

**ANALISIS KLASIFIKASI MENGGUNAKAN METODE REGRESI LOGISTIK  
BINER DAN BOOTSTRAP AGGREGATING CLASSIFICATION AND  
REGRESSION TREES (BAGGING CART)  
(Studi Kasus: Nasabah Koperasi Simpan Pinjam Dan Pembiayaan Syariah (KSPPS))**

**Salma Innassuraiya<sup>1\*</sup>, Tatik Widiharah<sup>2</sup>, Iut Tri Utami<sup>3</sup>**

<sup>1,2,3</sup>Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

\*email: innassuraiya@gmail.com

**ABSTRACT**

*The Save Loan and Sharia Financing Cooperatives (KSPPS) is a financial institution that offers deposits, loans, and financing to its members while adhering to Islamic sharia rules. Customers payment behaviour is influenced by their background differences, such as age, gender, occupation, and so on. The classification method is used to determine the characteristics of members who are currently in arrears or are stuck in arrears. Binary Logistic Regression and Bootstrap Aggregating Classification and Regression Trees were utilized as classification methods (BAGGING CART). A Logistic Regression with binary response variables is known as a Binary Logistic Regression. By resampling 50 times, the technique with the BAGGING process is used to improve the performance of the classification using CART. Customer data from one of the KSPPS in Central Java in 2021 was used in this investigation. Gender, age, marital status, employment, education level, time period, and income were the independent variables in this study, whereas payment status was the dependent variable (not stuck and stuck). The Binary Logistic Regression approach had an accuracy of 78.67 percent with an APER 21.33 percent, a Press's Q of 24.65, and a specificity of 98.30 percent, according to the classification accuracy statistics. The accuracy of the classification produced by CART with an accuracy value of 77.33 percent with an APER 22.67 percent, the value of Press's Q is 22,413, and specificity is 94.91 percent, then approached by BAGGING process the accuracy of the resulting classification by predicting data testing accuracy value of 78.67 percent with an APER 21.33 percent, press's Q value of 24.65, and specificity of 96.61 percent. Based on these findings, it can be inferred that using the BAGGING process can increase the CART method's performance to the point where it is nearly as good as Binary Logistic Regression, which has a slightly higher classification accuracy.*

**Keyword:** KSPPS, Binary Logistic Regression, CART, BAGGING CART, Accuracy, APER, Press's Q, Specificity.

## 1. PENDAHULUAN

Koperasi Simpan Pinjam dan Pembiayaan Syariah (KSPPS) adalah koperasi yang kegiatan usahanya meliputi simpanan, pinjaman dan pembiayaan sesuai prinsip syariah, termasuk mengelola zakat, infaq atau sedekah, dan wakaf. Setiap anggota koperasi memiliki berbagai latar belakang yang berbeda, hal tersebut dapat mempengaruhi anggota dalam memenuhi kewajibannya. Perbedaan latar belakang yang ada seperti usia, jenis kelamin, pekerjaan, dan sebagainya dapat digunakan dalam melakukan klasifikasi untuk mengetahui karakteristik anggota dengan status pembayaran lancar atau status pembayaran macet.

Klasifikasi merupakan salah satu metode statistika yang mengelompokkan data sesuai karakteristiknya secara sistematis dalam kelas yang telah ditentukan. Klasifikasi merupakan metode penelitian ketika suatu objek data dinilai untuk dimasukkan ke dalam suatu kelas tertentu dari sejumlah kelas yang ada atau tersedia (Prasetyo, 2012).

Metode klasifikasi yang digunakan adalah Regresi Logistik Biner dan *Classification And Regression Trees* (CART), namun hasil pohon klasifikasi cenderung tidak stabil karena setiap perubahan kecil pada data *learning* dapat menghasilkan perubahan besar dalam struktur pohon keputusan yang dapat menyampaikan hasil berbeda sehingga mempengaruhi

hasil akurasi prediksi. Metode *Bootstrap Aggregating* (BAGGING) dapat digunakan untuk memperbaiki stabilitas dan kekuatan akurasi pada CART.

## 2. TINJAUAN PUSTAKA

Pengertian Koperasi Simpan Pinjam dan Pembiayaan Syariah menurut Peraturan Menteri Koperasi dan Usaha Kecil dan Menengah Republik Indonesia No. 16 /Per/M.KUKM/IX/2015, Koperasi Simpan Pinjam dan Pembiayaan Syariah selanjutnya dalam peraturan ini disebut KSPPS adalah koperasi yang kegiatan usahanya meliputi simpanan, pinjaman dan pembiayaan sesuai prinsip syariah, termasuk mengelola zakat, infaq atau sedekah, dan wakaf. Modal sendiri KSPPS adalah jumlah simpanan pokok, simpanan wajib, cadangan yang disisihkan dari sisa hasil usaha, hibah, dan simpanan lain yang memiliki karakteristik sama dengan simpanan wajib.

Metode yang menghubungkan variabel dependen yang bersifat karegorik dengan variabel independen. Menurut Agresti (2007), pada pemodelan statistik variabel respons biner, di mana hasil respons untuk setiap subjek adalah "kesuksesan" atau "kegagalan". Bentuk model regresi logistiknya sebagai berikut:

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

dengan fungsi logitnya yaitu:

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

sehingga

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

Menurut Agresti (2007), untuk mengestimasi parameter dalam logistik biner dapat menggunakan metode penduga maksimum likelihood. Menurut Hosmer dan Lemeshow (2000), misalkan terdapat sampel dari  $n$  observasi independen  $(x_i, y_i)$ ,  $i = 1, 2, \dots, n$  dengan  $y_i$  dinotasikan dari variabel respon biner dan  $x_i$  adalah nilai dari variabel prediktor untuk subjek ke- $i$ , maka fungsi likelihoodnya merupakan perkalian masing-masing fungsi densitasnya

### 1. Uji Rasio Likelihood (Uji Serentak)

Menurut Hosmer dan Lemeshow (2000), Uji Rasio Likelihood bertujuan untuk mengetahui signifikansi koefisien  $\beta$  terhadap variabel dependen secara bersama-sama.

Hipotesis:

$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$  (Seluruh variabel independen tidak ada yang berpengaruh terhadap variabel dependen)

$H_1: \text{Paling sedikit ada satu } \beta_j \neq 0 \text{ dengan } j = 1, 2, \dots, p$  (Paling sedikit ada satu variabel independen yang berpengaruh terhadap variabel dependen)

Statistik Uji:

$$G = -2 \ln \left[ \frac{\text{likelihood tanpa variabel independen}}{\text{likelihood dengan variabel independen}} \right] \quad (4)$$

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

### 2. Uji Wald (Uji Parsial)

Menurut Hosmer dan Lemeshow (2000), Uji Wald bertujuan untuk mengetahui pengaruh setiap koefisien  $\beta_j$  secara individual dengan cara membandingkan terhadap standar errornya sehingga dapat diketahui apakah variabel independen didalam model berpengaruh secara signifikan terhadap variabel dependen.

Hipotesis:

$H_0: \beta_j = 0$  dengan  $j = 1, 2, \dots, p$  (Tidak ada pengaruh antara variabel independen ke- $j$  dengan variabel dependen)

$H_1: \beta_j \neq 0$  dengan  $j = 1, 2, \dots, p$  (Ada pengaruh antara variabel independen ke- $j$  dengan variabel dependen)

Statistik Uji:

$$W_j = \left\{ \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \right\}^2 \quad (5)$$

Pada taraf signifikansi  $\alpha$ ,  $H_0$  ditolak jika  $W_j > \chi_{(\alpha,1)}^2$  atau nilai  $p - value < \alpha$

### 3. Uji Hosmer dan Lemeshow (Uji Kesesuaian Model)

Menurut Hosmer dan Lemeshow (2000), mengatakan bahwa uji kesesuaian model merupakan uji yang dilakukan untuk mengetahui apakah ada perbedaan antara prediksi dan hasil observasi (model sesuai atau tidak).

Hipotesis:

$H_0$ : Model sesuai (Tidak ada perbedaan antara prediksi dengan hasil observasi)

$H_1$ : Model tidak sesuai (Ada perbedaan antara prediksi dengan hasil observasi)

Statistik Uji:

$$\hat{C} = \sum_{k=1}^g \frac{(o_k - n'_k \bar{\pi}_k)^2}{(n'_k \bar{\pi}_k)(1 - \bar{\pi}_k)} \quad (6)$$

Pada taraf signifikansi  $\alpha$ ,  $H_0$  ditolak jika  $\hat{C} > \chi_{(\alpha, g-2)}^2$  atau nilai  $p - value < \alpha$

*Classification and Regression Tree* atau yang dikenal dengan CART merupakan salah satu metode non parametrik yang digunakan untuk analisis klasifikasi yaitu dengan teknik pohon keputusan baik untuk variabel respon kategorik maupun kontinu. Apabila variabel respon berskala kontinu maka CART menghasilkan pohon regresi (*regression trees*), sedangkan jika variabel respon berskala kategorik maka CART menghasilkan pohon klasifikasi (*classification trees*). Pembentukan sebuah pohon klasifikasi CART pada dasarnya bekerja dalam empat langkah utama yaitu:

Proses pemilahan pada masing-masing simpul induk didasarkan pada *goodness of split criterion* (kriteria pemilihan terbaik). Kriteria pemilahan terbaik ini dibentuk berdasarkan fungsi *impurity* (fungsi keheterogenan) yaitu untuk mengukur tingkat keheterogenan suatu kelas dari suatu simpul tertentu dalam pohon klasifikasi. Fungsi keheterogenan yang digunakan adalah indeks Gini, yaitu:

$$i(t) = \sum_{j \neq i} p(j|t)p(i|t) \quad (7)$$

Proses pemecahan pada masing-masing simpul induk didasarkan pada aturan *goodness of split* (kriteria pemecahan terbaik). *Goodness of split* merupakan suatu evaluasi pemilahan oleh pemilah  $s$  pada simpul  $t$ , dirumuskan sebagai berikut:

$$\Delta_i(s, t) = i(t) - P_L i(t_L) - P_R i(t_R) \quad (8)$$

Pengembangan pohon dilakukan dengan mencari semua kemungkinan pemilah pada simpul  $t_1$  sehingga ditemukan pemilah  $s^*$  yang memberikan nilai penurunan keheterogenan tertinggi yaitu:

$$\Delta_i(s^*, t) = \max_j \Delta_i(s, t_1) \quad (9)$$

Kemudian  $t_1$  dipilah menjadi  $t_2$  dan  $t_3$  menggunakan pemilah  $s^*$  dan dengan prosedur yang sama diulangi pada simpul  $t_2$  dan  $t_3$  secara terpisah dan kemudian pada simpul-simpul selanjutnya sampai terbentuk pohon klasifikasi maksimal.

Pelabelan kelas dilakukan mulai dari awal pemilihan simpul hingga simpul akhir terbentuk, karena setiap simpul yang dibentuk memiliki kesempatan menjadi simpul akhir. Pelabelan tiap simpul akhir berdasarkan aturan jumlah anggota kelas terbanyak yaitu jika:

$$P(j_0|t) = \max_j P(j|t) = \max_j \frac{N_j(t)}{N(t)} \quad (10)$$

Label kelas untuk simpul terminal  $t$  adalah  $j_0$  yang memberikan nilai dugaan kesalahan pengklasifikasian pada simpul  $t$  terbesar.

Menurut Lewis (2000), suatu simpul  $t$  akan menjadi simpul akhir atau tidak akan dipilah kembali apabila hanya terdapat satu pengamatan dalam tiap simpul anak, semua pengamatan dalam tiap simpul anak memiliki distribusi variabel respon yang identik, dan adanya batasan jumlah kedalaman pohon maksimal yang ditentukan oleh peneliti.

Pohon klasifikasi maksimal yang terbentuk dimungkinkan berukuran sangat besar. Semakin banyak pemilahan yang dilakukan maka tingkat akurasi semakin tinggi, tetapi dengan ukuran yang sangat besar akan sulit dipahami sehingga menyebabkan *overfitting* (pencocokan nilai yang sangat kompleks) untuk data baru. Menurut Lewis (2000), masalah tersebut diatasi dengan pemangkasan pada pohon klasifikasi maksimal untuk mendapatkan pohon klasifikasi yang optimal.

Menurut Lewis (2000), untuk menghasilkan urutan pohon yang lebih sederhana, yang masing-masing adalah kandidat untuk pohon akhir yang sesuai, metode pemangkasan “*cost-complexity*” digunakan. Ukuran pemangkasan yang digunakan untuk memperoleh ukuran pohon yang layak adalah *Cost Complexity Minimum*, yaitu:

$$R(T) = \sum_{t \in \bar{T}} r(t)p(t) = \sum_{t \in \bar{T}} R(t) \quad (11)$$

dimana

$$r(t) = 1 - \max_j p(j|t) \quad (12)$$

Langkah awal proses pemangkasan dilakukan terhadap  $T_t$  yaitu subpohon dari  $T_{max}$  dengan mengambil  $t_L$  yang merupakan simpul anak kiri dan  $t_R$  yang merupakan simpul anak kanan hasil dari pemilihan simpul induk  $t$ , sehingga  $R(t_L)$  dan  $R(t_R)$  dapat dihitung menggunakan persamaan (11). Jika diperoleh dua simpul anak dan simpul induk yang memenuhi persamaan

$$R(t) = R(t_L) + R(t_R) \quad (13)$$

maka simpul anak  $t_L$  dan  $t_R$  dipangkas. Proses ini diulangi sampai tidak ada lagi pemangkasan yang mungkin.

Menurut Sutton (2005), *Bootstrap Aggregation* atau BAGGING adalah teknik yang diusulkan oleh Breiman (1996) yang dapat digunakan dengan banyak metode klasifikasi dan metode regresi untuk mereduksi variansi yang terkait dengan prediksi, dan dengan demikian meningkatkan proses prediksi. Terdapat dua tahapan, yaitu *bootstrap* yang tidak lain adalah pengambilan sampel dari data learning yang dimiliki (*resampling*) dan *aggregating* yaitu menggabungkan banyak nilai dugaan menjadi satu dugaan. Tahapan pembuatan dugaan secara BAGGING menggunakan pohon menurut Sutton (2005) sebagai berikut:

1. Tahapan *bootstrap*
  - a. Tarik sampel acak dengan pemulihan berukuran  $n$  dari gugus data *learning*.
  - b. Susun pohon terbaik berdasarkan data tersebut.
  - c. Ulangi langkah a-b sebanyak  $B$  kali sehingga diperoleh  $B$  buah pohon klasifikasi.
2. Tahapan *aggregating*

Lakukan pendugaan gabungan berdasarkan  $B$  buah pohon klasifikasi tersebut menggunakan aturan *majority vote* (suara terbanyak).

*Apparent Error Rate* (APER) merupakan suatu ukuran yang digunakan untuk menyatakan nilai proporsi sampel yang salah diklasifikasikan dalam proses klasifikasi. APER dapat dihitung dengan mudah yaitu dengan menggunakan matriks konfusi yang menyajikan anggota observasi dengan anggota prediksi. Misalkan suatu variabel respon memiliki 2 kategori yaitu  $Y= 1, 2$ . Jika  $n_1$  merupakan jumlah anggota observasi yang diklasifikasikan  $Y=1$  dan  $n_2$  merupakan jumlah anggota observasi yang diklasifikasikan  $Y=2$ , maka bentuk matriks konfusinya sebagai berikut:

**Tabel 1. Matriks Konfusi**

Observasi	Prediksi	
	Kelas 1	Kelas 2
Kelas 1	$n_{11}$	$n_{12}$
Kelas 2	$n_{21}$	$n_{22}$

$$APER = \frac{n_{12} + n_{21}}{n_1 + n_2} \quad (14)$$

Kerangka kerja statistik uji *Press's Q* yaitu dengan membandingkan antara banyaknya pengamatan yang diklasifikasi secara tepat dengan total pengamatan dan jumlah kelompok. Nilai yang diperoleh dari perhitungan kemudian dibandingkan dengan nilai kritis (*critical value*) yang diambil dari tabel *Chi-Square* dengan derajat kebebasan 1 dan tingkat keyakinan sesuai dengan yang ditetapkan. Statistik uji dari *Press's Q* dapat dirumuskan sebagai berikut :

$$Press's Q = \frac{[N - (rC)]^2}{N(C - 1)} \quad (15)$$

Proses klasifikasi akan dikatakan akurat secara statistik apabila nilai *Press's Q* lebih besar dari pada nilai kritis yang diambil dari tabel *Chi-Square* dengan derajat bebas bernilai 1 dan tingkat keyakinan sebesar  $\alpha$  ( $Press's Q > \chi^2_{\alpha,1}$ )

*Specificity* merupakan persentase kebenaran memprediksi data negatif dibandingkan dengan keseluruhan data negatif. *Specificity* adalah derajat keandalan model untuk mendeteksi data berlabel negatif dengan benar. *Specificity* dapat dipahami sebagai sensitivitas namun dalam konteks data berlabel negatif, pada penelitian ini kategori tidak macet yang bernilai 0 berlabel negatif sedangkan kategori macet yang bernilai 1 berlabel positif. Nilai *specificity* dapat dihitung dengan rumus berikut:

$$Specificity = \frac{TN}{TN + FP} \quad (16)$$

Proses klasifikasi akan dikatakan akurat secara statistik apabila nilai *Press's Q* lebih besar dari pada nilai kritis yang diambil dari tabel *Chi-Square* dengan derajat bebas bernilai 1 dan tingkat keyakinan sebesar  $\alpha$  ( $Press's Q > \chi^2_{\alpha,1}$ )

### 3. METODOLOGI PENELITIAN

Jenis data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari data nasabah pada salah satu Koperasi Simpan Pinjam dan Pembiayaan Syariah (KSPPS) di regional Provinsi Jawa Tengah pada tahun 2021.

Variabel yang digunakan dalam penelitian ini terdiri atas status pembayaran (Y), jenis kelamin ( $X_1$ ), usia ( $X_2$ ), status perkawinan ( $X_3$ ), pekerjaan ( $X_4$ ), jenjang pendidikan ( $X_5$ ), jangka waktu ( $X_6$ ), dan pendapatan ( $X_7$ ).

*Software* yang digunakan untuk pengolahan data dalam penelitian ini adalah Program R versi 4.0.4. Tahapan analisis yang dilakukan pada penelitian ini yaitu:

1. Membagi data menjadi data latih dan data uji. Perbandingan data latih dan data uji yang dipilih yaitu 80% : 20%.
2. Menentukan faktor-faktor yang mempengaruhi status pembayaran nasabah Koperasi Simpan Pinjam dan Pembiayaan Syariah (KSPPS) menggunakan metode Regresi Logistik Biner.
  - a. Membentuk model Regresi Logistik Biner.
  - b. Melakukan uji serentak dengan uji Rasio Likelihood.
  - c. Melakukan uji parsial dengan uji Wald.
  - d. Melakukan uji Hosmer dan Lemeshow.
  - e. Menentukan model akhir Regresi Logistik Biner.
  - f. Menghitung nilai  $\pi(x_i)$ ,  $\pi(x_i)$  adalah peluang kredit macet. Apabila nilai  $\pi(x_i) < 0,5$  maka masuk kedalam kelas 0 dan dikategorikan sebagai kredit tidak macet/ lancar.  
Apabila nilai  $\pi(x_i) \geq 0,5$  maka masuk kedalam kelas 1 dan dikategorikan sebagai kredit macet.
  - g. Membentuk matriks konfusi dan menghitung ketepatan klasifikasi menggunakan APER dan *Press's Q*.
3. Melakukan klasifikasi dengan menggunakan metode *Bootstrap Aggregating Classification and Regression Trees* (BAGGING CART).
  - a. Melakukan *resampling* menggunakan data latih.
  - b. Melakukan pembuatan pohon klasifikasi dengan algoritma CART dengan tahap-tahap Pemilahan simpul (*splitting nodes*), Pelabelan kelas (*class assignment*), Penghentian pembentukan pohon klasifikasi (*stop the splitting*), Pemangkasan pohon, jika dinilai pohon klasifikasi yang dihasilkan terlalu kompleks.
  - c. Pendugaan menggunakan data uji.
  - d. Proses BAGGING (*bootstrap aggregating*).
    1. Ulangi langkah b dan c sebanyak 50 kali, sehingga memperoleh pohon klasifikasi sebanyak 50 buah.
    2. Melakukan pendugaan gabungan berdasarkan hasil dari 50 buah pohon klasifikasi yang ada menggunakan aturan suara terbanyak (*majority vote*).
  - e. Menghitung ketepatan klasifikasi hasil dari pendugaan gabungan menggunakan APER dan *Press's Q*.
4. Membandingkan ketepatan klasifikasi metode Regresi Logistik Biner dan *Bootstrap Aggregating Classification and Regression Trees* (BAGGING CART).
5. Melakukan interpretasi hasil klasifikasi menggunakan metode Regresi Logistik Biner dan *Bootstrap Aggregating Classification and Regression Trees* (BAGGING CART).

#### 4. HASIL DAN PEMBAHASAN

Jumlah data yang digunakan adalah sebanyak 375 data nasabah yang terdiri dari 316 orang atau sebesar 84,27% tidak mengalami kredit macet dan 59 orang atau sebesar 15,73% mengalami kredit macet.

Model awal Regresi Logistik Biner disusun dengan menggunakan estimasi parameter. Model awal Regresi Logistik Biner yang terbentuk yaitu  $\pi(x_i) = \frac{e^{g(x_i)}}{1+e^{g(x_i)}}$  dengan  $g(x_i) = -1,59679 - 0.65092x_1(2) + 0.74834x_2(2) + 2.4308x_2(3) - 14.01046x_3(2) - 0.59205x_4(2) - 1.01883x_4(3) - 0.74421x_4(4) + 0.02175x_4(5) + 0.25717x_5(2) + 0.28091x_5(3) + 1.65824x_5(4) - 0.03576x_5(5) + 0.49024x_6(2) + 0.62740x_6(3) - 1.98802x_7(2) - 3.02501x_7(3)$

Hipotesis uji rasio Likelihood:

$H_0: \beta_1 = \beta_2 = \dots = \beta_{16} = 0$  (Seluruh variabel independen tidak ada yang berpengaruh terhadap status pembayaran)

$H_1: \text{Paling sedikit ada satu } \beta_j \neq 0 \text{ dengan } j = 1, 2, \dots, 16$  (Paling sedikit ada satu variabel independen yang berpengaruh terhadap variabel status kredit)

Statistik Uji:

$$G = -2 \ln \left[ \frac{\text{likelihood tanpa variabel independen}}{\text{likelihood dengan variabel independen}} \right] = 60.3634324$$

Pada taraf signifikansi  $\alpha = 5\%$ ,  $H_0$  ditolak karena  $G = 60.3634324 > \chi^2_{(16,0.05)} = 26.29623$ . Sehingga dapat disimpulkan bahwa paling sedikit ada satu variabel independen yang berpengaruh terhadap variabel status pembayaran.

Hipotesis Uji Wald:

$H_0: \beta_j = 0 \text{ dengan } j = 1, 2, \dots, 16$  (Tidak ada pengaruh antara variabel independen ke-j dengan variabel status pembayaran)

$H_1: \beta_j \neq 0 \text{ dengan } j = 1, 2, \dots, 16$  (Ada pengaruh antara variabel independen ke-j dengan variabel status pembayaran)

Statistik Uji:

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

dapat dibentuk tabel hasil uji Wald sebagai berikut:

**Tabel 2.** Hasil Uji Wald Model Awal

Variabel	Estimate	Std.Error	Nilai z	p-value	Keterangan
$x_1(2)$	-0.65092	0.40342	-1.614	0.10663	Tidak signifikan
$x_2(2)$	0.74834	0.73558	1.017	0.30899	Tidak signifikan
$x_2(3)$	2.43083	0.75783	3.208	0.00134	Signifikan
$x_3(2)$	-14.01046	1356.42475	-0.010	0.99176	Tidak signifikan
$x_4(2)$	-0.59205	1.59254	-0.372	0.71007	Tidak signifikan
$x_4(3)$	-1.01883	1.13546	-0.897	0.36957	Tidak signifikan
$x_4(4)$	-0.74421	0.75805	-0.982	0.32623	Tidak signifikan
$x_4(5)$	0.02175	0.77952	0.028	0.97774	Tidak signifikan
$x_5(2)$	0.25717	0.50082	0.513	0.60761	Tidak signifikan
$x_5(3)$	0.28091	0.56613	0.496	0.61975	Tidak signifikan
$x_5(4)$	1.65824	4182.25452	0.000	0.99968	Tidak signifikan
$x_5(5)$	-0.03576	1.33268	-0.027	0.97859	Tidak signifikan

$x_6(2)$	0.49024	0.51295	0.956	0.33922	Tidak signifikan
$x_6(3)$	0.62740	0.89059	0.704	0.48114	Tidak signifikan
$x_7(2)$	-1.98802	0.69734	-2.851	0.00436	Signifikan
$x_7(3)$	-3.02501	1.61408	-1.874	0.06091	Tidak signifikan
constant	-1.59679	0.99955	-1.598	0.11015	-

Pada taraf signifikansi  $\alpha = 5\%$   $H_0$  ditolak jika  $W_j > \chi^2_{(0.05,1)} = 3.841459$  atau  $H_0$  ditolak jika nilai  $p - value < \alpha$ , variabel independen yang berpengaruh secara signifikan terhadap status kredit yaitu variabel  $x_2(3) = \text{Usia } (>50 \text{ tahun})$  dan variabel  $x_7(2) = \text{Pendapatan (Rp3.000-000,00-4.000.000,00)}$ .

Model kedua Regresi Logistik Biner yang terbentuk yaitu  $\pi(x_i) = \frac{e^{g(x_i)}}{1+e^{g(x_i)}}$  sebagai peluang kredit macet dengan  $g(x_i) = -1,4222 + 0.8088x_2(2) + 2.4315x_2(3) - 2.0962x_7(2) - 3.4394x_7(3)$ .

Ketepatan klasifikasi dengan Regresi Logistik Biner menghasilkan matriks konfusi sebagai berikut:

**Tabel 3.** Matriks Konfusi Regresi Logistik Biner

Observasi	Prediksi	
	Kredit tidak macet (0)	Kredit macet (1)
Kredit tidak macet (0)	58	15
Kredit macet (1)	1	1

$$APER = \frac{15 + 1}{58 + 15 + 1 + 1} = 0.2133$$

$$Akurasi = 1 - APER = 1 - 0.2 = 0.7867$$

$$Press's Q = \frac{[75 - (59(2))]^2}{75(2 - 1)} = 24.65$$

$$Specificity = \frac{58}{58 + 1} = 0.9830$$

Besarnya ketepatan klasifikasi tersebut menunjukkan bahwa data uji berjumlah 75 terklasifikasikan secara benar sebanyak 78.67% dan terklasifikasikan secara salah sebanyak 21.33%. Nilai *Press's Q* lebih besar dari nilai kritis  $\chi^2_{\alpha,1}$  yaitu 3.841, sehingga dapat disimpulkan bahwa dengan tingkat keyakinan 95% metode Regresi Logistik Biner akurat secara statistik dalam mengklasifikasikan nasabah Koperasi Simpan Pinjam dan Pembiayaan Syariah (KSPPS). Nilai *specificity* sebesar 98.30% yang merupakan nilai ketepatan kelas tidak macet yang bernilai 0 atau berlabel negatif diprediksi secara benar. Nilai indeks gini pada masing-masing simpul disajikan dalam tabel berikut:

**Tabel 4.** Nilai Indeks Gini

Simpul	$i(t)$	$i(t_L)$	$i(t_R)$
1	0.3397	0.1830	0.4963
2	0.2796	0.1332	0.4260
3	0.2489	0.1511	0.3466
4	0.1259	0	0.2519

5	0.3492	0.2056	0.4928
6	0.2276	0.1975	0.2577
7	0.2447	0.2291	0.2604

sehingga nilai *goodness of split* untuk setiap calon simpul dapat dilihat pada tabel berikut:

**Tabel 5.** Nilai Goodness of Split

Simpul	$\Delta_i(s, t)$
1	-0.0006
2	0.0576
3	-0.1184
4	0.1196
5	0.0130
6	-0.0174
7	0.1201

Berdasarkan perhitungan nilai *goodness of split* pada **Tabel 5** di atas untuk setiap calon simpul yang tertinggi adalah calon simpul 7 sebesar 0.1201 yaitu variabel pendapatan dengan simpul kiri (Rp3.000.000,00-4.000.000,00 dan >Rp3.000.000,00) dan simpul kanan (<Rp3.000.000,00), sehingga calon simpul 7 ini akan menjadi *parent node*. Proses pemilihan akan terus berlanjut secara rekursif dengan cara yang sama pada *node* atau simpul yang lain sampai pada simpul akhir atau *terminal node*.

Penentuan label kelas pada simpul terminal berdasarkan jumlah terbanyak, sebagai contoh:

$$P(\text{Tidak Macet}|\text{simpul 1}) = \frac{257}{300} = 0.857$$

$$P(\text{Macet}|\text{simpul 1}) = \frac{43}{300} = 0.143$$

Simpul yang memiliki nilai probabilitas tidak macet lebih tinggi daripada probabilitas macet maka diberi label kelas 0, jika sebaliknya maka diberi label kelas 1. Tabel penandaan label kelas sebagai berikut:

**Tabel 6.** Penandaan Label Kelas

Simpul	Label Kelas
1	0
2	0
3	0
6	0
7	1
12	0
13	1
14	0
15	1
24	0
25	0
26	0

27	1
50	0
51	0
102	0
103	1

Berdasarkan **Tabel 13** di atas dapat dijelaskan bahwa simpul dengan label kelas tidak macet/ lancar berturut-turut adalah 1,2,3,6,12,14,24,25,26,50,51, dan 102. Sedangkan untuk simpul dengan label kelas macet berturut-turut adalah 7,13,15,27, dan 103.

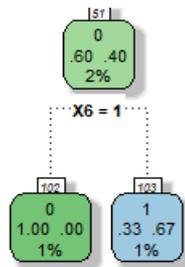
Penghentian kelas dilakukan pada pohon klasifikasi yang telah terbentuk karena pada simpul 14, 24, 26, 27, dan 102 hanya terdapat satu pengamatan dalam tiap simpulnya. Sedangkan pada simpul 2, 15, 25 terdapat variabel respon yang identik, dan pada simpul 103 terdapat 3 amatan pada kedalaman maksimal yaitu kedalaman enam sehingga proses pemilahan simpul dihentikan. Berikut simpul-simpul yang mengalami proses penghentian kelas:



**Gambar 1.** Simpul-Simpul Akhir pada Pohon Keputusan

Proses pemangkasan pohon klasifikasi maksimal dimulai dengan mengambil  $t_L$  yang merupakan simpul kiri dan  $t_R$  yang merupakan simpul kanan dari  $T_{max}$  yang dihasilkan dari simpul induk  $t$ . Jika diperoleh dua simpul anak dan simpul induk yang memenuhi persamaan  $R(t) = R(t_L) + R(t_R)$ , maka simpul anak  $t_L$  dan  $t_R$  dipangkas. Proses tersebut diulang sampai tidak ada lagi pemangkasan yang mungkin dilakukan. Contoh proses pemangkasan dilakukan dengan cara sebagai berikut:

Menghitung nilai  $R$  (simpul 51) sebagai simpul induk



**Gambar 2.** Simpul 51 pada Pohon Keputusan

$$r(\text{simpul } 51) = 1 - \max_j p(j|t) = 1 - 0.6 = 0.4$$

$$P(\text{simpul } 51) = \frac{5}{300} = 0.0167$$

$$R(\text{simpul } 51) = r(\text{simpul } 51) * P(\text{simpul } 51) = 0.4 * 0.0167 = 0.0067$$

Selanjutnya dihitung nilai  $R(t_L)$  dan  $R(t_R)$  pada simpul anak, yaitu simpul 102 dan simpul 103.

$$r(\text{simpul } 102) = 1 - \max_j p(j|t) = 1 - 1 = 0$$

$$P(\text{simpul } 102) = \frac{2}{300} = 0.0067$$

$$R(\text{simpul } 102) = r(\text{simpul } 102) * P(\text{simpul } 102) = 0 * 0.0067 = 0$$

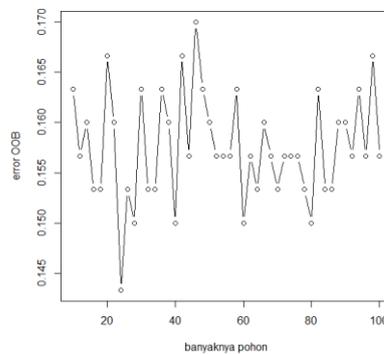
$$r(\text{simpul } 103) = 1 - \max_j p(j|t) = 1 - 0.667 = 0.333$$

$$P(\text{simpul } 103) = \frac{3}{300} = 0.01$$

$$R(\text{simpul } 103) = r(\text{simpul } 103) * P(\text{simpul } 103) = 0.333 * 0.01 = 0.003$$

Dengan demikian  $R(\text{simpul } 102) + R(\text{simpul } 103) \neq R(\text{simpul } 51)$  yaitu  $0 + 0.003 \neq 0.0067$  sehingga simpul 102 dan simpul 103 tidak dipangkas.

Proses perbaikan pada *Classification And Regression Tree* (CART) dilakukan dengan cara *bootstrap* yaitu melakukan *resampling* pada data *training* sebanyak 50 kali secara berulang dengan pengembalian dari sampel asli. Setelah didapatkan pohon klasifikasi sebanyak 50 buah pohon dilakukan pengambilan suara terbanyak (*majority vote*) untuk menggabungkan nilai dugaan dari beberapa pohon menjadi satu dugaan akhir. Penentuan proses *resampling* sebanyak 50 kali didasarkan pada nilai amatan yang tidak terambil atau *out of bag* (OOB) seperti pada gambar berikut:



**Gambar 3.** Nilai Error Out of Bag (OOB)

**Gambar 3** mengilustrasikan bagaimana nilai *error* OOB untuk banyaknya pohon 10 hingga 100, pada angka sekitar 50 nilai OOB sudah relatif stabil. Tingkat kesalahan prediksi didasarkan pada amatan yang tidak terpilih atau OOB saat proses *bootstrap*. Amatan yang tidak terpilih diprediksi dan dilakukan proses *aggregating* atau pengambilan suara terbanyak berdasarkan 50 pohon klasifikasi yang telah terbentuk.

Ketepatan klasifikasi dengan BAGGING CART menghasilkan matriks konfusi sebagai berikut:

**Tabel 7.** Matriks Konfusi BAGGING CART

Observasi	Prediksi	
	Kredit tidak macet (0)	Kredit macet (1)
Kredit tidak macet (0)	57	15
Kredit macet (1)	2	1

$$APER = \frac{15 + 1}{58 + 15 + 1 + 1} = 0.2133$$

$$Akurasi = 1 - APER = 1 - 0.2 = 0.7867$$

$$Press's Q = \frac{[75 - (59(2))]^2}{75(2 - 1)} = 24.65$$

$$Specificity = \frac{57}{57 + 2} = 0.9661$$

Besarnya ketepatan klasifikasi tersebut menunjukkan bahwa data uji berjumlah 75 terklasifikasikan secara benar sebanyak 78.67% dan terklasifikasikan secara salah sebanyak 21.33%. Nilai *Press's Q* lebih besar dari nilai kritis  $\chi^2_{\alpha,1}$  yaitu 3.841, sehingga dapat disimpulkan bahwa dengan tingkat keyakinan 95% metode BAGGING CART akurat secara statistik dalam mengklasifikasikan nasabah Koperasi Simpan Pinjam dan Pembiayaan Syariah (KSPPS). Nilai *specificity* sebesar 96.61% yang merupakan nilai ketepatan kelas tidak macet yang bernilai 0 atau berlabel negatif diprediksi secara benar

## 5. KESIMPULAN

ketepatan klasifikasi menunjukkan bahwa metode Regresi Logistik Biner menghasilkan nilai akurasi sebesar 78.67% dengan nilai APER 21.33%, *Press's Q* sebesar 24.65, dan *specificity* sebesar 98.30%. Ketepatan klasifikasi yang dihasilkan menggunakan CART dengan nilai akurasi sebesar 77.33% dengan nilai APER 22.67%, nilai *Press's Q* yaitu 22.413, dan *specificity* sebesar 94.91%, kemudian dilakukan pendekatan dengan proses BAGGING ketepatan klasifikasi yang dihasilkan dengan memprediksi data testing nilai akurasi sebesar 78.67% dengan nilai APER 21.33%, nilai *Press's Q* yaitu 24.65, dan *specificity* sebesar 96.61%.

## DAFTAR PUSTAKA

- Agresti, A. (2007). *An Introduction to Categorical Data Analysis*. New York: John Wiley and Sons.
- Breiman, L., Friedman, J. H., Olshen, R., & Stone, C. (1993). *Classification and Regression Trees*. New York: Chapman and Hall (Wadsworth, Inc.).
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression*. New York: John Wiley & Sons.
- Lewis, R. J. (2000). *An Introduction to Classification and Regression Trees (CART)*. California: Presented at the 2000 Annual Meeting of Society for Academic Emergency Medicine of Sanfransisco.
- Republik Indonesia. (2015). Peraturan Menteri Koperasi dan Usaha Kecil dan Menengah Nomor 16 /Per/M.KUKM/IX/2015. *Tentang Pelaksanaan Kegiatan Usaha Simpan Pinjam dan Pembiayaan*. Jakarta: Menteri Koperasi dan Usaha Kecil dan Menengah.
- Sutton, C. D. (2005). *Classification and Regression Trees, Bagging, and Boosting*. *Handbook of Statistics*, Vol. 24, 303-329.