

Model Credit Scoring Menggunakan Metode Classification and Regression Trees (CART) pada Data Kartu Kredit

¹Rifani Yunindya, ²Abdul Kudus, ³Teti Sofia Yanti

^{1,2,3}Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung,
Jl. Tamansari No.1 Bandung 40116

email: ¹rifaniyunindya23@gmail.com, ²akudus69@yahoo.com, ³tetisofiyanti@gmail.com

Abstract. Credit scoring is a tool and prediction technique that helps financial institutions to lend. The purpose of credit scoring is to assign prospective customers or customers to one group of "good customer" or "bad customer". One method that can be used to evaluate credit scoring is Classification and Regression Trees (CART). Classification and Regression Trees (CART) is a statistical method used to perform classification analysis. This paper discusses how to model credit scoring using the Classification and Regression Trees (CART) method. The calculation of credit scoring data is based on the credit history data of the customer. In this paper the data used are credit card customer payment data from April 2005 to September 2005 in Taiwan. The influential independent variables amount of bill statement in April (X_{17}), amount paid in May, 2005 (X_{22}), the repayment status in May, 2005 (X_{10}), the repayment status in July, 2005 (X_{20}), the repayment status in Agustus, 2005 (X_7) And the repayment status in September, 2005 (X_6). In this method the classification of credit customers by Classification and Regression Trees (CART) method gives 78.4 percent classification accuracy for training data and 78.6 percent for data testing.

Keywords: Credit scoring, Classification and Regression Trees, Credit card.

Abstrak. *Credit scoring* merupakan suatu alat dan teknik prediksi yang membantu lembaga keuangan dalam pemberian kredit. Tujuan dari *credit scoring* yaitu untuk menetapkan calon nasabah atau nasabah ke salah satu kelompok yaitu "nasabah yang baik" atau "nasabah macet". Salah satu metode yang dapat digunakan untuk mengevaluasi *credit scoring* yaitu *Classification and Regression Trees* (CART). *Classification and Regression Trees* (CART) adalah metode statistik yang digunakan untuk melakukan analisis klasifikasi. Makalah ini membahas cara memodelkan *credit scoring* dengan menggunakan metode *Classification and Regression Trees* (CART). Perhitungan data *credit scoring* didasarkan pada data riwayat pembayaran kredit nasabah. Dalam makalah ini data yang digunakan adalah data pembayaran pelanggan kartu kredit pada bulan April 2005 sampai dengan bulan September 2005 di Taiwan. Variabel bebas yang berpengaruh adalah jumlah tagihan bulan April 2005 (X_{17}), jumlah pembayaran bulan Mei 2005 (X_{22}), riwayat pembayaran bulan Mei 2005 (X_{10}), jumlah pembayaran bulan Juli 2005 (X_{20}), riwayat pembayaran bulan Agustus 2005 (X_7), dan riwayat pembayaran bulan September 2005 (X_6). Dalam metode ini pengklasifikasian nasabah kredit dengan metode *Classification and Regression Trees* (CART) menghasilkan ketepatan klasifikasi sebesar 78.4 persen untuk data *training* dan 78.6 persen untuk data *testing*.

Kata kunci: Credit scoring, Classification and Regression Trees, Kartu Kredit.

A. Pendahuluan

Kegiatan pinjam-meminjam uang telah dilakukan sejak lama dalam kehidupan masyarakat. Keputusan evaluasi risiko kredit sangat penting bagi lembaga keuangan karena risiko tinggi yang terkait dengan keputusan kredit yang tidak pantas. *Credit scoring* merupakan suatu alat prediksi yang membantu lembaga keuangan dalam pemberian kredit (Rezac, 2011). Oleh karena itu, model *credit scoring* dengan berbagai macam teknik telah dikembangkan oleh bank dan para peneliti untuk memecahkan masalah yang terlibat selama proses evaluasi. Tujuan dari model *credit scoring* adalah untuk mengelompokkan "nasabah yang baik" yaitu kelompok nasabah yang akan melunasi pinjaman atau "nasabah macet" yaitu kelompok nasabah yang kemungkinan besar tidak akan melunasi pinjaman.

Pada awalnya, lembaga keuangan biasanya mengadopsi aturan atau prinsip-prinsip yang dibangun oleh analis untuk memutuskan siapa yang harus diterima atau ditolak ajuan kreditnya. Sangat dibutuhkan proses otomatis untuk evaluasi kredit.

Seiring dengan pertumbuhan volume kredit, banyak model penilaian kredit yang telah dikembangkan oleh pihak bank maupun para peneliti dalam rangka untuk mengevaluasi calon nasabah menggunakan metode *credit scoring*. Model tersebut antara lain menggunakan Analisis Diskriminasi Linear (ADL), Analisis Regresi Logistik (ARL), *Multiplicative Adaptive Regression Spline* (MARS), *Artificial Neural Network* (ANN). *Classification and Regression Trees* (CART) merupakan metode klasifikasi yang berstruktur pohon. Berdasarkan kelebihan dan kelemahan dari metode CART maka penulis memilih menggunakan metode CART.

Metode CART adalah metode klasifikasi berstruktur pohon yang diperkenalkan oleh Leo Breiman, et al (1984). Bagi sekumpulan data yang terdiri dari p buah variabel independen dan sebuah variabel dependen, maka ketika yang dimiliki bertipe kategorik CART menghasilkan pohon klasifikasi (*classification trees*), sedangkan jika variabel dependen yang dimiliki bertipe kontinu atau numerik maka CART menghasilkan pohon regresi (*regression trees*). Tujuan dari CART adalah mengklasifikasikan suatu kelompok observasi atau sebuah observasi ke dalam suatu sub kelompok dari kelas-kelas yang diketahui.

Berdasarkan uraian dari latar belakang di atas, maka dari makalah ini adalah untuk mengetahui cara memodelkan *credit scoring* berdasarkan metode CART.

B. Landasan Teori

Classification and Regression Trees

CART (*Classification And Regresion Trees*) adalah metode statistik yang digunakan untuk melakukan analisis klasifikasi. CART pertama kali diperkenalkan pada tahun 1984 oleh empat ilmuwan Amerika serikat yaitu Leo Breiman, Jerome H. Friedman, Richard A. Olshen, dan Charles J. Stone. Tujuan dari CART adalah untuk mengklasifikasikan obyek menjadi dua atau lebih kelompok. CART terdiri dari dua analisis yaitu *classification trees* dan *regression trees*. Bagi sekumpulan data yang terdiri dari p buah variabel independen dan sebuah variabel dependen, maka ketika yang dimiliki bertipe kategorik CART menghasilkan pohon klasifikasi (*classification trees*), sedangkan jika variabel dependen yang dimiliki bertipe kontinu atau numerik maka CART menghasilkan pohon regresi (*regression trees*).

Teknik atau proses kerja dari CART dalam membuat sebuah pohon klasifikasi dikenal dengan dengan istilah *Binary Recursive Partitioning*. Proses disebut *binary* karena setiap simpul data (*node*) akan selalu mengalami pemisahan kedalam dua *child node*. Sedangkan *recursive* berarti bahwa proses pemisahan tersebut diulang kembali pada setiap *child node* sebagai hasil pemisahan terdahulu, sedangkan *child node* tersebut sekarang menjadi *parent node*. Proses pemisahan ini akan terus dilakukan sampai tidak ada kesempatan lagi untuk melakukan pemisahan berikutnya. Dan istilah *Partitioning* berarti bahwa testing sample yang dimiliki dipisah kedalam bagian-bagian atau partisi-partisi yang lebih kecil (Lewis, R.J, 2000).

Algoritma CART merupakan prosedur rekursif, mulai dari *root node* dan pada setiap *node* internal dipilih variabel independen tunggal dan nilai pemisahan point s untuk membagi sekelompok individu menjadi 2 kelompok individu yang masing-masing menjadi *child node*. CART yang menghasilkan pohon yang besar disebut *maximal tree*, dimana terminal *nodenya* tidak dapat dibagi lebih lanjut.

CART memberikan beberapa kemungkinan metode pemisahan seperti Entropi, Gini, dan Twoing. Setiap pilihan metode pemisahan dapat diadopsi bersama dengan struktur biaya kesalahan klasifikasinya.

Proses Pemisahan Node

Salah satu fungsi *impurity* utama yang digunakan dalam metode CART untuk mengukur tingkat keragaman variabel dependen dalam suatu *node* adalah: Indeks keragaman Gini pada *cost* yang sama

$$d(t) = \sum_{j \neq i} p(j|t)p(i|t) = (1 - \sum_j p^2(j|t)) \quad \dots(1)$$

Untuk mengukur tingkat keragaman variabel tak bebas dalam suatu *node* menggunakan Indeks keragaman Gini pada *cost* yang berbeda:

$$d(t) = (1 - \sum p^2(j|t))[C(j|i) + C(i|j)] \quad \dots(2)$$

Goodness of split merupakan sebuah evaluasi pemilihan oleh pemisah s pada *node* t . Jika suatu pemisah s dalam *node* t akan membagi data ke dalam t_R (dengan proporsi banyaknya objek dalam t_R adalah p_R), dan t_L (dengan proporsi banyaknya objek dalam t_L adalah p_L), maka didefinisikan *decrease impurity* (pengurangan keragaman) (Breiman et al, 1993)

$$\Delta d(s, t) = d(t) - p_L d(t_L) - p_R d(t_R) \quad \dots(3)$$

Suatu pemisah s akan digunakan untuk memisah *node* t menjadi dua buah *node* yaitu *node* t_R dan t_L jika s memaksimalkan nilai

$$\Delta d(s^*, t) = \max(s) \Delta d(s, t) \quad \dots(4)$$

Pelabelan Kelas

Pelabelan kelas adalah proses pengidentifikasian tiap *node* pada suatu kelas tertentu. Pelabelan kelas tidak hanya diberlakukan untuk *terminal node* saja, *non-terminal node* bahkan *root node* mengalami proses ini. Hal ini dikarenakan setiap *non-terminal node* memiliki kesempatan untuk menjadi *terminal node*. Sehingga proses pelabelan kelas akan terus dilakukan selama proses pemisahan masih berlanjut.

Pelabelan tiap *terminal node* berdasarkan aturan jumlah anggota kelas terbanyak yaitu:

$$p(j|t) = \max_i p(i|t) = \max_i \frac{n_i(t)}{n(t)} \quad \dots(5)$$

maka label kelas untuk *terminal node* t adalah j .

Proses Penghentian Pemisahan

Menurut Vayssieres (Vayssieres et al, 2000), proses pemisahan atau pembuatan pohon klasifikasi akan berhenti apabila sudah tidak dimungkinkan lagi dilakukan proses pemisahan atau bisa dikatakan sebuah *node* tidak dapat dibagi lebih lanjut. Sebuah *node* tidak dapat dibagi lebih lanjut jika:

1. Hanya terdapat sebuah nilai variabel dependen pada *node* tersebut,
2. Jika dipisah tidak ada penurunan tingkat keragaman antara keragaman *parent node* dan dua *node* anaknya, atau
3. Jika pada *node* tersebut hanya berisi 1 kasus.

Proses Pemangkasan Pohon

Metode yang digunakan dalam proses pemangkasan pohon adalah berdasarkan kriteria minimal *cost complexity pruning*.

$$R(T) = \sum_{t \in T} r(t)p(t) = \sum_{t \in T} R(t) \quad \dots(6)$$

$R(T)$ adalah *tree misclassification cost*, sedangkan $R(t)$ disebut *node misclassification cost*, $p(t)$ adalah proporsi yang masuk dalam *node* t , dan $r(t)$ adalah probabilitas terjadinya kesalahan klasifikasi di dalam sebuah *node* t tertentu yang didefinisikan sebagai berikut:

$$r(t) = 1 - \max_j p(j|t) \quad \dots(7)$$

Untuk subtree $T < T_{max}$ didefinisikan kompleksitas dari subtree ini adalah $|T^t|$, yaitu banyaknya *terminal node* yang dimiliki T . $\alpha \geq 0$ adalah *node complexity* dan $R_\alpha(T)$ adalah *cost complexity measure* (ukuran ongkos kompleksitas), maka:

$$R_\alpha(T) = R(T) + \alpha|T^t| \quad \dots(8)$$

Proses pemangkasan pohon klasifikasi dimulai dengan mengambil t_R (*right child node*) dan t_L (*left child node*) dari T_{max} yang dihasilkan dari *parent node* t . Jika diperoleh dua *child node* dan *parent node* yang memenuhi persamaan

$$R(t) = R(t_R) + R(t_L) \quad \dots(9)$$

maka *child node* t_R dan t_L dipangkas.

Hasilnya adalah pohon T_1 yang memenuhi kriteria $R(T_1) = R(T_{max})$. Proses tersebut diulang sampai tidak ada lagi pemangkasan yang mungkin terjadi.

Pohon Klasifikasi Optimal

Dalam menggambarkan struktur data, pohon klasifikasi yang terbentuk bisa saja berukuran besar dan sangat kompleks. Sehingga perlu dipilih pohon optimal yang berukuran sederhana tetapi memberikan nilai penduga pengganti yang cukup kecil. Pohon optimal dihasilkan dari pemangkasan.

Test Sample Estimate

Test sample digunakan jika ukuran data besar (\mathcal{L} besar). *Test sample* uji dibagi menjadi dua himpunan data. Misal \mathcal{L}_1 sebanyak ⁽¹⁾ objek sebagai *training* dan \mathcal{L}_2 sebanyak ⁽²⁾ objek sebagai *testing*. Buatlah T_{max} menggunakan \mathcal{L}_1 dan pangkas hingga diperoleh $T_k = T_1 > T_2 > T_3 > \dots > \{t\}$. Gunakan \mathcal{L}_2 pada masing-masing *tree* dan hitunglah berapa banyak objek yang mengalami kesalahan klasifikasi (*misclassification*).

Jika probabilitas *prior* di estimasi dari data sampel diperoleh $\pi_j = \frac{N_j^{(2)}}{N^{(2)}}$ maka bisa dihitung oleh

$$R^{ts}(T) = \frac{1}{N^{(2)}} \sum_{i,j} C(i|j) N_{ij}^{(2)} \quad \dots(10)$$

Test sample estimate dapat digunakan untuk memilih pohon klasifikasi yang optimum dari $\{T_k\}$, misal *tree optimum* adalah T_{k0}

$$R^{ts}(T_{k0}) = \min_k R^{ts}(T_k)$$

Ketepatan Klasifikasi

Pohon klasifikasi optimal yang telah terbentuk dilakukan evaluasi dari hasil klasifikasi. Cara untuk mengevaluasi hasil klasifikasi adalah dengan menghitung akurasi klasifikasi *sensitivity* menggambarkan akurasi pada sampel kelas i , sedangkan *specificity* menggambarkan bagaimana akurasi pada sampel kelas j . G-means dapat menggambarkan bagaimana sebuah metode klasifikasi mampu mengukur *sensitivity* dan *specificity*.

Tabel 1. Struktur Data Hasil Klasifikasi Optimal

	Prediksi		Total
	Kelas 1	Kelas 2	
Observasi			
Kelas 1	n_{11}	n_{12}	n_1
Kelas 2	n_{21}	n_{22}	n_2

$$\text{Tingkat akurasi total} = 1 - (\text{Total Tingkat Kesalahan}) = \frac{n_{11}+n_{22}}{n} \quad \dots(11)$$

$$\text{Total Tingkat Kesalahan} = \frac{n_{12}+n_{21}}{n} \quad \dots(12)$$

$$Se = \frac{n_{11}}{n_1} \quad \dots(13)$$

$$Sp = \frac{n_{22}}{n_2} \quad \dots(14)$$

Kredit Perbankan dan Kredit Bermasalah

Menurut Triandaru (2006), kredit merupakan pemberian fasilitas pinjaman kepada nasabah, baik berupa fasilitas pinjaman tunai (*cash loan*) maupun pinjaman nontunai (*non-cash loan*). Pemberian kredit, dalam pengertian sebagai *cash loan* merupakan salah satu bentuk usaha yang dilakukan oleh sebuah bank. Berdasarkan UU Nomor 10 tahun 1998 tentang pembahasan UU Nomor 7 tahun 1992 tentang perbankan, yang dimaksud dengan kredit 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. Dalam memberikan kredit atau pembiayaan, bank umum wajib mempunyai keyakinan berdasarkan analisis yang mendalam atas itikad dan kemampuan serta kesanggupan nasabah debitur untuk melunasi utangnya dan mengembalikan pembiayaan yang dimaksud sesuai dengan perjanjian yang disepakati.

Dalam pelepasan kartu kredit selalu terkandung risiko yang akan ditanggung oleh bank. Risiko-risiko yang berkaitan dengan pelepasan kredit nasabah dapat dikelompokkan ke dalam 4 kelompok yaitu *Credit Risk*, *Liquidity Risk*, *Price Risk*, dan *Prepayment Risk* (Prasetya, 2006).

C. Hasil Penelitian dan Pembahasan

Pembersihan Data

Data mengenai kartu kredit ini ada sebanyak 30000 data. Dalam data X3 (Pendidikan) terdapat 5 kategori, tetapi kategori 5 tidak diketahui keterangannya maka kategori 5 tidak dipakai. Dalam data X6 sampai dengan X11 (riwayat pembayaran) terdapat 12 kategori yaitu -2, -1, 0, 1, 2, 3, 4, 5, 6, 7, 8, dan 11, tetapi kategori -2 dan 0 tidak diketahui keterangannya maka kategori -2 dan 0 tidak dipakai. Sehingga data yang dipakai sebanyak 4030 data.

Data Training dan Data Testing

Tahapan pertama adalah membagi data menjadi dua kelompok yaitu data *training* sebesar 95% dan data *testing* sebesar 5%.

Dengan demikian,

$$\text{Jumlah } training = \frac{4030}{0.95} = 3829$$

$$\text{Jumlah } testing = \frac{4030}{0.05} = 201$$

Analisis Classification and Regression Trees

Proses pembentukan pohon memakai software Statistica 8.

Proses Pemisahan Node

Sebelum dilakukan proses pemisahan *node*, terlebih dahulu memilih variabel pemisah terbaik dari 23 variabel tak bebas menggunakan indeks keragaman Gini. Suatu pemisah akan digunakan untuk memecah *node* t menjadi dua buah *node* yaitu *node* t dan

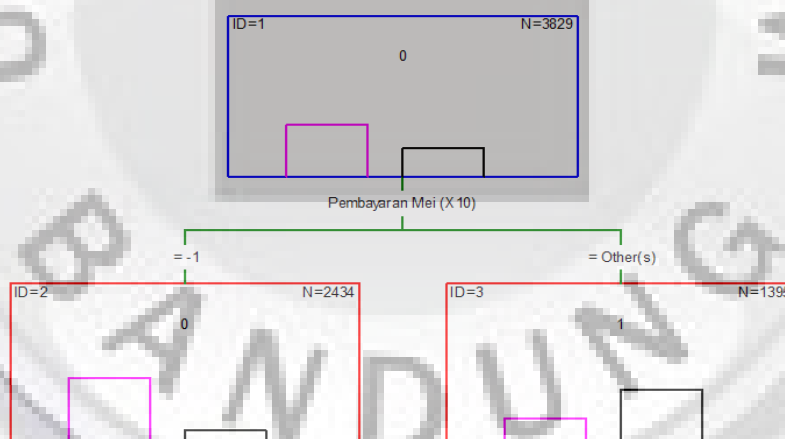
$node t_L$ jika s memaksimalkan nilai $\Delta i(s, t)$. Nilai *goodness of fit* dapat dilihat pada Tabel 2.

Tabel 2. *Goodness of Split* Variabel X_{10} (Riwayat Pembayaran Bulan April 2005)

No	Node Kiri	Node Kanan	<i>Goodness of Split</i>
1	-1	Lainnya	0.294
2	2	Lainnya	0.212
3	3	Lainnya	0.016
4	4	Lainnya	0.008
5	5	Lainnya	0.017
6	6	Lainnya	0.0002
7	7	Lainnya	0.016
8	8	Lainnya	0.0003

Sumber: Hasil Pengolahan Statistica 8

Setelah terbentuk dan terpilih kriteria pemisahan terbaik, maka *node* utama yang berisi 3829 objek data dipisah menjadi dua buah *terminal nodes*. Pemisah terbaik untuk *Root node* adalah variabel X_{10} (Riwayat Pembayaran pada Bulan April 2005) adalah dengan kriteria pemisahan -1 pada *node* kiri dan lainnya pada *node* kanan. Proses pemisahan dapat dilihat pada Gambar 4.1. Proses serupa terus berjalan pada *node-node* lainnya, sehingga terpenuhinya *stopping rule*. Adapun *stopping rule* yang digunakan adalah minimum jumlah sampel pada *node* tersebut 10% dari data atau semua objek yang berada pada dalam sebuah *node* merupakan anggota dari kelas yang sama.



Gambar 1. Pemilihan *Root Node* pada Pohon Klasifikasi Pertama

Proses Pelabelan Kelas

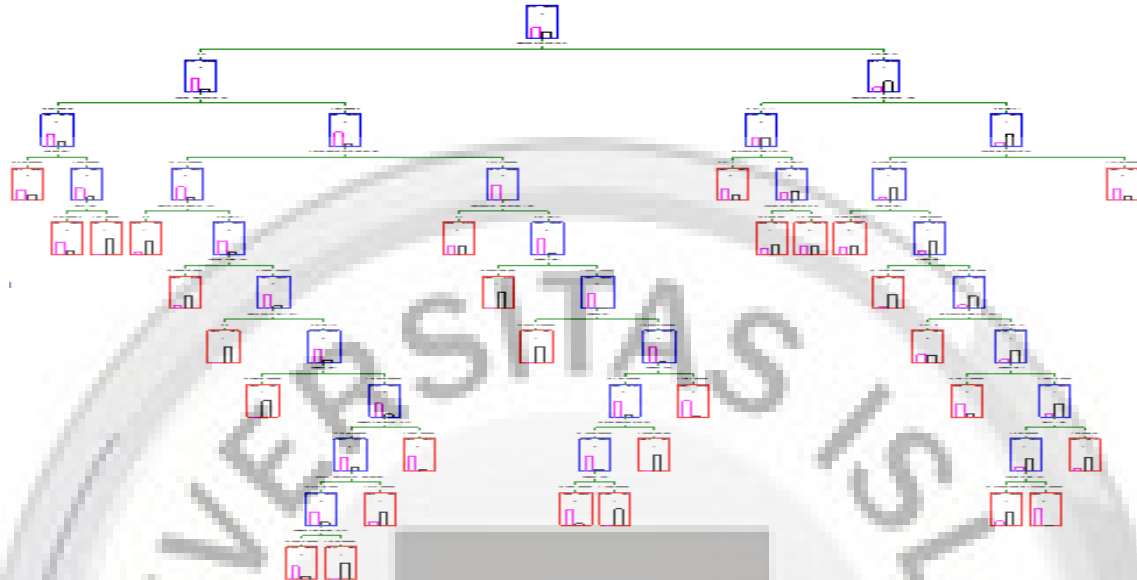
Proses pelabelan kelas pada *node-node* yang terbentuk berdasarkan aturan jumlah anggota kelas terbanyak yaitu jika $p(j_0|t) = \max_j p(j|t)$, maka label kelas untuk *terminal node* t adalah j . Sebagai contoh yaitu *node* 1 pada Gambar 1.

$$P(\text{Lancar}|\text{node } 1) = \frac{2462}{3829} = 0.643 \quad P(\text{Macet}|\text{node } 1) = \frac{1367}{3829} = 0.357$$

sehingga *root node* diberi label kelas “lancar”, karena peluang kelas “lancar” lebih besar dari peluang kelas “macet”.

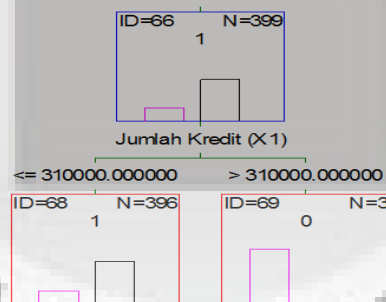
Proses Penghentian Pemisahan

Proses pemisahan *node* yang berulang-ulang akan berhenti setelah memenuhi *stopping rule* dan hasilnya berupa pohon maksimal. Pohon klasifikasi maksimal memiliki 28 *non-terminal node* dan 29 *terminal node*.



Gambar 2. Pohon Klasifikasi T (Pohon Klasifikasi Maksimal)

Proses penghentian pemisahan dapat dilihat pada contoh *node* 68 dan *node* 69 pada Gambar 2.



Gambar 3. Terminal Node Terakhir

Proses Pemangkasan Pohon

Pohon klasifikasi maksimal yang disajikan dalam Gambar 4.2 selanjutnya akan dipangkas untuk mendapatkan pohon yang optimal. Jika diperoleh dua *child node* dan *parent node* yang memenuhi persamaan $R(t) = R(t_R) + R(t_L)$, maka *child node* t_R dan t_L dipangkas.

Sebagai contoh proses pemangkasan dapat dilihat dari perhitungan dibawah ini, *Parent Node (Node 42)*:

Nilai $R(t) = r(t)p(t)$

Dengan $r(t) = 1 - \max_j p(j|t)$ dan $p(j|t)$ adalah peluang beberapa objek dalam *node* t .

Dalam *node* 42 terdapat dua kelas yaitu kelas macet dan kelas lancar. Nilai probabilitas tiap kelas dalam *node* 42 adalah sebagai berikut:

$$p(\text{macet}|\text{node 42}) = \frac{43}{496} = 0.087 \qquad p(\text{lancar}|\text{node 42}) = \frac{453}{496} = 0.913$$

sehingga dengan menggunakan rumus $r(t) = 1 - \max_j p(j|t)$, maka diperoleh

$$r(\text{node } 42) = 1 - 0.913 = 0.087$$

Nilai probabilitas objek yang berada dalam *node* 42 adalah

$$p(\text{node } 42) = \frac{N(\text{node } 42)}{N} = \frac{496}{3829} = 0.13$$

Oleh karena itu,

$$R(\text{node } 42) = r(\text{node } 42) p(\text{node } 42) = 0.087 * 0.13 = 0.011$$

Selanjutnya pada child node, yaitu terminal node 44 dan terminal node 45.

Dengan cara perhitungan yang sama maka, *terminal node* 44 memiliki nilai

$$R(t_L) = r(t_L) p(t_L) = 0.085 * 0.13 = 0.011$$

Untuk *terminal node* 45 memiliki nilai

$$R(t_R) = r(t_R) p(t_R) = 0 * 0.0002 = 0$$

Dengan demikian persamaan

$$R(t) = R(t_L) + R(t_R) \\ 0.011 = 0.011 + 0$$

Terpenuhi untuk *node* 42, sehingga dilakukan pemangkasan. Proses pemangkasan pohon klasifikasi maksimal menghasilkan 11 *subtrees*.

Pohon Klasifikasi Optimal

Dalam mencari pohon klasifikasi optimal digunakan metode *Test Sample Estimate*, karena ukuran data yang besar. Nilai penduga pengganti sample uji (*test sample estimate*) dari masing-masing *subtree* terlihat pada tabel 3.

Tabel 3. Test Set Relative Pohon Klasifikasi

Tree Number	Terminal Nodes	Test Set Relative	Resubstitution Cost	Node Complexity
Tree 1*	29	0.371	0.181028	0.000000
Tree 2	23	0.388	0.181818	0.000132
Tree 3	18	0.299	0.183300	0.000296
Tree 4	12	0.306	0.185277	0.000329
Tree 5**	10	0.289	0.186166	0.000445
Tree 6	8	0.316	0.187154	0.000494
Tree 7	5	0.301	0.189526	0.000791
Tree 8	4	0.299	0.191107	0.001581
Tree 9	3	0.299	0.193972	0.002866
Tree 10	2	0.303	0.216798	0.022826
Tree 11	1	1	0.270158	0.053360

Sumber: Hasil Pengolahan Statistica 8

*Pohon Maksimal

**Pohon Optimal

Pada Tabel 4.3 terlihat bahwa dari kesebelas *subtree* yang terbentuk, *subtree* nomor 5 dengan 10 *terminal nodes* adalah pohon klasifikasi optimal, hal ini dikarenakan memenuhi kriteria nilai *test set relative cost* $R^{ts}(T_{k0}) = \min_k R^{ts}(T_k)$.

Ketepatan Klasifikasi

Pohon klasifikasi optimal yang telah terbentuk dilakukan evaluasi dari hasil klasifikasi. Struktur data hasil klasifikasi optimal pada data *training* dan data *testing* dapat dilihat pada Tabel 4.

Tabel 4. Struktur Data *Training* dan Data *Testing* Hasil Klasifikasi Optimal

		Data <i>Training</i>		Data <i>Testing</i>	
		Prediksi (\hat{y})		Prediksi (\hat{y})	
		Lancar (0)	Macet (1)	Lancar (0)	Macet (1)
Observasi (y)	Lancar (0)	2234	228	122	10
	Macet (1)	600	767	33	36
Tingkat Akurasi Total		0.784		0.786	
Total Tingkat Kesalahan		0.216		0.214	
Sensitivity (se)		0.907		0.924	
Specificity (sp)		0.561		0.522	

Sumber: Hasil Pengolahan Statistica 8

Artinya pohon klasifikasi yang terbentuk dari data *training* mampu memprediksi dengan tepat pengamatan sebesar 78.4 persen, sedangkan pohon klasifikasi yang terbentuk dari data *testing* mampu memprediksi dengan tepat pengamatan sebesar 78.6 persen. Tingkat akurasi pada nasabah kredit yang tergolong lancar untuk data *training* sebesar 90.7 persen dan data *testing* 92.4 persen. Tingkat akurasi pada nasabah kredit yang tergolong macet untuk data *training* sebesar 56 persen dan data *testing* sebesar 52.2 persen.

Tingkat akurasi total pada data *training* dan data *testing* seimbang sehingga dapat dikatakan klasifikasi pohon optimal yang terbentuk sudah baik (Pratiwi, F.E dan Zain, I, 2014).

D. Kesimpulan dan Saran

Kesimpulan

Dalam skripsi ini telah dibahas mengenai pemodelan menggunakan metode *Classification and Regression Trees* (CART) pada data nasabah kredit. Berdasarkan hasil dan pembahasan dapat ditarik kesimpulan bahwa metode *Classification and Regression Trees* (CART) dapat diterapkan untuk mengklasifikasikan nasabah kredit. Hasil pengklasifikasian tersebut selanjutnya dapat digunakan misalnya untuk memberikan penawaran peningkatan limit kredit pada bulan selanjutnya bagi nasabah-nasabah yang diprediksi lancar, sehingga dapat meminimalisir terjadinya kerugian akibat dari pembayaran kartu kredit yang macet. Dalam penelitian ini, pengklasifikasian nasabah kredit dengan metode *Classification and Regression Trees* (CART) menghasilkan ketepatan klasifikasi sebesar 78.4 persen untuk data *training* dan 78.6 persen untuk data *testing*.

Saran

Saran yang dapat dikemukakan dalam penulisan skripsi ini adalah:

1. Jika akan memprediksi calon nasabah baru, sebaiknya menambahkan variabel lainnya dari nasabah tersebut. Variabel lainnya tersebut seperti pekerjaan, pendapatan, status kepemilikan rumah, jumlah anak, dan lain sebagainya.
2. Kepada peneliti lain, disarankan untuk meneliti menggunakan metode yang berbeda seperti CHAID, MARS, ataupun SVM. Hal tersebut bertujuan agar didapatkan ketepatan klasifikasi yang lebih tinggi.

Daftar Pustaka

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. (1984). *Classification and Regression Trees*. Wadsworth Inc.

- Damayanti, L. (2011). *Aplikasi Algoritma CART untuk Mengklasifikasikan Data Nasabah Asuransi Jiwa Bersama BUMIPUTERA 1912 Surakarta*. Program Sarjana, Fakultas Mipa, Universitas Sebelas Maret, Surakarta.
- Feldman, D dan Gross, S. (2005). *Mortgage Default: Classification Trees Analysis*. The Journal of Real Estate Finance and Economics, 369-396.
- Kudus, A., et al. (2009). *Decision tree for prognostic of multivariate survival data and competing*. Jurnal Recent Advances In Technologises, 1-33. cdn.intechopen.com
- Lee, T.S., et al. (2002). *Mining the customer credit using classification and regression tree and multivariate adaptive regression splines*. Computational Statistics & Data Analysis, 1113-1130.
- Marc, Vayssieres, Richard E. Plant, Barbara H. Allen-Diaz, *Classification Trees: An Alternative Non-Parametric Approach for Predicting Species Distributions*, Journal of Vegetation Science, Vol. 11, No.5, pp. 679-694, Blackwell Publishing, 2000.
- Negara, D. (2016). *Perbandingan Credit Scoring Yang Dihasilkan oleh Model Regresi Logistik dan Cox Proporsional Hazard dengan Menggunakan Kriteria Mean Cost*. Program Sarjana, Program Studi Statistika, Universitas Islam Bandung.
- Prasetya, Laniati. (2006). *Penerapan Metode Survival Analysis dan KPR (Studi Kasus pada PT. Bank ABC, Tbk)*, Program Pasca Sarjana, Magister Manajemen, Universitas Indonesia.
- Pratiwi, F.E dan Zain, I. (2014). *Klasifikasi Pengangguran Terbuka Menggunakan CART (Classification and Regression Tree) di Provinsi Sulawesi Utara*. Program Sarjana, Jurusan Statistika, FMIPA, Institut Teknologi Sepuluh Nopember (ITS).
- Rezac, M. (2011). *How to Measure the Quality of Credit Scoring Models*. *Jurnal of Economics and Finance*, 5.
- Webb, P., and I. Yohannes. (1999). *Classification And Regression Trees, CART*, International Food Policy Research Institute, Washington D.C.