

**PERBANDINGAN KLASIFIKASI NASABAH KREDIT  
MENGUNAKAN REGRESI LOGISTIK BINER DAN CART  
(CLASSIFICATION AND REGRESSION TREES)**

**Agung Waluyo<sup>1</sup>, Moch. Abdul Mukid<sup>2</sup>, Triastuti Wuryandari<sup>2</sup>**

<sup>1</sup>Alumni Jurusan Statistika FSM UNDIP

<sup>2</sup>Dosen Jurusan Statistika FSM UNDIP

**Abstract**

Credit is the greatest asset managed the bank and also the most dominant contributor to the bank's revenue. Debtors to pay their credit to the bank may smoothly or jammed. This study aims to identify the variables that affect a debtor's credit status and compare the acuration of classification method both classification and regression trees (CART) and logistic regression. The variables used were debtor's gender, education level, occupation, marital status, and income. By using logistic regression, it was known that only the debtor's income effect their credit status with the classification accuracy reach into 80%. By using CART, there were some variables affect the credit status and the classification accuracy 80,9%. This paper showed that the performance of CART in classifying the credit status of debtors was better than logistic regression.

**Keywords:** Credit Status, Logistic Regression, CART

## 1. Pendahuluan

Bank menurut Undang-undang RI nomor 10 tahun 1998 tanggal 10 November 1998 tentang perbankan adalah Badan usaha yang menghimpun dana dari masyarakat dalam bentuk simpanan dan menyalurkan kepada masyarakat dalam bentuk kredit dan atau bentuk-bentuk lainnya dalam rangka meningkatkan taraf hidup rakyat banyak. Bank merupakan perusahaan yang bergerak dalam bidang keuangan, artinya usaha perbankan selalu berkaitan dengan masalah di bidang keuangan. Kegiatan menghimpun dan menyalurkan dana merupakan kegiatan pokok perbankan, sedangkan kegiatan memberikan jasa-jasa bank hanyalah merupakan kegiatan pendukung. Fungsi utama bank dalam suatu perekonomian adalah untuk memobilisasi dana masyarakat, secara tepat dan cepat menyalurkan dana tersebut kepada pengguna atau investasi yang efektif dan efisien<sup>[4]</sup>.

Penerapan prinsip kehati-hatian oleh bank diantaranya diimplementasikan melalui kemampuan bank untuk mengelola portofolio kredit yang dimiliki sehingga resiko yang berpotensi untuk terjadi (*credit risk*) dapat diukur dan dikontrol. Kredit merupakan asset yang paling besar yang dikelola bank dan juga merupakan kontributor yang paling dominan terhadap pendapatan bank. Model skor kredit merupakan alat bantu dalam melakukan analisa kelayakan kredit berguna sebagai langkah awal dalam mengurangi resiko terjadinya kegagalan pemenuhan kewajiban oleh debitur. *Credit scoring* (skor kredit) adalah metode yang digunakan untuk mengevaluasi resiko kredit dalam hal permohonan pinjaman dari konsumen<sup>[6]</sup>.

Metode klasifikasi dapat dilakukan dengan pendekatan parametrik dan nonparametrik. Dalam pendekatan parametrik terdapat beberapa metode klasifikasi, salah satunya adalah analisis regresi logistik. Menurut Hosmer dan Lemeshow, metode regresi logistik adalah suatu metode analisis statistika yang mendeskripsikan hubungan antara variabel respon yang memiliki dua kategori atau lebih dengan satu atau lebih variabel penjelas berskala kategori atau interval<sup>[3]</sup>.

Regresi logistik biner adalah salah satu metode statistika yang sering digunakan untuk mengklasifikasikan sejumlah pengamatan dengan respon biner ke dalam beberapa kelompok berdasarkan satu atau lebih variabel prediktor. Melalui metode ini akan dihasilkan peluang dari masing-masing kategori respon yang akan dijadikan sebagai pedoman pengklasifikasian dan suatu pengamatan akan masuk ke dalam respon kategori tertentu berdasarkan nilai peluang yang terbesar.

Ada beberapa metode klasifikasi dengan pendekatan nonparametrik, salah satunya metode klasifikasi berstruktur pohon yang diperkenalkan oleh Leo Breiman, Jerome H. Friedman, Richard A. Olshen, dan Charles J. Stone [2]. Pada tahun 1984, keempat ilmuwan tersebut memperkenalkan metode klasifikasi CART (*Classification And Regression Trees*) yaitu metode pohon regresi dan pohon Klasifikasi. Jika variabel respon yang dimiliki bertipe kategorik maka CART menghasilkan pohon klasifikasi (*classification trees*). Sedangkan jika variabel respon yang dimiliki bertipe kontinu atau numerik maka CART menghasilkan pohon regresi (*regression trees*).

Pada tulisan ini akan dilakukan pengklasifikasian dengan variabel prediktor jenis kelamin, jenjang pendidikan, status pernikahan, jenis pekerjaan, dan pendapatan debitor. Disamping itu juga akan dilakukan perbandingan tingkat akurasi analisis regresi logistik biner dan CART dalam pengklasifikasian nasabah kredit di suatu bank.

## 2. Tinjauan Pustaka

### 2.1. Pengertian Kredit

Pengertian kredit dibakukan oleh pemerintah dengan dikeluarkan Undang-Undang No.7 Tahun 1992 tentang Perbankan sebagaimana telah diubah dengan Undang-Undang No.10 tahun 1998, yang mendefinisikan pengertian kredit adalah sebagai berikut (pasal 1 Ayat 12) adalah penyediaan uang atau tagihan yang dapat disamakan dengan bentuk kredit berdasarkan persetujuan atau kesepakatan pinjam-meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam untuk melunasi hutangnya setelah jangka waktu tertentu dengan jumlah bunga, imbalan atau pembagian hasil keuntungan.

### 2.2. Model Regresi Logistik Biner

Model regresi logistik biner digunakan untuk menganalisa hubungan antara satu variabel respon (variabel tak bebas) dan beberapa variabel bebas, dengan variabel responnya berupa data kualitatif dikotomi yaitu bernilai 1 untuk menyatakan keberadaan sebuah karakteristik dan bernilai 0 untuk menyatakan ketidakberadaan sebuah karakteristik<sup>[1]</sup>.

Jika diketahui  $Y$  variabel tak bebas bernilai 0 dan 1, maka

$$P(Y = 1|X = x) = \pi(x) \text{ dan } P(Y = 0|X = x) = 1 - \pi(x)$$

Sehingga diperoleh model regresi logistik:

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

dan logit dari  $\pi(x)$  adalah:

$$\ln \left( \frac{\pi(x)}{1-\pi(x)} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

### 2.3. CART (*Classification And Regression Trees*)

CART (*Classification And Regression Trees*) adalah metode statistik non parametrik 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. CART terdiri dari dua analisis yaitu

*classification trees* dan *regression trees*. Jika variabel yang dimiliki bertipe kategorik maka 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 induk akan selalu mengalami pemecahan kedalam tepat dua simpul anak. Sedangkan *recursive* berarti bahwa proses pemecahan tersebut diulang kembali pada setiap simpul anak sebagai hasil pemecahan terdahulu, sedangkan simpul-simpul anak tersebut sekarang menjadi simpul induk. Proses pemecahan ini akan terus dilakukan sampai tidak ada kesempatan lagi untuk melakukan pemecahan berikutnya. Dan istilah *Partitioning* berarti bahwa learning sample yang dimiliki dipecah kedalam bagian-bagian atau partisi-partisi yang lebih kecil<sup>[5]</sup>.

Pada dasarnya dalam membuat sebuah pohon klasifikasi, CART bekerja dalam empat langkah utama. Langkah pertama adalah *tree building process* yaitu proses pembentukan atau pembuatan sebuah pohon klasifikasi. Terdiri dari proses pemecahan simpul induk menjadi dua buah simpul anak melalui aturan pemecahan (*splitting rules*) tertentu dan dilakukan secara berulang serta proses *class assignment* yaitu proses mengidentifikasi simpul-simpul yang terbentuk pada suatu kelas tertentu melalui aturan pengidentifikasian. Langkah kedua adalah proses penghentian pembuatan atau pembentukan pohon klasifikasi (*stopping the trees building process*). Pada tahap ini pohon terakhir atau maximal tree ( $T_{max}$ ) telah terbentuk. Langkah ketiga adalah *pruning the tree* yaitu proses pemangkasan atau pemotongan  $T_{max}$  menjadi pohon yang lebih kecil. Langkah terakhir adalah interpretasi dan menghitung akurasi pohon.

### 2.3.1. Proses Pemecahan Simpul

Proses pemecahan pada masing-masing simpul induk didasarkan pada *goodness of split criterion* (kreteria uji pemecah terbaik). Dalam penelitian ini digunakan *twoing criterion*, yaitu:

$$\Phi(s, t) = \frac{P_L P_R}{4} \left[ \sum_j |p(j | t_L) - p(j | t_R)| \right]^2$$

dengan:

$\Phi(s, t)$  = nilai *goodness of split* di simpul  $t$  pada suatu pemecah  $s$

$t_L$  = simpul kiri dari simpul  $t$

$t_R$  = simpul kanan dari simpul  $t$

$P_L$  = peluang objek yang berada pada simpul kiri

$P_R$  = peluang objek yang berada pada simpul kanan

$p(j | t_L)$  = peluang bahwa sebuah objek adalah anggota kelas  $j$  jika diketahui objek ini berada dalam simpul kiri

$p(j | t_R)$  = peluang bahwa sebuah objek adalah anggota kelas  $j$  jika diketahui objek ini berada dalam simpul kanan

### 2.3.2. Pelabelan Kelas (*Class Assignment*)

Pelabelan kelas adalah proses pengidentifikasian tiap simpul-simpul dengan suatu kelas tertentu. Penandaan label tiap simpul akhir berdasarkan aturan jumlah anggota kelas terbanyak yaitu labeli simpul  $t$  dengan kelas  $j = j_0$  jika

$$p(j_0 | t) = \max_j p(j | t) = \max_j \frac{N_j(t)}{N(t)}$$

dengan

$p(j_0 | t)$  adalah proporsi kelas  $j_0$  pada simpul  $t$

$p(j | t)$  adalah proporsi kelas  $j$  pada simpul  $t$

$N_j(t)$  adalah jumlah pengamatan kelas  $j$  pada simpul  $t$

$N(t)$  adalah jumlah pengamatan pada simpul  $t$

### 2.3.3. Proses Penghentian Pemecahan

Proses pembuatan pohon klasifikasi akan berhenti apabila sudah tidak dimungkinkan lagi dilakukan proses pemecahan. Proses pemecahan akan berhenti apabila hanya tersisa satu objek saja yang ada di dalam simpul terakhir atau semua objek yang berada di dalam sebuah simpul merupakan anggota kelas yang sama (homogen). Pohon klasifikasi yang terbentuk sebagai hasil proses ini dinamakan “*maximal*” atau “*largest*” *tree* ( $T_{max}$ ).

### 2.3.4. Proses Pemangkasan Pohon

*Maximal tree* ( $T_{max}$ ) yang dihasilkan dari proses pembentukan pohon klasifikasi dapat mengakibatkan dua masalah.

1. Meskipun  $T_{max}$  memiliki tingkat akurasi yang tinggi dengan nilai *misclassification cost* yang kecil bahkan bernilai nol (dihitung dengan metode *resubstitution estimate*), namun  $T_{max}$  dapat menyebabkan *over-fitting* atau pencocokan nilai yang terlalu kompleks pada data-data baru. Hal ini dikarenakan terlalu banyaknya variabel independen yang digunakan untuk melakukan prediksi terhadap data baru tersebut.
2. Sebagai akibat dari poin 1,  $T_{max}$  akan sulit untuk dipahami atau diinterpretasi. Sehingga  $T_{max}$  akan lebih tepat dikatakan sebagai *complex tree* yang kompleksitasnya ditentukan dari banyaknya Simpul akhir yang dimiliki.

Proses *tree pruning* (pemangkasan pohon) dimaksudkan untuk menghilangkan kedua masalah tersebut dengan cara memotong atau memangkas  $T_{max}$  menjadi beberapa pohon klasifikasi ( $T$ ) yang ukurannya lebih kecil (*subtrees*). Untuk mempermudah dalam memahami proses *tree pruning* ini, berikut diberikan penjelasan awal mengenai proses ini.

Metode yang digunakan dalam proses pemangkasan pohon berdasarkan pada *minimal cost complexity pruning*, yaitu:

$$R(T) = \sum_{t \in \tilde{T}} r(t)P(t) = \sum_{t \in \tilde{T}} R(t)$$

dengan  $r(t) = 1 - \max_j P(j | t)$

$R(T)$  adalah *tree misclassification cost* atau *tree resubstitution cost* (proporsi kesalahan pada sub pohon),  $\tilde{T}$  adalah himpunan simpul akhir (*terminal node*) sedangkan  $r(t)$  disebut simpul *misclassification cost* (Brieman et al., 1984).  $P(t)$  adalah peluang sebuah objek akan berada dalam simpul  $t$ .

$$\begin{aligned} P(t) &= \sum_j^J P(j, t) = P(1, t) + P(2, t) + \dots + P(J, t) \\ &= \frac{N_1(t)}{N} + \frac{N_2(t)}{N} + \dots + \frac{N_j(t)}{N} = \frac{N(t)}{N} \end{aligned}$$

Jika didefinisikan  $P(j | t) =$  peluang bahwa sebuah objek adalah anggota kelas  $j$  dan jika diketahui objek ini berada dalam simpul  $t$ , maka diperoleh

$$P(j|t) = \frac{P(j, t)}{P(t)} = \frac{\frac{N_j(t)}{N}}{\frac{N(t)}{N}} = \frac{N_j(t)}{N(t)}$$

dengan

$\frac{N_j(t)}{N(t)}$  = proporsi objek-objek dalam kelas  $j$  yang berada di simpul  $t$

$P(j,t)$  = probabilitas bahwa sebuah objek adalah anggota kelas  $j$  dan berada dalam simpul  $t$ .

Proses pemangkasan pohon klasifikasi dimulai dengan mengambil  $t_R$  yang merupakan simpul anak kanan dan  $t_L$  yang merupakan simpul anak kiri dari simpul  $t$ . Jika  $R(t) = R(t_R) + R(t_L)$ , maka simpul anak  $t_R$  dan  $t_L$  dipangkas. Proses tersebut diulang sampai tidak ada lagi pemangkasan yang mungkin.

### 3. Metodologi Penelitian

#### 3.1. Sumber Data

Data yang digunakan pada penelitian ini bersumber dari data debitur di suatu bank pada tahun 2011, yang merupakan data debitur kredit lancar dan macet. Unit observasi dalam penelitian ini adalah pada regional Lampung. Data debitur tersebut berjumlah 1000 debitur.

#### 3.2. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah status kredit (lancar dan macet), jenis kelamin (laki-laki dan wanita), jenjang pendidikan (sarjana, diploma, SMA dan lainnya), status pernikahan (menikah, belum menikah dan cerai), jenis pekerjaan (pegawai swasta, pegawai BUMN/BUMD, pegawai negeri sipil dan lainnya) dan pendapatan (< 3000000, 3000000 - 5000000, > 5000000)

#### 3.3. Metode Penelitian

Metode analisis yang digunakan dalam penelitian ini diuraikan sebagai berikut :

1. Pengolahan data dengan analisis regresi logistik biner.
2. Mengintrepretasikan hasil analisis regresi logistik biner.
3. Menghitung nilai akurasi klasifikasi regresi logistik biner.
4. Pengolahan data menggunakan CART (*Classification And Regression Trees*)
5. Mengintrepretasikan hasil pohon klasifikasi algoritma CART
6. Menghitung nilai akurasi ketepatan klasifikasi pohon tersebut.
7. Membandingkan hasil langkah (3) dan (6)
8. Mengambil kesimpulan dari analisis yang dilakukan

### 4. Pembahasan

Pada bagian ini akan diidentifikasi variabel-variabel yang mempengaruhi status kredit seorang debitur menggunakan model regresi logistik biner dan model berstruktur pohon CART.

#### 4.1. Metode Regresi Logistik Biner

Setelah dilakukan uji signifikansi terhadap model, baik secara keseluruhan maupun individual diperoleh hasil bahwa variabel pendapatan berpengaruh signifikan terhadap variabel status kredit. Oleh karena itu, untuk memperoleh model akhir yang sesuai

dilakukan analisis regresi logistik biner kembali dengan tidak mengikutsertakan variabel yang tidak berpengaruh signifikan. Berikut adalah model terbaiknya

$$\pi(x) = \frac{e^{-1,260+0,587 \text{ pendapatan}(1)+ 0,325 \text{ pendapatan}(2)}}{1 + e^{-1,260+0,587 \text{ pendapatan}(1)+ 0,325 \text{ pendapatan}(2)}}$$

Setelah dilakukan uji signifikansi parameter selanjutnya akan dilakukan perhitungan klasifikasi untuk melihat ketepatan klasifikasi model. Berikut adalah hasil ketepatan klasifikasi data kredit nasabah bank:

**Tabel 1.** Tabel Ketepatan Klasifikasi Regresi Logistik Biner

$f_{ij}$		Kelas Hasil Prediksi ( $j$ )	
		Kelas lancar	Kelas macet
Kelas asli ( $i$ )	Kelas lancar	800	0
	Kelas macet	200	0

$$\text{Akurasi} = \frac{800+0}{800+0+200+0} \times 100 = 80,0\%$$

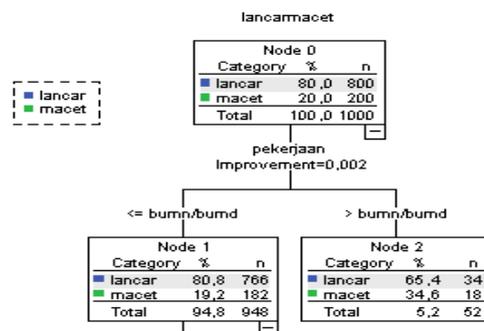
Berdasarkan Tabel 1 diketahui bahwa ketepatan klasifikasi dengan regresi logistik biner sebesar 80%.

#### 4.2 Analisis Status Kredit Bank dengan dengan Metode *Clasification And Regression Tress*

CART (*Clasification And regression Tress*) termasuk dalam anggota analisis klasifikasi yang disebut *decision tress*, karena proses analisis dari CART digambarkan dalam bentuk atau struktur yang menyerupai sebuah pohon, lebih tepatnya pohon klasifikasi berbentuk biner.

##### 4.2.1. Proses *Splitting Simpul*

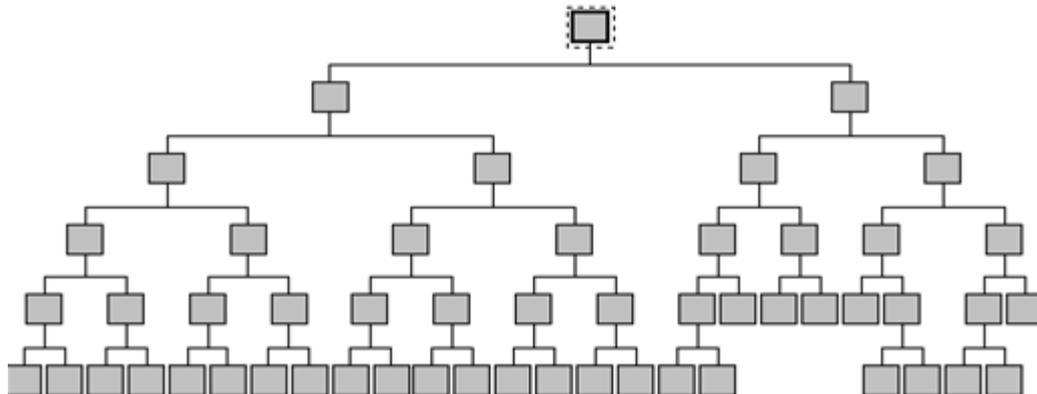
Berdasarkan nilai *goodness of split (towing criterion)*, variabel jenis pekerjaan terpilih sebagai pemilah terbaik yang digunakan untuk memecah simpul akar menjadi simpul  $t_L$  (pegawai swasta dan pegawai BUMD/BUMD) dan simpul  $t_R$  (pegawai negri sipil dan lainnya). Nilai *goodness of split* nya sebesar 0,0011716



**Gambar 1.** Variabel Pemilah Simpul Akar

Proses pemecahan simpul akar dapat dilihat pada Gambar 1. Proses serupa terus berjalan pada simpul-simpul lainnya. Proses pemilahan yang berulang-ulang akan berhenti apabila sudah tidak dimungkinkan lagi dilakukan proses pemilahan karena pada ujung

pohon klasifikasi terdapat simpul akhir yang memiliki anggota kelas yang sama (homogen). Pohon klasifikasi maksimum tampak seperti pada Gambar 2.



**Gambar 2.** Pohon Klasifikasi Maksimum

#### 4.2.2. Pelabelan Kelas (*Class Assignment*)

Pada bagian ini dibahas mengenai pemberian label kelas pada simpul-simpul yang telah terbentuk. Prosedur pemberian label sebuah simpul (utamanya simpul akhir) berdasarkan *class assignment rule*, yaitu jika  $P(j_0 | t) = \max_j P(j | t)$  maka simpul  $t$  tersebut diberi label  $j = j_0$ . Sebagai contoh, pada simpul akar:

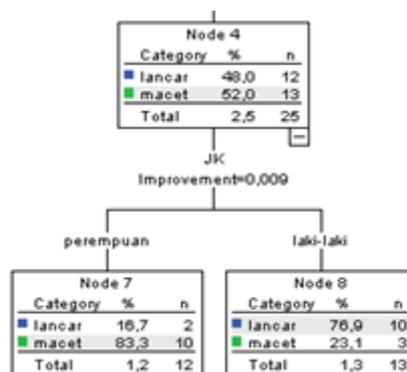
$$P(\text{kredit lancar} \mid \text{simpul } 0) = 800/1000 = 0,80$$

$$P(\text{kredit macet} \mid \text{simpul } 0) = 200/1000 = 0,20$$

sehingga simpul akar diberi label “kredit lancar”, karena peluang kelas kredit lancar lebih besar dari peluang kelas kredit macet. Proses pelabelan kelas ini berlaku pada semua simpul terutama simpul akhir, karena simpul akhir adalah simpul yang sangat penting dalam memprediksi suatu objek pada kelas tertentu jika objek berada pada simpul akhir tersebut.

#### 4.2.3. Proses Penghentian Pemecahan (*Stop The Split*)

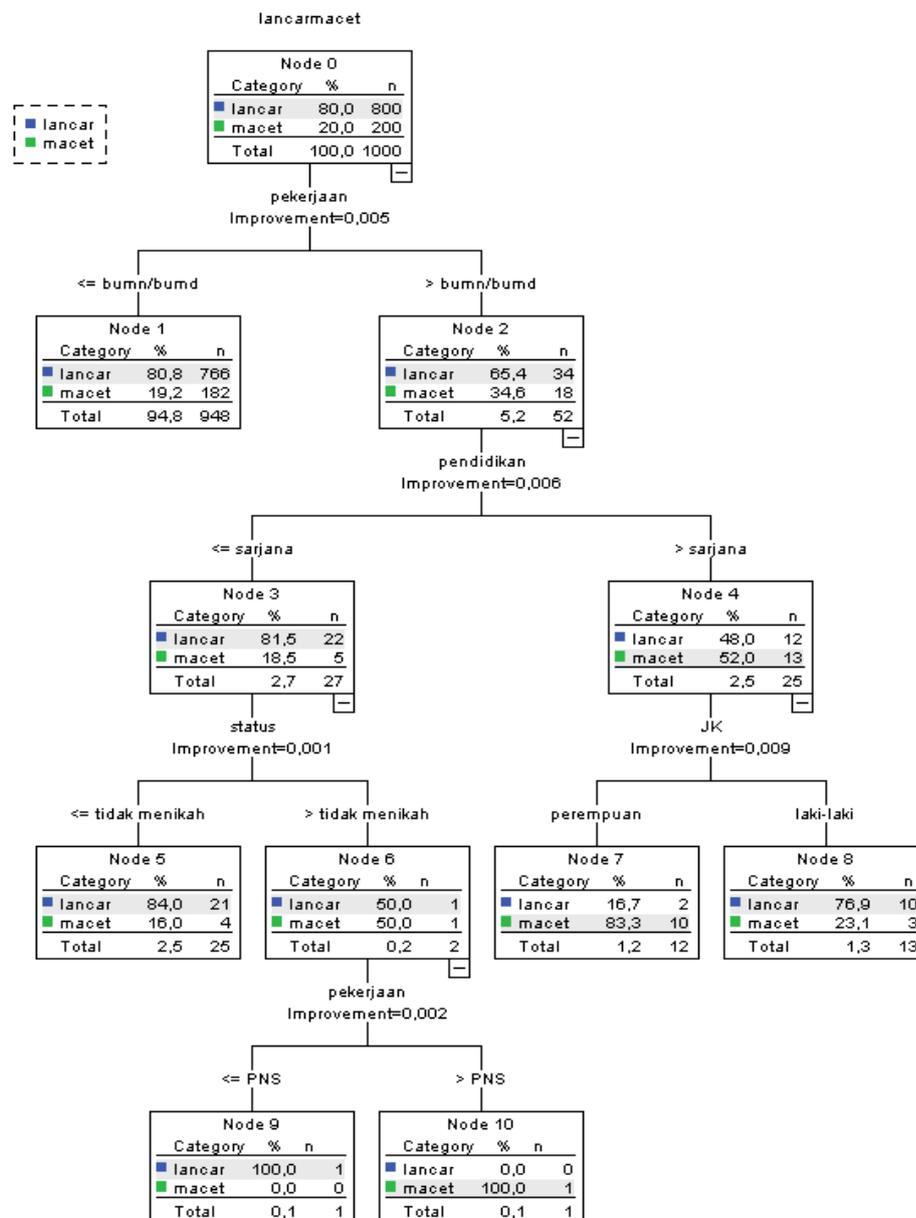
Proses pemecahan simpul yang berulang-ulang akan berhenti apabila sudah tidak dimungkinkan lagi dilakukan proses pemecahan. Proses *splitting* simpul berhenti jika pada ujung pohon klasifikasi simpul akhir hanya memiliki satu anggota atau sudah tidak ada lagi variabel yang menjadi pemecah. Gambar 3 berikut menampilkan beberapa simpul akhir dari pohon klasifikasi yang dihasilkan. Simpul 7 dan 8 menjadi simpul akhir karena sudah tidak ada lagi variabel yang menjadi pemecah.



**Gambar 3.** Simpul Akhir 7 dan 8

#### 4.2.4. Proses Pemangkasan Pohon ( *Trees Pruning Process* )

Pemangkasan (*pruning*) dilakukan untuk mengurangi kompleksitas pohon agar menjadi lebih sederhana. Dengan pemangkasan jumlah simpul akan menjadi berkurang sehingga jumlah simpul akhir juga akan berkurang. Proses pemangkasan pohon klasifikasi dimulai dengan mengambil  $t_R$  yang merupakan simpul anak kanan dan  $t_L$  yang merupakan simpul anak kiri dari  $T_{max}$  yang dihasilkan dari simpul induk  $t$ . Jika diperoleh dua simpul anak dan simpul induk yang memenuhi persamaan  $R(t) = R(t_R) + R(t_L)$ , maka simpul anak  $t_R$  dan  $t_L$  dipangkas. Proses tersebut diulang sampai tidak ada lagi pemangkasan yang mungkin dilakukan. Berikut ini pada Gambar 4 merupakan bentuk pohon klasifikasi setelah dipangkas.



Gambar 4. Pohon Klasifikasi Setelah Dipangkas

#### 4.2.5. Interpretasi Pohon Klasifikasi

Dari hasil pohon klasifikasi setelah proses pemangkasan, variabel pekerjaan merupakan variabel independen yang paling berpengaruh, sehingga menjadi pemilah terbaik dari simpul akar. Pohon klasifikasi ini memiliki 6 simpul akhir dengan profile sebagai berikut:

1. Simpul akhir 1 memiliki label status kredit lancar.  
Simpul akhir 1 ini memiliki profil debitur dengan pekerjaan sebagai pegawai swasta atau pegawai bumn/bumd dengan potensi lancar 80,8% dan potensi macet 19,2%.
2. Simpul akhir 5 memiliki label status kredit lancar.  
Simpul akhir 5 ini memiliki profil debitur dengan pekerjaan sebagai pegawai negeri sipil atau lainnya, berpendidikan sarjana dan belum menikah dengan potensi lancar 84% dan potensi macet 16%.
3. Simpul akhir 7 memiliki label status kredit macet.  
Simpul akhir 7 ini memiliki profil debitur dengan pekerjaan sebagai pegawai negeri sipil atau lainnya yang berpendidikan diploma, SMA atau lainnya berjenis kelamin perempuan dengan potensi lancar 16,7% dan potensi macet 83,3%.
4. Simpul akhir 8 memiliki label status kredit lancar.  
Simpul akhir 8 ini memiliki profil debitur dengan pekerjaan sebagai pegawai negeri sipil atau lainnya yang berpendidikan diploma, SMA atau lainnya berjenis kelamin laki-laki dengan potensi lancar 76,9% dan potensi macet 23,1%.
5. Simpul akhir 9 memiliki label kelas kredit lancar.  
Simpul akhir 9 ini memiliki profil debitur dengan pekerjaan sebagai pegawai negeri sipil yang berpendidikan sarjana dan belum menikah dengan potensi lancar 100%.
6. Simpul akhir 10 memiliki label kelas kredit macet.  
Simpul akhir 10 ini memiliki profil debitur dengan pekerjaan sebagai lainnya yang berpendidikan sarjana dan belum menikah dengan potensi macet 100%.

#### 4.2.6. Ukuran Ketepatan Klasifikasi

Pohon klasifikasi yang telah dihasilkan diuji tingkat ketepatan atau akurasi dalam mengklasifikasi data *training*. Uji ketepatan pohon klasifikasi dilakukan menggunakan matriks konfusi seperti pada Tabel 2 berikut ini.

**Tabel 2.** Tabel Ketepatan Klasifikasi CART

$f_{ij}$		Kelas Hasil Prediksi ( $j$ )	
		Kelas lancar	Kelas macet
Kelas asli ( $i$ )	Kelas lancar	798	2
	Kelas macet	189	11

$$\text{Akurasi} = \frac{798+11}{798+2+189+11} = 0,809$$

## 5. Kesimpulan

Kredit merupakan asset yang paling besar yang dikelola bank dan juga merupakan kontributor yang paling dominan terhadap pendapatan bank. Oleh sebab itu setiap bank menerapkan prinsip kehati-hatian dalam mengalokasikan kreditnya. Setelah melakukan analisis menggunakan CART (*Classification And Regression Trees*) diketahui bahwa

variabel yang berpengaruh terhadap status kredit seorang debitur adalah jenis pekerjaan, jenjang pendidikan, status pernikahan dan jenis kelamin. Dengan menggunakan metode ini diketahui pula bahwa profile debitur yang memiliki potensi bayar paling tinggi dan potensi macet paling rendah adalah debitur dengan profile pekerjaan sebagai PNS, berpendidikan sarjana dan belum menikah. Dengan menggunakan regresi logistik biner diketahui bahwa variabel yang mempengaruhi status kredit seorang debitur adalah jenis pekerjaan. Pada kasus klasifikasi nasabah kredit, kinerja CART lebih baik jika dibandingkan dengan regresi logistik biner.

#### DAFTAR PUSTAKA

1. Agresti, A., *An Introduction to Categorical Data Analysis*, John Wiley & Sons. Inc., New York, 1990.
2. Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J., *Classification and Regression Trees*, Chapman & Hall, New York, 1984.
3. Hosmer, D.W. and Lemeshow, S, *Apllied Logistic Regression*, John Wiley & Sons, New York, 1989.
4. Kasmir, *Manajemen Perbankan*, Penerbit PT Rajagrafindo Persada, Jakarta, 2002.
5. Lewis, R.J, *An Introduction to Classification And Regression Tree (CART)*, Analysis, Annual Meeting of the Society For Academic Emergency Medicine in San Fransisco, California, Department of Emergency Medicine, California. 2000,
6. Pandia, F., *Manajemen Dana Dan Kesehatan BANK*, Penerbit RINEKA CIPTA, Jakarta, 2012.