,7533	0,7533

Berdasarkan Tabel 11, pemodelan dengan algoritma genetika berdasarkan $fitness$ MSE menghasilkan model terbaik dengan nilai R_L^2 tertinggi dan nilai MSE terendah, dengan ketepatan klasifikasi 75,33%. Sehingga model $\pi(x) = \frac{e^{g(x)}}{1+e^{g(x)}}$ dengan $g(x) = -2,568 - 1,601 X_2(1) + 2,795 X_6(3) + 2,124 X_6(6) + 3,114 X_9(4)$ dapat digunakan oleh PT BPR Gunung Lawu untuk memprediksi risiko kredit calon debitur.

Berdasarkan model yang diperoleh, dapat diketahui peluang terbesar untuk digolongkan sebagai debitur dengan kredit bermasalah adalah debitur laki-laki yang berprofesi sebagai petani/pekebun, dan tidak memiliki usaha yang sedang dijalankan, yaitu sebesar 0,947. Peluang terkecil untuk digolongkan sebagai debitur dengan kredit bermasalah adalah debitur perempuan yang berprofesi sebagai PNS, karyawan swasta, wiraswasta, pedagang, atau sopir (selain petani/pekebun dan buruh harian lepas), dan sedang menjalankan usaha di bidang industri, pertanian/perkebunan, perdagangan, atau jasa, yaitu sebesar 0,015.

5. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah dipaparkan, dapat diambil kesimpulan sebagai berikut:

1. Pemodelan dengan regresi logistik menghasilkan model terbaik dengan empat faktor yang mempengaruhi penggolongan kredit di PT BPR Gunung Lawu, yaitu debitur berjenis kelamin perempuan, pekerjaan petani/pekebun, pekerjaan buruh harian lepas, dan tidak adanya jenis usaha yang sedang dijalankan.
2. Pemodelan dengan metode regresi logistik menghasilkan model dengan nilai R_L^2 yang relatif rendah, yaitu 0,1865 dengan nilai MSE sebesar 0,1666, serta ketepatan klasifikasi sebesar 75,33%.
3. Model terbaik dihasilkan oleh pemodelan dengan algoritma genetika berdasarkan $fitness$ MSE. Model menghasilkan nilai R_L^2 tertinggi sebesar 0,1958 dan nilai MSE terendah sebesar 0,1648 dengan ketepatan klasifikasi sebesar 75,33% sehingga model

$\pi(x) = \frac{e^{g(x)}}{1+e^{g(x)}}$ dengan $g(x) = -2,568 - 1,601 X_2(1) + 2,795 X_6(3) + 2,124 X_6(6) + 3,114 X_9(4)$ dapat digunakan oleh PT BPR Gunung Lawu untuk memprediksi risiko kredit calon debitur.

DAFTAR PUSTAKA

- Agresti, A. 1990. *Categorical Data Analysis*. New York: John Wiley & Sons.
- Desiani, A dan Arhami, M. 2006. *Konsep Kecerdasan Buatan*. Yogyakarta: Andi Offset.
- Haupt, S.E. dan Haupt, R.L. 2004. *Practical Genetic Algorithms*. New Jersey: John Wiley & Sons Inc.
- Hosmer, D.W. dan Lemeshow, S. 1989. *Applied Logistic Regression*. USA: John Wiley and Sons.
- Kusumadewi, S. 2003. *Artificial Intelligence (Teknik dan Aplikasinya)*. Yogyakarta: Graha Ilmu.
- Rencher, A.C., 2002. *Methods of Multivariate Analysis*. USA : John Wiley and Sons.