

**PENERAPAN METODE CART (CLASSIFICATION AND REGRESSION TREES)
UNTUK MENENTUKAN FAKTOR-FAKTOR YANG MEMPENGARUHI
PEMBAYARAN KREDIT OLEH NASABAH
(STUDI KASUS BANK BRI UNIT AEK TARUM – SUMATERA UTARA)**

Rezeki Handayani Tanjung¹, Kartiko²

^{1,2}Jurusan Statistika, FST, Institut Sains & Teknologi AKPRIND Yogyakarta

Email : rejekihandayanirh@gmail.com

Abstract

This study aims to classified credit payment by debtors at Bank BRI Unit Aek Tarum – North Sumatera. The results are the factors that influence to credit payment by the debtors.

To know the results of credit payment classification using CART analysis, tree classification approach was applied.

The result of CART classification using tree analysis shows that credit payment was influenced by the income amount, age and plafond with the accuracy classification 84,2%. The debtor characteristic that has a disposed stuck payment was debtor with maximal income Rp 4.550.000, maximal plafond Rp 17.500.000 and the maximum age of 39,5 years old with percentase 20.5%, debtor with income more than Rp 4.550.000 with percentase 18.1%, and debtor with maximal income Rp 4.550.000 and plafond more than Rp 17.500.000 with percentase 1.4%.

Keywords: *CART, Classification Tree, Bank Credit, Plafond*

ABSTRAK

Penelitian ini bertujuan untuk mengklasifikasi pembayaran kredit nasabah Bank BRI Unit Aek Tarum – Sumatera Utara. Hasil pengklasifikasian tersebut, berupa faktor-faktor penciri yang mempengaruhi pembayaran kredit oleh nasabah.

Untuk mengetahui hasil pengklasifikasian pembayaran kredit, digunakan analisis CART dengan pendekatan analisis pohon klasifikasi.

Hasil analisis pohon klasifikasi CART menunjukkan bahwa status pembayaran kredit dipengaruhi oleh jumlah penghasilan, usia dan plafond dengan ketepatan hasil klasifikasi yang terbentuk sebesar 84,2%. Karakteristik nasabah yang memiliki status pembayaran kredit cenderung tidak lancar adalah nasabah yang mempunyai jumlah penghasilan maksimum Rp 4.550.000, dengan plafond maksimum Rp 17.500.000 usia maksimum 39,5 tahun dengan persentase 20,5%, nasabah dengan jumlah penghasilan lebih dari Rp 4.550.000 dengan persentase 18,1%, serta nasabah dengan jumlah penghasilan maksimum Rp 4.550.000 dengan plafond lebih dari Rp 17.500.000 persentase 1,4%.

Kata kunci: *CART, Pohon Klasifikasi, Kredit Bank, Plafond*

1. Pendahuluan

Metode CART (*Classification and Regression Trees*) adalah metode klasifikasi berstruktur pohon yang pertama kali diperkenalkan oleh [1]. Klasifikasi pada metode ini dilakukan dengan membangun sebuah pohon klasifikasi yang diperoleh melalui penyekatan berulang terhadap sebuah himpunan data (dalam hal ini diistilahkan sebagai simpul) menjadi dua simpul baru. Metode CART efektif bila diterapkan pada data dengan pengamatan yang relatif banyak [3].

Kemampuan pengklasifikasian suatu metode sangat membantu untuk pemecahan persoalan dalam suatu institusi yang membutuhkan. Bank adalah suatu institusi kreditur yang sering mengalami kredit yang bermasalah sehingga membuat pihak bank harus menentukan debitur yang layak untuk melakukan kredit.

Kondisi yang terjadi saat ini adalah pihak bank dalam melakukan pengecekan data kredit dengan status lancar dan tidak lancar adalah secara manual sehingga tidak efisien dan dapat terjadi kesalahan. Oleh karena itu diperlukan suatu metode untuk mengklasifikasikan data secara otomatis menggunakan alat bantu *SPSS 18.0*.

Penelitian sebelumnya yang berkaitan dengan pengklasifikasian diantaranya: [2]“Aplikasi Algoritma *CART* untuk Mengklasifikasikan Data Nasabah Asuransi Jiwa Bersama Bumiputera 1912 Surakarta”, [6]“ Penerapan Metode *CART* (*Classification And Regression Trees*) dan Analisis Regresi Logistik Biner pada Klasifikasi Profil Mahasiswa FMIPA Universitas Brawijaya, [5]” *Using Classification and Regression Trees (CART) in SAS® Enterprise Miner™ For Applications in Public Health*”. Berdasarkan penelitian-penelitian tersebut dapat disimpulkan bahwa metode *CART* (*Classification and Regression Trees*) efektif dalam mengklasifikasikan suatu data.

Berdasarkan pertimbangan tersebut, maka dilakukan penelitian dengan judul “Penerapan Metode *CART* Dalam Menentukan Faktor-Faktor Yang Mempengaruhi Pembayaran Kredit Oleh Nasabah Dengan Studi Kasus Bank Bri Unit Aek Tarum – Sumatera Utara”. Tujuan penelitian ini untuk mengetahui prosedur analisis metode *CART* dan mendapatkan hasil analisis serta mengetahui peubah apa saja yang menjadi faktor pencari pada kasus pembayaran kredit. Data yang digunakan merupakan data pembayaran kredit pada tahun 2016 yang diperoleh dari Bank BRI Unit Aek Tarum – Sumatera Utara.

2. Metode

2.1 Sumber Data

Penelitian ini menggunakan data sekunder sebanyak 650 data yaitu data yang didapat dari pembayaran kredit nasabah oleh Bank BRI unit Aek Tarum – Sumatera Utara tahun 2016.

2.2 Peubah Penelitian

Peubah dalam penelitian ini terdiri atas peubah respon (*Y*) yaitu status pembayaran kredit berskala kategorik, dan peubah penjelas (*X*) adalah jenis kelamin, usia, anggota keluarga, jumlah penghasilan, plafond dan lama pinjaman (dalam bulan) adalah berskala kategorik dan numerik.

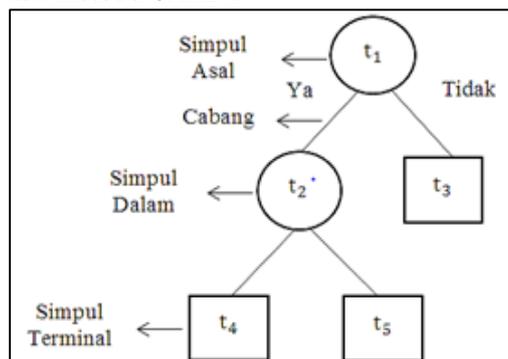
2.3 Metode Analisis Data

2.3.1 Pohon Klasifikasi dan Pohon Regresi

Pohon klasifikasi dan pohon regresi atau yang dalam bahasa inggrisnya *Classification and Regression Trees (CART)* merupakan metode statistika nonparametrik yang dikembangkan untuk keperluan analisis klasifikasi, baik untuk peubah respon kategorik maupun kontinu [1].

Pohon klasifikasi merupakan metode penyekatan data secara berulang sehingga menghasilkan pohon yang tersusun atas banyak simpul (*node*) yang terbentuk dari proses pemilahan *rekursif biner*.

Berikut ini skema pohon pada metode *CART* :



Gambar 1. Skema Pohon Klasifikasi dan Regresi

Dengan t_1 = Simpul asal, t_2 = Simpul dalam, dan t_3, t_4, t_5 = Simpul terminal

Dalam pembangunan pohon klasifikasi *CART* terdapat tiga tahapan yaitu:

a. Pemilihan Pemilah

Pemilihan pemilah pada setiap simpul bertujuan mendapatkan pemilah yang mampu menghasilkan simpul dengan tingkat kehomogenan nilai peubah respon paling tinggi. Keheterogenan suatu simpul diukur berdasarkan nilai impuritasnya.

Metode pemilihan pemilah menggunakan *impurity measure* $i(t)$ yaitu pengukuran tingkat heterogenan suatu kelas dari suatu simpul tertentu dalam pohon klasifikasi yang dapat menemukan fungsi pemilah optimal. Fungsi *impurity measure* $i(t)$ yang digunakan adalah fungsi *indeks Gini* di tunjukkan pada rumus:

$$i(t) = 1 - \sum_j P^2(j|t) \quad (1)$$

dengan $P(j|t)$ adalah proporsi pada kelas j pada simpul t .

Pemilah dimulai dengan cara memeriksa nilai-nilai pada setiap peubah penjelasnya untuk mendapatkan *split-point* yang diperoleh dengan mencari nilai tengah dari 2 nilai atribut yang sudah diurutkan terlebih dahulu.

Goodness of Split merupakan suatu evaluasi pemilahan oleh pemilah s pada simpul t yang didefinisikan sebagai penurunan heterogenan dan didefinisikan sebagai:

$$\Delta i(s, t) = i(t) - P_L i(t_L) - P_R i(t_R) \quad (2)$$

dengan $P_L i(t_L)$ adalah proporsi pengamatan dari simpul t menuju kiri, $P_R i(t_R)$ adalah proporsi pengamatan dari simpul t menuju kanan.

Pemilah terbaik adalah pemilah yang memberikan nilai penurunan heterogenan atau nilai penurunan impuritas tertinggi [1]. Pohon dikembangkan dengan cara pada simpul t_1 , carilah s^* yang menghasilkan nilai penurunanan impuritas tertinggi yaitu:

$$\Delta i(s^*, t_1) = \max_{s \in S} \Delta i(s, t_1) \quad (3)$$

dengan demikian t_1 dipilih menjadi t_2 dan t_3 menggunakan s^* . Dengan cara yang sama dilakukan juga pencarian pemilah terbaik pada t_2 dan t_3 secara terpisah dan demikian seterusnya.

b. Penentuan Simpul Terminal

Simpul dikatakan sebagai simpul terminal ketika suatu simpul t mencapai batas akhir yang ditentukan sehingga tidak terdapat penurunan impuritas secara berarti. Simpul t tidak dipilah lagi tetapi dijadikan simpul terminal dan pembentukan pohon berhenti [1], sesuai dengan peraturan pemberhentian di bawah ini :

- Tidak ada lagi peubah respon yang signifikan menunjukkan perbedaan terhadap peubah penjelas.
- Jika pohon sekarang mencapai batas nilai maksimum pohon dari spesifikasi, maka proses pertumbuhan akan berhenti. Misalkan ditetapkan batas kedalaman pertumbuhan pohon klasifikasi adalah 3, ketika pertumbuhan pohon sudah mencapai kedalaman 3 maka pertumbuhan pohon klasifikasi akan berhenti.
- Jika ukuran dari *child node* kurang dari nilai ukuran *child node* minimum spesifikasi, atau berisi pengamatan-pengamatan dengan jumlah yang terlalu sedikit maka *node* tidak akan terpilah. Misalkan ditetapkan ukuran minimal *child node* adalah 50, ketika ukuran *child node* kurang dari 50, maka node tersebut tidak akan terpilah.

c. Penandaan Label Kelas

Label kelas pada simpul terminal t ditentukan melalui aturan jumlah terbanyak:

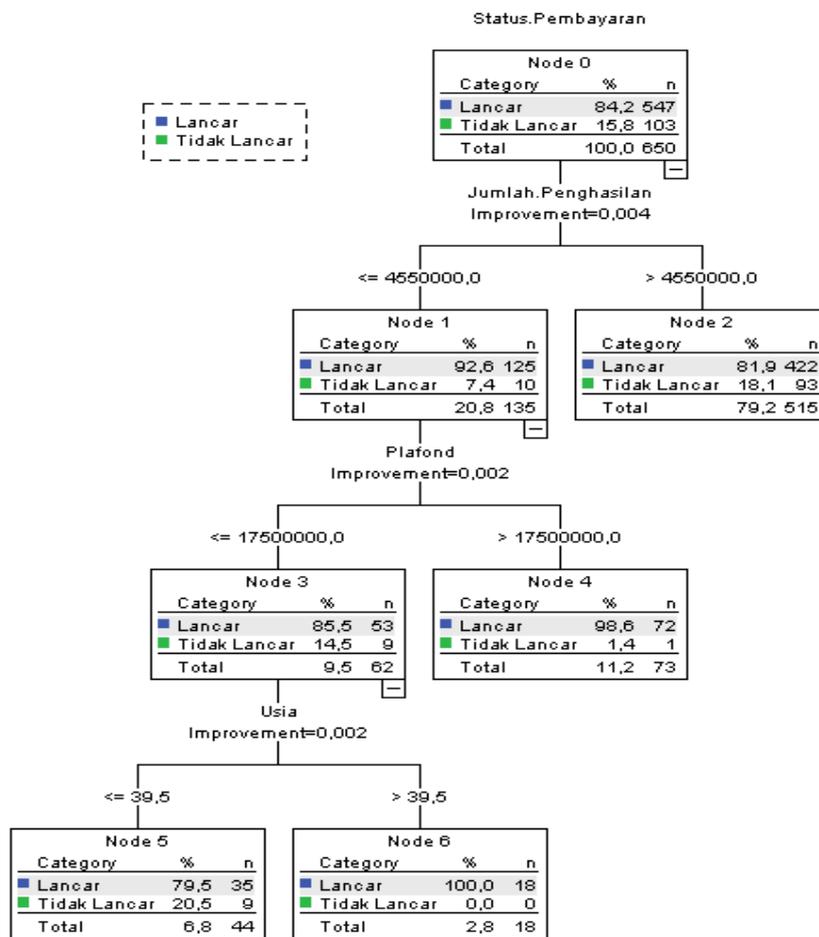
$$P(j_0|t) = \max_j P(j|t) = \max_j \frac{N_j(t)}{N(t)} \tag{4}$$

dengan $P(j|t)$ adalah proporsi kelas j pada simpul t , $N_j(t)$ adalah jumlah pengamatan pada kelas j pada simpul t , dan $N(t)$ adalah jumlah pengamatan pada simpul t . Label kelas simpul terminal t adalah j_0 yang memberikan nilai dugaan pengklasifikasian simpul t terbesar [1].

3. Hasil Dan Pembahasan

3.1 Analisis Pohon Klasifikasi *CART*

Berdasarkan pohon klasifikasi yang terbentuk menghasilkan 7 simpul yang terdiri atas 1 simpul asal, 2 simpul dalam, dan 4 simpul terminal dengan kedalaman pohon sebesar 3 serta peubah penjelas yang masuk ke dalam pohon klasifikasi adalah penghasilan, plafond, dan usia pada Gambar 2.



Gambar 2. Pohon klasifikasi *CART* nasabah bank

Berdasarkan pohon klasifikasi yang terbentuk pada Gambar 2, pemilah dengan nilai *improvement* tertinggi digunakan sebagai pemilah awal dalam pohon klasifikasi *CART*. Dapat dijelaskan bahwa pemilah awal pada simpul asal (simpul 0) adalah penghasilan nasabah yaitu nasabah yang memiliki penghasilan maksimum Rp 4.550.000 pada simpul 1 dan nasabah yang

memiliki penghasilan lebih dari Rp 4.550.000 pada simpul 2 dengan penurunan *improvement* sebesar 0,004.

Selanjutnya pada simpul 1 menghasilkan pemilah berdasarkan plafond nasabah yaitu nasabah yang memiliki plafond (pinjaman) maksimum Rp17.500.000 pada simpul 3 dan nasabah yang memiliki plafond (pinjaman) lebih dari Rp 17.500.000 pada simpul 4 dengan nilai *improvement* sebesar 0,002. Dan nasabah yang memiliki palfond (pinjaman) pada simpul 3 dipilah kembali berdasarkan usia nasabah yaitu nasabah yang memiliki usia maksimum 39,5 tahun pada simpul 5 dan nasabah yang memiliki usia lebih dari 39,5 tahun pada simpul 6 dengan nilai *improvement* sebesar 0,002.

Dan ketepatan prediksi hasil klasifikasi yang terbentuk dari pohon klasifikasi *CART* adalah sebesar 84,2 % pada Tabel 1.

Tabel 1. Ketepatan prediksi hasil klasifikasi

Observed	Predicted		
	Lancar	Tidak Lancar	Percent Correct
Lancar	547	0	100,0%
Tidak Lancar	103	0	,0%
Overall Percentage	100,0%	,0%	84,2%

Berikut ini merupakan Tabel 2 yaitu pembagian segmentasi dari nasabah bank yang diperoleh berdasarkan Gambar 2.

Tabel 2. Pembagian segmentasi nasabah bank

Segmen Ke-	Kriteria Nasabah Bank
1	Nasabah dengan jumlah penghasilan lebih dari Rp 4.550.000
2	Nasabah dengan jumlah penghasilan maksimum Rp 4.550.000 dengan plafond lebih dari Rp 17.500.000
3	Nasabah yang mempunyai jumlah penghasilan maksimum Rp4.550.000 dengan plafond maksimum Rp 17.500.000 adalah nasabah yang berusia maksimum 39,5 tahun.
4	Nasabah yang mempunyai jumlah penghasilan maksimum Rp4.550.000 dengan plafond maksimum Rp 17.500.000 adalah nasabah yang berusia lebih dari 39,5 tahun.

Segmentasi yang terbentuk terdiri atas 4 segmen dengan kriteria nasabah bank yang berbeda-beda. Segmentasi tersebut dapat ditabulasikan jumlah dan persentase masing-masing nasabah bank dengan status lancar dan tidak lancar sebagaimana disajikan pada Tabel 3.

Tabel 3. Distribusi Nasabah Bank Menurut Status Kredit Berdasarkan Segmen yang Telah Terbentuk

Segmen Ke-	Status Kredit			
	Lancar		Tidak Lancar	
	Jumlah	Presentase	Jumlah	Presentase
1	422	81,9%	93	18,1%
2	72	98,6%	1	1,4%
3	35	79,5%	9	20,5%
4	18	100%	0	0,0%

Dapat disimpulkan bahwa persentase nasabah bank yang memiliki status kredit lancar berdasarkan Tabel 2 dan Tabel 3 adalah pada segmen ke-4 yaitu nasabah yang mempunyai jumlah penghasilan maksimum Rp 4.550.000, dengan plafond maksimum sebesar Rp 17.500.000 dan berusia lebih dari 39,5 tahun dengan persentase 100%. Pada segmen ke-2 yaitu nasabah dengan jumlah penghasilan maksimum Rp 4.550.000 dengan plafond lebih dari Rp 17.500.000 dengan persentase 98,6%.

Persentase nasabah bank yang memiliki status kredit tidak lancar terdapat pada segmen ke-3 yaitu nasabah yang mempunyai jumlah penghasilan maksimum Rp 4.550.000, dengan plafond maksimum sebesar Rp 17.500.000 dan berusia maksimum 39,5 tahun yaitu dengan persentase 20,5%. Pada segmen ke-1 yaitu nasabah dengan jumlah penghasilan lebih dari Rp 4.550.000 dengan persentase 18,1%.

4. Kesimpulan

Berdasarkan hasil analisis dan pembahasan, didapat ketepatan klasifikasi sebesar 84,5% menggunakan analisis *CART* dengan 3 peubah penjelas yang mempengaruhi status pembayaran kredit pada nasabah bank yaitu jumlah penghasilan, plafond dan usia. Analisis *CART* membagi nasabah bank menjadi 4 segmen. Berdasarkan segmentasi yang terbentuk dari pohon klasifikasi, karakteristik nasabah yang cenderung mengalami pembayaran kredit berstatus tidak lancar adalah:

1. Nasabah yang mempunyai jumlah penghasilan maksimum Rp 4.550.000 dengan plafond maksimum Rp 17.500.000 adalah nasabah yang berusia maksimum 39,5 tahun dengan persentase 20,5%.
2. Nasabah dengan jumlah penghasilan lebih dari Rp 4.550.000 dengan persentase 18,1%.
3. Nasabah dengan jumlah penghasilan maksimum Rp 4.550.000 dengan plafond lebih dari Rp 17.500.000 dengan persentase 1,4%.

Ucapan Terima Kasih

Dalam penyusunan tulisan ini, banyak pihak yang telah memberikan dukungan kepada penulis. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan terima kasih kepada seluruh dosen dan pimpinan Jurusan Statistika Institut Sains & Teknologi AKPRIND Yogyakarta.

DAFTAR PUSTAKA

- [1] Breiman L, Friedman JH, Olshen RA & Stone CJ, 1984, *Classification and Regression Trees*, Chapman and Hall, New York.
- [2] Damayanti LK, 2011, *Aplikasi Algoritma CART untuk Mengklasifikasikan Data Nasabah Asuransi Jiwa Bersama Bumiputera 1912 Surakarta*, Skripsi Universitas Sebelas Maret, Surakarta.
- [3] Du Toit SHC, Steyn AGW & Stumph RH, 1986, *Graphical Exploratory Data Analysis*, Springer-Verlag, New York.
- [4] Gio PU, 2013, *Aplikasi Statistika dalam SPSS*, USU Press, Sumatera Utara.
- [5] Gordon L, 2013, *Using Classification and Regression Trees (CART) in SAS® Enterprise Miner™ For Applications in Public Health*, University of Kentucky, Lexington KY.
- [6] Margasari A, *Penerapan Metode CART (Classification And Regression Trees) dan Analisis Regresi Logistik Biner Pada Klasifikasi Profil Mahasiswa FMIPA Universitas Brawijaya*, Jurnal Universitas Brawijaya, Malang.
- [7] McClave JT & Terry Sincich, 2006, "*Statistics*", *tenth edition*, Pearson Prentice Hall, New Jersey.
- [8] Webb P & Yohannes, I, 1999, *Classification and Regression Trees, CART™*, International Food Policy Research Institute, Washington D.C.