

## KLASIFIKASI PEMBERIAN KREDIT SEPEDA MOTOR MENGGUNAKAN METODE REGRESI LOGISTIK BINER DAN *CHI-SQUARED AUTOMATIC INTERACTION DETECTION* (CHAID) DENGAN GUI R (Studi Kasus: Kredit Sepeda Motor di PT X)

Chalimatus Sa'diah<sup>1</sup>, Tatik Widiharih<sup>2</sup>, Arief Rachman Hakim<sup>3</sup>

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

e-mail: [chalimatussadiyah@students.undip.ac.id](mailto:chalimatussadiyah@students.undip.ac.id)

### ABSTRACT

One of the factors causing the bankruptcy of a company is bad credit. Therefore, prospective customers need to be selected so that bad credit cases can be minimized. This study aims to determine the classification of credit granting to prospective customers of company X in order to reduce the risk of bad credit. The method used is the binary logistic regression method and the Chi-Squared Automatic Interaction Detection (CHAID) method. In this study, data used in November 2019 were 690 motorcycle credit data for company X in Gresik. The independent variables in this study are the factors that affect bad credit such as gender, marital status, education, employment, income, expenses, home ownership status and the dependent variable is credit status (bad and current). The analysis results show that the binary logistic regression has an accuracy value of 76.38% with an APER of 23.62%, while CHAID has an accuracy value of 93.19% with an APER of 6.81%. The accuracy value of the CHAID method is greater than the binary logistic regression method, while the APER value of the CHAID method is smaller than the binary logistic regression method. So it can be concluded that the CHAID method is better than the binary logistic regression method in classifying bad credit at company X.

**Keywords:** Credit, Classification, Binary Logistic Regression, CHAID.

### 1. PENDAHULUAN

Tingginya pembelian sepeda motor di Indonesia diakibatkan oleh banyaknya perusahaan pembiayaan yang menawarkan dananya untuk membiayai pembelian sepeda motor. Beberapa lembaga keuangan seperti bank dan non-bank (*multifinance*) menawarkan penyaluran kredit motor. Perusahaan X merupakan perusahaan pembiayaan yang memiliki berbagai bidang usaha kredit yang dibutuhkan oleh masyarakat salah satunya yaitu kredit sepeda motor. Beberapa nasabah tidak membayar cicilan sesuai jangka waktu yang telah ditetapkan (kredit tersebut macet). Kebangkrutan pada suatu perusahaan dapat disebabkan oleh kredit macet. Oleh karena itu perusahaan X harus melakukan seleksi terhadap calon nasabahnya dengan menggunakan analisis klasifikasi kredit supaya kasus kredit macet dapat diminimalisir. Dalam penelitian ini analisis klasifikasi yang digunakan adalah metode regresi logistik biner dan *Chi-squared Automatic Interaction Detection* (CHAID). Tujuan penelitian ini yaitu untuk membandingkan metode mana yang lebih tepat dalam mengklasifikasikan kredit macet pada perusahaan X.

Pada penelitian ini digunakan dua penelitian terdahulu sebagai rujukan ilmiah. Penelitian pertama dilakukan oleh Diaprina pada tahun 2014 dengan judul "Analisis Klasifikasi Kredit Menggunakan Regresi Logistik Biner dan *Radial Basis Function Network* (RBFN) di Bank X Cabang Kediri" dan dapat disimpulkan bahwa metode yang lebih baik digunakan untuk kasus klasifikasi yaitu menggunakan metode regresi logistik biner karena memiliki nilai ketepatan klasifikasi yang lebih besar dibandingkan dengan menggunakan metode RBFN. Penelitian kedua dilakukan oleh Muhajir pada tahun 2016 dengan judul "Metode Improved CHAID (*Chi-squared Automatic Interaction Detection*) pada Analisis Kredit Macet BMT (Baitul Mal wa Tamwil). Pada kasus BMT El Bummi 372 didapatkan enam segmen dengan tingkat akurasi analisis *Improved CHAID* yaitu sebesar 90,5% yang berarti metode CHAID sudah cukup baik jika digunakan dalam pengklasifikasian calon nasabah peminjam kredit.

## 2. TINJAUAN PUSTAKA

Berdasarkan Surat Keputusan Menteri Keuangan No.84/PMK.012/2006, perusahaan pembiayaan adalah lembaga di luar bank dan bukan lembaga keuangan bank yang didirikan khusus untuk melakukan kegiatan yang termasuk dalam bidang usaha lembaga pembiayaan.

Kredit berasal dari bahasa Yunani yaitu “*credere*” yang artinya kepercayaan dan menurut bahasa latin yaitu “*creditum*” artinya kepercayaan atau kebenaran. Kepercayaan merupakan dasar dari kredit. Kredit adalah tagihan yang nilainya dapat diukur dengan uang, misalnya bank membiayai kredit untuk pembelian mobil atau rumah (Kasmir, 2000).

### 2.1. Regresi Logistik Biner

Regresi logistik biner adalah metode yang menjelaskan hubungan antara variabel dependen yang bersifat kategorik dengan variabel independen (Hosmer dan Lemeshow, 2000). Model regresi logistik biner digunakan untuk menganalisis hubungan antara variabel status kredit (Y) yang bernilai 1 (macet) dan bernilai 0 (tidak macet) dengan faktor-faktor yang mempengaruhi kredit macet (X). Berikut merupakan model regresi logistiknya:

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

Sedangkan logit dari  $\pi(x_i)$  adalah

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

Menurut Agresti (2002), untuk mengestimasi parameter dalam logistik biner dapat menggunakan metode penduga maksimum likelihood.

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

Hipotesis :

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

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

Statistik uji :

$$G = -2 \ln \left[ \frac{\left(\frac{n_1}{n}\right)^{n_1} \left(\frac{n_0}{n}\right)^{n_0}}{\prod_{i=1}^n \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{1 - y_i}} \right]$$

dengan  $n_0 = \sum_{i=1}^n (1 - y_i)$ ,  $n_1 = \sum_{i=1}^n y_i$ , dan  $n = n_0 + n_1$

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

Daerah kritis :

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

#### 2.1.2. Uji Wald (Uji Parsial)

Hipotesis :

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

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

Statistik uji :

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

Daerah kritis :

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

#### 2.1.3. Uji Hosmer and Lemeshow (Uji Kesesuaian Model)

Hipotesis :

$H_0 : \text{Model sesuai}$  (Tidak ada perbedaan antara prediksi dengan hasil observasi)

H<sub>1</sub> : Model tidak sesuai (Ada perbedaan antara prediksi dengan hasil observasi)

Statistik uji :

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

Daerah kritis :

H<sub>0</sub> ditolak jika  $\hat{C} > \chi^2_{(\alpha, g-2)}$  atau H<sub>0</sub> ditolak jika nilai *p-value* <  $\alpha$

## 2.2. CHAID (*Chi-Squared Automatic Interaction Detection*)

*Chi-squared Automatic Interaction Detection* (CHAID) pertama kali dikenalkan oleh Dr.G.V.Kass dalam artikel dengan judul “*An Exploratory Technique for Investigating Large Quantities of Categorical Data*” pada tahun 1980. Menurut Gallagher (2000), CHAID adalah suatu teknik iteratif yang menguji satu-persatu variabel independen dalam klasifikasi, kemudian disusun berdasarkan pada tingkat signifikansi *chi-square* terhadap variabel dependennya.

### 2.2.1. Variabel-Variabel dalam CHAID

Menurut Gallagher (2000), terdapat tiga bentuk variabel yang berbeda dalam CHAID:

1. Variabel Monotonik adalah variabel independen yang kategorinya dapat digabungkan atau dikombinasikan hanya jika kedua variabel tersebut berdekatan satu sama lain atau mengikuti urutan aslinya (data ordinal).
2. Variabel Bebas adalah variabel independen yang kategorinya dapat digabungkan atau dikombinasikan ketika kedua variabel berdekatan maupun tidak berdekatan (data nominal).
3. Variabel Mengambang (*floating*) adalah variabel independen yang kategorinya dapat diperlakukan seperti monotonik kecuali untuk kategori yang terakhir (*missing value*) yang dapat dikombinasikan dengan kategori manapun.

### 2.2.2. Uji *Chi-Square* (Independensi)

Menurut Kunto (2006), teknik uji *chi-square* memungkinkan untuk mengetahui independensi antara dua variabel pada tiap levelnya.

Hipotesis :

H<sub>0</sub> : Tidak terdapat hubungan antara variabel independen dengan variabel dependen (Kedua kriteria klasifikasi bebas)

H<sub>1</sub> : Terdapat hubungan antara variabel independen dengan variabel dependen (Kedua kriteria klasifikasi tidak bebas)

Statistik uji :

$$\chi^2_{hitung} = \sum_{k=1}^r \sum_{l=1}^c \left[ \frac{(n_{kl} - E_{kl})^2}{E_{kl}} \right]$$

$$E_{kl} = \frac{n_{k.} n_{.l}}{n}, \quad k = 1, 2, \dots, r \quad \text{dan} \quad l = 1, 2, \dots, c.$$

Daerah kritis :

H<sub>0</sub> ditolak jika  $\chi^2_{hitung} > \chi^2_{\alpha : (r-1)(c-1)}$  atau H<sub>0</sub> ditolak jika nilai *p-value* <  $\alpha$

### 2.2.3. Koreksi Bonferroni

Menurut Gallagher (2000), pengali bonferroni untuk masing-masing jenis variabel independen adalah sebagai berikut:

1. Variabel monotonik

$$M = \binom{c-1}{r-1}$$

2. Variabel bebas

$$M = \sum_{i=1}^{r-1} (-1)^i \frac{(r-i)^c}{i!(r-i)!}$$

3. Variabel mengambang (*floating*)

$$M = \binom{c-2}{r-2} + r \binom{c-2}{r-1}$$

dengan :

$M$  = pengali bonferroni

$c$  = banyaknya kategori variabel independen awal

$r$  = banyaknya kategori variabel independen setelah penggabungan

#### 2.2.4. Algoritma CHAID

Algoritma CHAID dibagi menjadi tiga tahap yaitu tahap penggabungan (*merging*), tahap pemisahan (*splitting*), dan tahap penghentian (*stopping*).

##### 1. Tahap Penggabungan (*merging*)

Menurut Gallagher (2000), tahap penggabungan adalah sebagai berikut:

- Hitung statistik *chi-square* untuk setiap variabel independen.
- Gabungkan pasangan yang tidak signifikan yaitu pasangan yang memiliki nilai *chi-square* terkecil dan nilai *p-value* terbesar menjadi kategori tunggal dari variabel independen yang mempunyai 3 atau lebih kategori asal.
- Periksa kembali signifikansi kategori baru setelah penggabungan, apabila masih terdapat pasangan kategori yang belum signifikan maka ulangi ke langkah a dan jika semua pasangan kategori sudah signifikan, maka lanjutkan ke langkah berikutnya.
- Hitung koreksi bonferroni didasarkan pada kategori yang sudah digabungkan.

##### 2. Tahap Pemisahan (*splitting*)

Tahap pemisahan dilakukan dengan cara memilih variabel independen terbaik yaitu variabel dengan nilai *chi-square* terbesar atau nilai *p-value* terkecil, selanjutnya membagi kelompok variabel independen dengan menggunakan masing-masing kategori variabel independen yang telah digabungkan untuk menentukan sub pembagian dari kelompok induk menjadi sub kelompok baru dan apabila sudah tidak ada variabel independen yang memiliki nilai *chi-square* atau nilai *p-value* yang signifikan, maka jangan memulai pembagian kelompok tersebut.

##### 3. Tahap Penghentian (*stopping*)

Menurut Rokach dan Maimon (2008), penyekatan simpul dalam pembentukan pohon harus dihentikan apabila memenuhi aturan sebagai berikut :

- Terdapat kasus yang berasal dari salah satu kelas variabel dependen pada simpul.
- Telah tercapai kedalaman pohon yang maksimal.
- Jumlah minimal kasus untuk menjadi simpul induk lebih besar dari jumlah kasus di simpul terminal.
- Apabila simpul disekat, jumlah kasus dalam satu atau lebih simpul anak akan kurang dari jumlah minimal kasus untuk menjadi simpul anak maka proses penyekatan simpul berhenti.
- Simpul tidak dapat disekat apabila semua variabel independen mempunyai nilai signifikan lebih besar dari nilai  $\alpha$  yang ditentukan.

#### 2.3. Ketepatan Klasifikasi

Menurut Johnson dan Wichern (1992), *Apparent Error Rate* (APER) adalah prosedur evaluasi yang digunakan untuk melihat kesalahan klasifikasi yang dilakukan oleh suatu fungsi klasifikasi. Nilai APER menyatakan nilai proporsi sampel yang salah diklasifikasikan oleh fungsi klasifikasi. Kesalahan klasifikasi dapat dilihat pada tabel matriks konfusi berikut:

**Tabel 1.** Matriks konfusi

Hasil Observasi	Hasil Prediksi	
	Kelas A	Kelas B
Kelas A	$f_{11}$	$f_{10}$
Kelas B	$f_{01}$	$f_{00}$

Maka dapat dihitung tingkat akurasi dan tingkat kesalahan prediksi (APER) sebagai berikut:

$$\text{Tingkat akurasi} = \frac{f_{11}+f_{00}}{f_{11}+f_{10}+f_{01}+f_{00}}$$

$$\text{Tingkat kesalahan prediksi (APER)} = \frac{f_{10}+f_{01}}{f_{11}+f_{10}+f_{01}+f_{00}}$$

## 2.4. Program R

R adalah bahasa pemrograman open source yang memiliki hubungan dengan komputasi dan pengolahan data statistika yang menampilkan grafik dengan menggunakan tools yang telah disediakan oleh paket-paketnya yang sangat berguna di dalam penelitian dan industri (Budiharto dan Rachmawati, 2013). R memiliki dua bentuk umum pengguna dalam komputasi statistika yaitu *Command Line Interface* (CLI) dan *Graphical User Interface* (GUI). R-shiny memungkinkan orang membuat laman web (*web pages*) interaktif sehingga kemampuan R yang pada dasarnya bersifat CLI bisa diakses melalui menu web secara GUI web (Tirta, 2015).

## 3. METODE PENELITIAN

### 3.1. Sumber Data

Sumber data yang digunakan dalam penulisan ini adalah data sekunder yang diperoleh dari hasil pendataan kredit yang dilakukan oleh suatu perusahaan X di Gresik. Data yang digunakan adalah data kredit sepeda motor bulan November tahun 2019 sebanyak 690 data

### 3.2. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini yaitu status kredit sebagai variabel dependen (Y) dan jenis kelamin ( $X_1$ ), status perkawinan ( $X_2$ ), pendidikan ( $X_3$ ), pekerjaan ( $X_4$ ), penghasilan ( $X_5$ ), pengeluaran ( $X_6$ ), serta status kepemilikan rumah ( $X_7$ ) sebagai variabel independennya.

### 3.3. Metode Analisis Data

Metode analisis yang digunakan dalam penelitian ini adalah metode CHAID dan Regresi logistik biner.

1. Tahapan analisis data yang dilakukan dengan menggunakan metode Regresi logistik biner yaitu:
  - a. Membentuk model regresi logistik biner.
  - b. Melakukan uji rasio likelihood.
  - c. Melakukan uji wald.
  - d. Melakukan uji hosmer and lemeshow terhadap data.
  - e. Menentukan dan menginterpretasi 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.
    - Apabila nilai  $\pi(x_i) > 0,5$  maka masuk kedalam kelas 1 dan dikategorikan sebagai kredit macet.
  - g. Membentuk matriks konfusi regresi logistik biner
  - h. Menghitung nilai ketepatan klasifikasi
2. Tahapan analisis data yang dilakukan dengan menggunakan metode CHAID yaitu:
  - a. Menghitung nilai *chi-square* untuk setiap pasangan kategori variabel independen dan variabel dependen.
  - b. Gabungkan kategori dalam variabel independen yang memiliki nilai *chi-square* terkecil.
  - c. Menghitung kembali nilai *chi-square*
  - d. Melakukan koreksi bonferroni

- e. Pemilihan variabel independen yang memiliki nilai *chi-square* terbesar sebagai pemisah untuk membentuk sub kelompok. Proses pemilihan variabel terus berjalan hingga mencapai peraturan penghentian yang diberlakukan.
- f. Menginterpretasi pohon klasifikasi.
- g. Menghitung nilai ketepatan klasifikasi.

## 4. HASIL DAN PEMBAHASAN

### 4.1. Analisis Regresi Logistik Biner

#### 4.1.1. Model Awal

Model awal regresi logistik biner yang terbentuk yaitu  $\pi(x_i) = \frac{e^{g(x_i)}}{1+e^{g(x_i)}}$  sebagai peluang kredit macet dengan

$$g(x_i) = 1,51337 - 1,28594 x_1(1) - 0,33265 x_2(2) - 3,36832 x_2(3) + 0,39564 x_3(2) + 0,18899 x_3(3) + 0,06162 x_4(2) + 0,34316 x_4(3) + 1,20134 x_4(4) - 1,55094 x_4(5) - 22,65312 x_5(2) - 43,19720 x_5(3) + 19,14282 x_6(2) + 22,77250 x_6(3) + 0,24887 x_7(2) + 1,27742 x_7(3)$$

1. Uji rasio likelihood model awal

Hipotesis :

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_{15} = 0$$

$H_1$  : Paling sedikit ada satu  $\beta_j \neq 0$  dengan  $j = 1, 2, \dots, 15$

$$\text{Nilai kritis} : \chi^2_{(p,\alpha)} = \chi^2_{(15:0,05)} = 24,9958$$

Statistik uji :

$$G = -2 \ln \left[ \frac{\text{likelihood tanpa variabel bebas}}{\text{likelihood dengan variabel bebas}} \right] = 508,6089361$$

$H_0$  ditolak karena  $G = 508,6089361 > \chi^2_{(15:0,05)} = 24,9958$ . Sehingga dapat disimpulkan bahwa paling sedikit ada satu variabel independen yang berpengaruh terhadap variabel status kredit.

2. Uji wald model awal

Hipotesis :

$H_0 : \beta_j = 0$  dengan  $j = 1, 2, \dots, 15$  (Tidak ada pengaruh antara variabel independen dengan variabel status kredit)

$H_1 : \beta_j \neq 0$  dengan  $j = 1, 2, \dots, 15$  (Ada pengaruh antara variabel independen dengan variabel status kredit)

$$\text{Nilai kritis} : \chi^2_{(\alpha,1)} = \chi^2_{(0,05:1)} = 3,84146$$

Statistik uji :

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

Kesimpulan :

Pada taraf signifikansi  $\alpha = 5\%$ ,  $H_0$  ditolak untuk variabel  $X_1(1)$  = jenis kelamin (laki-laki),  $X_2(3)$  = status perkawinan (belum kawin), dan  $X_7(3)$  = status kepemilikan rumah (rumah sewa/kos/kontrak). Untuk variabel  $X_2(2)$  = status perkawinan (cerai) dianggap signifikan karena masih termasuk kategori dari variabel status perkawinan dan  $X_7(2)$  = status kepemilikan rumah (rumah sendiri) dianggap signifikan karena masih termasuk kategori dari variabel status kepemilikan rumah. Sehingga ada pengaruh antara variabel independen dengan variabel status kredit.

Setelah melakukan uji signifikansi secara keseluruhan maupun secara individu terhadap model regresi logistik biner, didapatkan hasil bahwa variabel independen yang berpengaruh secara signifikan terhadap status kredit yaitu variabel jenis kelamin ( $X_1$ ), status perkawinan ( $X_2$ ) dan status kepemilikan rumah ( $X_7$ ). Karena terdapat beberapa variabel independen yang

tidak signifikan maka akan dibentuk model regresi logistik biner yang baru dengan tidak mengikutsertakan variabel independen yang tidak signifikan.

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

Variabel	Estimate	Std.Error	Nilai z	p-value	Keterangan
X <sub>1</sub> (1)	-1,28594	0,335161	-3,657	0,000255	Signifikan
X <sub>2</sub> (2)	-0,33265	0,58354	-0,570	0,568641	Tidak Signifikan
X <sub>2</sub> (3)	-3,36832	1,12131	-3,004	0,002665	Signifikan
X <sub>3</sub> (2)	0,39564	0,43224	0,915	0,360020	Tidak Signifikan
X <sub>3</sub> (3)	0,18899	0,87022	0,217	0,828071	Tidak Signifikan
X <sub>4</sub> (2)	0,06162	1,13785	0,054	0,956811	Tidak Signifikan
X <sub>4</sub> (3)	0,034316	1,14571	0,300	0,764547	Tidak Signifikan
X <sub>4</sub> (4)	1,20134	1,29693	0,926	0,354292	Tidak Signifikan
X <sub>4</sub> (5)	-1,55094	1,40704	-1,102	0,270344	Tidak Signifikan
X <sub>5</sub> (2)	-22,65312	1133,40573	-0,020	0,984054	Tidak Signifikan
X <sub>5</sub> (3)	-43,19720	1984,77639	-0,022	0,982636	Tidak Signifikan
X <sub>6</sub> (2)	19,14282	1133,40579	0,017	0,986525	Tidak Signifikan
X <sub>6</sub> (3)	22,77250	1133,40604	0,020	0,983970	Tidak Signifikan
X <sub>7</sub> (2)	0,24887	0,40678	0,612	0,540668	Tidak Signifikan
X <sub>7</sub> (3)	1,27742	0,55549	2,300	0,021468	Signifikan
Constant	1,51337	1,50064	1,008	0,313222	-

3. Uji Hosmer and Lemeshow

Hipotesis :

H<sub>0</sub> : Model sesuai (Tidak ada perbedaan antara prediksi dengan hasil observasi)

H<sub>1</sub> : Model tidak sesuai (Ada perbedaan antara prediksi dengan hasil observasi)

Nilai kritis :  $\chi^2_{(\alpha, g-2)} = \chi^2_{(0,05;8)} = 15,50731$

Statistik uji :

$$\hat{C} = \sum_{k=1}^g \frac{(o_k - n_k \bar{\pi}_k)^2}{(n_k \bar{\pi}_k)(1 - \bar{\pi}_k)} = 0,19791 ; p\text{-value} = 1$$

Kesimpulan :

Pada taraf signifikansi  $\alpha = 5\%$ , H<sub>0</sub> diterima karena  $\hat{C} = 0,19791 < \chi^2_{(0,05;8)} = 15,50731$  atau  $p\text{-value} = 1 > \alpha = 0,05$ . Sehingga dapat disimpulkan bahwa model sesuai (Tidak ada perbedaan antara prediksi dengan hasil observasi).

4. Model akhir

Berdasarkan hasil uji rasio likelihood, uji wald dan uji hosmer and lemeshow maka diperoleh model akhir regresi logistik biner untuk status kredit yaitu sebagai berikut:

$$\pi(x_i) = \frac{e^{g(x_i)}}{1 + e^{g(x_i)}} \text{ dengan}$$

$$g(x_i) = 0,02484 - 1,82978 x_1(1) - 0,49936 x_2(2) - 2,42599 x_2(3)$$

**4.1.2. Ketepatan Klasifikasi Regresi Logistik Biner**

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

Prediksi	Observasi	
	Kredit tidak macet (0)	Kredit macet (1)
Kredit tidak macet (0)	415	57
Kredit macet (1)	106	112

Berdasarkan Tabel 3, maka dapat dihitung nilai akurasi ketepatan klasifikasi dan tingkat kesalahan prediksi (APER) sebagai berikut :

$$\text{Akurasi} = \frac{415+112}{415+57+106+112} = 0,7637681 = 76,38\%$$

$$\text{APER} = \frac{57+106}{415+57+106+112} = 0,2362319 = 23,62\%$$

## 4.2. Analisis CHAID (*Chi-squared Automatic Interaction Detection*)

### 4.2.1. Tahap Penggabungan

Pada penelitian ini variabel pendidikan mengalami proses penggabungan dikarenakan variabel pendidikan memiliki nilai *chi-square* paling kecil. Berikut merupakan hasil uji statistik *chi-square*:

**Tabel 4.** Variabel Status Kredit dan Variabel Pendidikan

Status Kredit	Pendidikan		Total
	SLTA (2); Diploma, Sarjana (3)	SD, SLTP (1)	
Kredit tidak macet (0)	424	97	521
Kredit macet (1)	126	43	169
Total	550	140	690

$$\chi^2_{hitung} = \sum_{i=1}^r \sum_{j=1}^c \left[ \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \right] = 3,6761$$

Berdasarkan perhitungan diatas  $H_0$  diterima karena  $\chi^2_{hitung} = 3,6761 < \chi^2_{(0,05;1)} = 3,841$ . Sehingga dapat disimpulkan bahwa tidak terdapat hubungan antara variabel status kredit dan variabel pendidikan kategori (SLTA; diploma, sarjana dan SD, SLTP) (saling bebas) maka seluruh kategori variabel pendidikan digabungkan menjadi satu kategori tunggal dan kemudian dihitung koreksi bonferroni.

$$M = \binom{c-1}{r-1} = \binom{3-1}{1-1} = \frac{2!}{0!(2-0)!} = 1$$

Hasil dari koreksi bonferroni adalah  $2,2 \times 10^{-16} \times 1 = 2,2 \times 10^{-16}$ . Karena nilai *p-value* terkoreksi =  $2,2 \times 10^{-16} < \alpha = 0,05$  maka dapat disimpulkan bahwa variabel pendidikan signifikan (tidak saling bebas) terhadap variabel status kredit.

### 4.2.2. Tahap Pemisahan

**Tabel 5.** Uji *Chi-square* Tahap Pemisahan Simpul 1 ( $t_1$ )

Variabel	Kategori	<i>Chi-square</i>	Keputusan	Kesimpulan
Jenis Kelamin ( $X_1$ )	Perempuan (0)	97,621	$H_0$ ditolak	signifikan
	Laki-laki (1)			
Status Perkawinan ( $X_2$ )	Kawin (1)	34,182	$H_0$ ditolak	Signifikan
	Cerai (2)			
	Belum Kawin (3)			
Pendidikan ( $X_3$ )	SD, SLTP; SLTA; Diploma, Sarjana (1)	179,57	$H_0$ ditolak	Signifikan
Pekerjaan ( $X_4$ )	PNS/TNI/polisi/BUMN/pegawai pemerintah (1)	50,767	$H_0$ ditolak	Signifikan
	Karyawan Swasta (2)			
	Wiraswasta/Pedagang (3)			
	Petani/Peternak/Nelayan (4)			
	Lainnya (5)			



Penghasilan (X <sub>5</sub> )	< Rp. 3.000.000,00 (1)	372,15	H <sub>0</sub> ditolak	Signifikan
	Rp. 3.000.000,00 – Rp. 5.000.000,00 (2)			
	> Rp. 5.000.000,00 (3)			
Pengeluaran (X <sub>6</sub> )	< Rp. 2.000.000,00 (1)	19,418	H <sub>0</sub> ditolak	Signifikan
	Rp. 2.000.000,00 – Rp. 3.000.000,00 (2)			
	> Rp. 3.000.000,00 (3)			
Status Kepemilikan Rumah (X <sub>7</sub> )	Rumah Keluarga/Rumah Dinas (1)	22,244	H <sub>0</sub> ditolak	Signifikan
	Rumah Sendiri (2)			
	Rumah Sewa/Kos/Kontrak (3)			

Berdasarkan Tabel 5, variabel penghasilan merupakan variabel pemisah terbaik pada simpul 1 (t<sub>1</sub>) karena mempunyai nilai *chi-square* terbesar yaitu 372,15. Proses pemisahan simpul terus dilakukan dan berhenti ketika memenuhi syarat tahap penghentian.

#### 4.2.3. Tahap Penghentian

Tahap penghentian terjadi pada simpul status perkawinan kategori kawin, status perkawinan kategori cerai, status perkawinan kategori belum kawin, pengeluaran kategori < Rp. 2.000.000,00, jenis kelamin kategori perempuan, jenis kelamin kategori laki-laki, pengeluaran kategori >Rp. 3.000.000,00 dan penghasilan kategori > Rp. 5.000.00,00. Hal tersebut terjadi disebabkan oleh faktor sebagai berikut:

1. Proses penghentian terjadi pada simpul status perkawinan kategori cerai, status perkawinan kategori belum kawin, jenis kelamin kategori perempuan, jenis kelamin kategori laki-laki, dan pengeluaran kategori >Rp. 3.000.000,00 disebabkan karena sudah tidak ada lagi variabel independen yang signifikan.
2. Proses penghentian terjadi pada simpul status perkawinan kategori kawin, pengeluaran kategori < Rp. 2.000.000,00 dan penghasilan kategori > Rp. 5.000.00,00 disebabkan karena hanya terdapat 1 kategori variabel dependen.

#### 4.2.4. Ketepatan Klasifikasi CHAID

Tabel 6. Matriks Konfusi Regresi Logistik Biner

Prediksi	Observasi	
	Kredit tidak macet (0)	Kredit macet (1)
Kredit tidak macet (0)	516	42
Kredit macet (1)	5	127

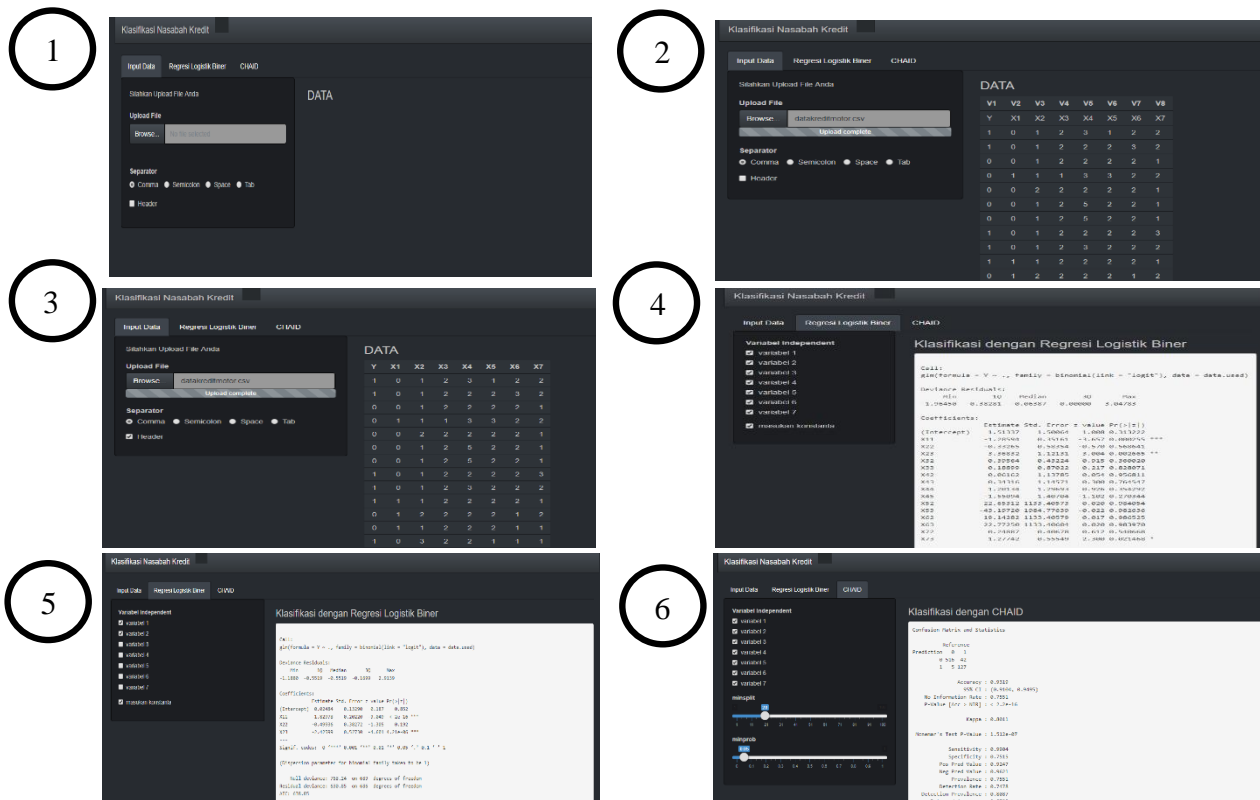
Berdasarkan Tabel 6, maka dapat dihitung nilai akurasi ketepatan klasifikasi dan tingkat kesalahan prediksi (APER) dan sebagai berikut :

$$\text{APER} = \frac{42+5}{516+42+5+127} = 0,0681 = 6,81\%$$

$$\text{Akurasi} = \frac{516+127}{516+42+5+127} = 0,9319 = 93,19\%$$

### 4.3. Langkah-langkah menggunakan *Graphical User Interface (GUI) R*

Berikut merupakan langkah-langkah untuk menggunakan GUI:



1. Klik tombol browse untuk memasukkan file (dalam format .csv)
2. Tunggu sampai file berhasil diupload.
3. Setelah file berhasil diupload, klik header supaya baris pertama adalah data yang akan digunakan.
4. Klik Regresi Logistik Biner. Kemudian centang semua variabel independent untuk model awal. Apabila ingin memasukkan intersep (konstanta) dalam model maka centang “masukkan konstanta”.
5. Apabila ada variabel yang tidak signifikan maka variabel tersebut dibuang dan tidak dimasukkan kedalam model regresi logistik biner. Kemudian membuat model baru dengan mencentang variabel yang signifikan. Jika ingin memasukkan intersep (konstanta) dalam model maka centang “masukkan konstanta”.
6. Klik CHAID. Kemudian centang variabel independent sesuai model yang diinginkan. Atur minsplit dan minprob sesuai yang diinginkan dengan cara menggeser slider.

### 5. KESIMPULAN

Berdasarkan hasil dan pembahasan dapat diambil kesimpulan sebagai berikut:

1. Faktor-faktor yang mempengaruhi kredit macet di perusahaan X dengan metode regresi logistik biner yaitu variabel jenis kelamin dan variabel status perkawinan. Sedangkan untuk metode CHAID faktor-faktor yang mempengaruhi yaitu variabel jenis kelamin, variabel status perkawinan, variabel penghasilan dan variabel pengeluaran.
2. Metode regresi logistik biner memiliki nilai akurasi sebesar 76,38% dan APER sebesar 23,62% dalam mengklasifikasikan kredit macet di perusahaan X. Sedangkan metode CHAID memiliki nilai akurasi sebesar 93,19% dan APER sebesar 6,81% dalam mengklasifikasikan kredit macet di perusahaan X.

3. Berdasarkan hasil perbandingan ketepatan klasifikasi dengan metode regresi logistik biner dan metode CHAID diperoleh hasil bahwa metode CHAID lebih baik dari regresi logistik biner dalam mengklasifikasikan kredit macet di perusahaan X.

#### **DAFTAR PUSTAKA**

- Agresti, A.2002. *An Introduction to Categorical Data Analysis*. New York: John Wiley & Sons. Inc.
- Budiharto, W., dan Rachmawati, N. R. 2013. *Pengantar Praktis Pemrograman R untuk Ilmu Komputer*. Jakarta: Halaman Moeka.
- Gallagher, C.A. 2000. *An Iterative Approach to Classification Analysis*. [www.casact.org/library/ratemaking/90dp237.pdf](http://www.casact.org/library/ratemaking/90dp237.pdf)
- Hosmer, D.W., dan Lemeshow. 2000. *Applied Logistic Regression*. USA: John Willey and Sons.
- Johnson, R.A. dan Winchern, DW. 2007. *Applied Multivariate Statistical Analysis Sixth Edition*. New Jersey: Pearson.
- Kasmir. 2000. *Bank dan Lembaga Keuangan Lainnya*. Edisi keenam. Salemba Empat, Jakarta.
- Kunto, Y.S. 2006. *Analisis CHAID Sebagai Alat Bantu Statistik untuk Segmentasi Pasar*. Jurnal Manajemen Pemasaran, Vol. 01, No. 02. Surabaya: Universitas Kristen Petra.
- Rokach, L. dan Maimon, O. 2008. *Data Mining with Decision Trees "Theory and Application"*. USA: World Scientific Publishing Co. Pte. Ltd.
- Surat Keputusan Menteri Keuangan No.84/PMK.012/2006 Tentang Kegiatan Usaha Perusahaan Pembiayaan.
- Tirta, I. M. 2006. *Mendesain Paket Analisis dan Media Pembelajaran Statistika..* Jember: Universitas Jember.
- Tirta, I.M. 2015. *Panduan Penyusunan Modul dan Analisis Data Online Berbasis Web Interaktif Menggunakan R dalam Rangka Melengkapi K Statistika Virtual (Online, Interaktif, Terintegrasi)*. Jember: Universitas Jember.