

ANALISIS KLASIFIKASI NASABAH KREDIT MENGGUNAKAN *BOOTSTRAP AGGREGATING CLASSIFICATION AND REGRESSION TREES* (*BAGGING CART*)

Desy Ratnaningrum¹, Moch. Abdul Mukid², Triastuti Wuryandari³

¹Mahasiswa Jurusan Statistika FSM Universitas Diponegoro

^{2,3}Staff Pengajar Jurusan Statistika FSM Universitas Diponegoro

ABSTRACT

Credit is one of the facilities provided by banks to lend money to someone or a business entity within the prescribed period. The smooth repayment of credit is essential for the bank because it influences the performance as well as its presence in daily life. Acceptance of prospective credit customers should be considered to minimize the occurrence of bad credit. Classification and Regression Trees (CART) is a statistical method that can be used to identify potency of credit customer status such as current credit and bad credit. The predictor variables used in this study are gender, age, marital status, number of children, occupation, income, tenor / period, and home ownership. To improve the stability and accuracy of the prediction were used the Bootstrap Aggregating Classification and Regression Trees (Bagging CART) method. The classification of credit customers using Bagging CART gives the classification accuracy 81,44%.

Key words : Credit, Bootstrap Aggregating Classification and Regression Trees (Bagging CART), Classification Accuracy

1. PENDAHULUAN

Salah satu lembaga keuangan yang dikenal oleh masyarakat adalah bank. Di Indonesia, terdapat berbagai jenis bank yang menyediakan berbagai layanan jasa keuangan yang dibutuhkan baik oleh masyarakat maupun perusahaan. Seseorang atau kelompok yang memiliki usaha atau sedang akan merintis usaha tidak perlu khawatir lagi jika merasa tidak memiliki modal atau kekurangan modal, karena bank tidak hanya berfungsi sebagai tempat menyimpan uang tetapi menyediakan jasa keuangan lainnya yang dapat dimanfaatkan. Salah satu jasa keuangan yang kini sangat diminati oleh masyarakat maupun perusahaan adalah layanan jasa kredit.

Undang-Undang No. 10 tahun 1998 menyebutkan bahwa 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^[2]. Namun pada kenyataannya, tidak semua orang taat untuk membayar angsuran kredit pada waktu yang ditentukan. Hal ini dapat merugikan pihak bank jika terjadi dalam waktu berkepanjangan sehingga pihak bank harus selektif dalam menerima calon debitur untuk meminimalkan resiko kredit macet.

Salah satu metode statistika yang dapat digunakan untuk mengklasifikasikan nasabah kredit yaitu kredit lancar dan kredit macet adalah *Classification and Regression Trees* (CART). Metode ini merupakan metode yang dapat diterapkan pada data yang memiliki jumlah besar dan variabel yang banyak. Namun demikian, hasil pohon klasifikasi CART cenderung tidak stabil, karena perubahan-perubahan kecil pada data *learning* akan

mempengaruhi hasil akurasi prediksi. Untuk memperbaiki stabilitas dan kekuatan prediksi CART dapat digunakan metode *bootstrap aggregating (bagging)* CART^[2].

2. TINJAUAN PUSTAKA

2.1. Kredit

Kredit berasal dari kata "*credere*" yang artinya adalah kepercayaan. Maksudnya adalah apabila seseorang memperoleh kredit, maka artinya mereka memperoleh kepercayaan, sedangkan bagi si pemberi kredit artinya memberikan kepercayaan kepada seseorang bahwa uang yang dipinjamkan pasti kembali. Pengertian kredit menurut Undang-Undang RI Perbankan nomor 10 tahun 1998 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 melunasi utangnya setelah jangka waktu tertentu dengan pemberian bunga^[2].

2.2. Classification and Regression Trees (CART)

CART (*Classification and Regression Trees*) adalah salah satu metode nonparametrik yang digunakan untuk analisis klasifikasi yaitu dengan teknik pohon keputusan baik untuk variabel respon kategorik maupun kontinu. Apabila variabel respon berskala kontinu maka CART menghasilkan pohon regresi (*regression trees*), sedangkan apabila variabel respon berskala kategorik, maka CART menghasilkan pohon klasifikasi (*classification trees*).

Tujuan utama CART adalah untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian dan menentukan prediksi struktur data^[7]. Beberapa keunggulan dari metode CART adalah tidak ada asumsi yang harus dipenuhi sehingga CART termasuk dalam kelompok metode statistika nonparametrik. CART dapat digunakan pada data kontinu maupun kategorik, dapat diterapkan pada data dengan jumlah besar dan variabel yang banyak, lebih cepat perhitungannya, dan hasil lebih mudah untuk diinterpretasikan^[6].

2.3. Langkah Kerja CART

2.3.1. Proses Pemilahan Simpul (*Splitting Nodes*)

Proses pemilahan dimulai dari simpul utama yang terdiri dari data yang akan dipilah. Pemilahan dilakukan untuk memilah data menjadi dua kelompok yaitu kelompok yang masuk simpul kiri dan yang masuk simpul kanan. Aturan pemilahannya adalah sebagai berikut :

1. Tiap pemilahan bergantung pada satu nilai pemilah yang hanya berasal dari satu variabel prediktor.
2. Untuk variabel prediktor kontinu X_g , pemilahan berasal dari pertanyaan "Apakah $x_g \leq c_n$?" jika ruang sampel berukuran N dan terdapat n nilai amatan yang berbeda pada variabel X_g dengan nilai $c_n = (-\infty, \infty)$ dan c_n adalah nilai tengah antara dua nilai amatan variabel berukuran X_g berbeda.
3. Untuk variabel kategorik, pemilahan berasal dari semua kemungkinan pemilahan. Apabila X_g merupakan variabel kategorik bertaraf L , maka akan diperoleh pemilahan sebanyak $2^{L-1} - 1$.

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

$$i(t) = \sum_{\substack{j=1, \\ j \neq k}}^J P(j|t)P(k|t) \quad (1)$$

dengan $P(j|t)$ merupakan proporsi kelas j pada simpul t , $P(k|t)$ merupakan proporsi kelas k pada simpul t .

Goodness of split merupakan suatu evaluasi pemilihan oleh pemilah s pada simpul t . *Goodness of split* $\Delta i(s, t)$ didefinisikan sebagai penurunan keheterogenan dan dinyatakan sebagai berikut :

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

Nilai $\Delta i(s, t)$ digunakan sebagai uji *goodness of split criterion* (kriteria uji pemilahan terbaik). Pengembangan pohon dilakukan dengan mencari semua kemungkinan pemilah pada simpul t_1 sehingga ditemukan pemilah s^* yang memberikan nilai penurunan keheterogenan tertinggi yaitu :

$$\Delta i(s^*, t) = \max_j \Delta i(s, t_1) \quad (3)$$

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

2.3.2. Proses Pelabelan Kelas (*Class Assignment*)

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

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

maka label kelas untuk simpul akhir t adalah j_0 .

2.3.3. Proses Penghentian Pemilahan (*Stop the Splitting*)

Suatu simpul t akan menjadi simpul akhir atau tidak akan dipilah kembali apabila hanya terdapat satu pengamatan dalam tiap simpul anak, semua pengamatan dalam tiap simpul anak memiliki distribusi variabel respon yang identik, dan adanya batasan jumlah kedalaman pohon maksimal yang ditentukan oleh peneliti^[6]. Apabila hal tersebut terpenuhi, maka pengembangan pohon dihentikan dan diperoleh pohon klasifikasi maksimal atau *maximal tree* (T_{max}).

2.3.4. Proses Pemangkasan Pohon Klasifikasi (*Pruning*)

Pohon klasifikasi maksimal yang terbentuk dimungkinkan berukuran sangat besar. Semakin banyak pemilahan yang dilakukan maka tingkat akurasi semakin tinggi, tetapi dengan ukuran yang sangat besar akan sulit dipahami sehingga menyebabkan *overfitting* (pencocokan nilai yang sangat kompleks) untuk data baru. Masalah tersebut diatasi dengan pemangkasan pada pohon klasifikasi maksimal untuk mendapatkan pohon klasifikasi yang optimal^[6].

Ukuran pemangkasan yang digunakan untuk memperoleh ukuran pohon yang layak yaitu^[1] :

$$R(T) = \sum_{t \in T} r(t)P(t) = \sum_{t \in T} R(t) \quad (5)$$

dengan *resubstitution estimate* $R(T)$ adalah *tree missclassification cost* atau *tree resubstitution cost* (probabilitas terjadinya kesalahan klasifikasi yang disebabkan oleh pohon klasifikasi yang terbentuk), $P(t)$ adalah proporsi amatan yang masuk dalam simpul t , \tilde{T} merupakan himpunan simpul akhir, sedangkan *resubstitution estimate* $r(t)$ adalah probabilitas terjadinya kesalahan klasifikasi di dalam sebuah simpul t tertentu yang didefinisikan sebagai berikut :

$$r(t) = 1 - \max_j P(j|t) \quad (6)$$

Langkah awal proses pemangkasan dilakukan terhadap T_t yaitu subpohon dari T_{max} dengan mengambil t_L yang merupakan simpul anak kiri dan t_R yang merupakan simpul anak kanan hasil dari pemilahan simpul induk t . Jika diperoleh dua simpul anak dan simpul induk yang memenuhi persamaan

$$R(t) = R(t_L) + R(t_R) \quad (7)$$

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

2.4. Bootstrap Aggregating (Bagging)

Bootstrap aggregating, bagging, merupakan salah satu teknik yang diusulkan oleh Breiman (1996) yang dapat digunakan pada beberapa metode klasifikasi dan regresi untuk mereduksi variansi dari suatu prediktor, dan dengan demikian dapat memperbaiki proses pendugaan^[8]. Metode *bootstrap* ini digunakan untuk melakukan suatu *resampling* atau pengambilan data sampel yang saling bebas dan dilakukan secara berulang-ulang, yang kemudian diestimasi tingkat kesalahannya (*error*) dari pengulangan tersebut^[3]. Sampel *bootstrap* diperoleh dengan cara mengambil sampel secara random dengan pengembalian dari sampel asli.

Proses pembuatan dugaan secara *bagging* menggunakan pohon adalah sebagai berikut^[8] :

1. Tahapan *bootstrap*
 - a. Tarik sampel acak dengan pemulihan berukuran n dari gugus data *learning*.
 - b. Susun pohon terbaik berdasarkan data tersebut.
 - c. Ulangi langkah a-b sebanyak B kali sehingga diperoleh B buah pohon klasifikasi.
2. Tahapan *aggregating*
Lakukan pendugaan gabungan berdasarkan B buah pohon klasifikasi tersebut menggunakan aturan *majority vote* (suara terbanyak).

Penggunaan *bagging* ini sangat membantu terutama mengatasi sifat ketidakstabilan pohon klasifikasi dan pohon regresi. Pada banyak gugus data yang dicoba, *bagging* mampu mengurangi tingkat kesalahan klasifikasi pada kasus klasifikasi^[2]. Mengenai berapa banyak pengulangan *bootstrap* yang diperlukan, menggunakan 50 kali untuk kasus klasifikasi dan 25 kali untuk kasus regresi dapat memberikan hasil yang memuaskan^[2].

2.5. Ketepatan Klasifikasi

Apparent Error Rate (APER) merupakan suatu ukuran yang digunakan untuk menyatakan nilai proporsi sampel yang salah diklasifikasikan^[4]. Jika n_1 merupakan jumlah anggota observasi yang diklasifikasikan $Y = 1$ dan n_2 merupakan jumlah anggota observasi yang diklasifikasikan $Y = 2$, maka bentuk matriks konfusinya yaitu seperti pada Tabel 1.

Tabel 1. Matriks Konfusi

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

$$APER = \frac{n_{12} + n_{21}}{n_1 + n_2} \quad (8)$$

dengan

n_{11} = jumlah amatan kelas 1 yang tepat diklasifikasikan sebagai kelas 1

n_{12} = jumlah amatan kelas 1 yang salah diklasifikasikan sebagai kelas 2

n_{22} = jumlah amatan kelas 2 yang tepat diklasifikasikan sebagai kelas 2

n_{21} = jumlah amatan kelas 2 yang salah diklasifikasikan sebagai kelas 1

Maka, untuk menghitung nilai ketepatan klasifikasi adalah $1 - APER$.

3. METODE PENELITIAN

3.1. Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang berasal dari hasil pendataan nasabah kredit suatu bank di Provinsi Lampung tahun 2011. Ukuran sampelnya sebanyak 974 data.

3.2. Variabel Penelitian

Variabel respon (Y) yang dianalisis adalah status kredit nasabah yang dikategorikan sebagai berikut :

Y = 1, untuk kredit lancar

Y = 2, untuk kredit macet

dan terdiri dari delapan variabel prediktor (X) yang diduga memiliki pengaruh terhadap status kredit nasabah yaitu seperti pada Tabel 2.

3.3. Langkah-langkah Analisis Data

Metode yang akan digunakan yaitu metode *Bootstrap Aggregrating Classification Analysis Regression Tree (Bagging CART)* dengan bantuan *software* Ms. Excel 2007, IBM SPSS 22, dan Minitab 14 dengan tahap-tahap sebagai berikut :

1. Mengumpulkan data sekunder.
2. Menentukan variabel respon dan variabel prediktor.
3. Analisis deskriptif data nasabah berdasarkan status kredit.
4. Menentukan proporsi data *learning* : data *testing* yaitu 90% : 10%.
5. Dari 90% data *learning*, kemudian dilakukan *resampling* secara acak dan dengan pemulihan menggunakan bantuan *software* Minitab 14.
6. Melakukan pembuatan pohon klasifikasi dengan algoritma CART dengan tahap-tahap sebagai berikut :
 - a. Proses pemilahan simpul (*splitting nodes*) dengan kriteria pemilahan terbaik (*goodness of split*) untuk memilah/memecah sebuah simpul.
 - b. Proses pelabelan kelas (*class assignment*) dengan aturan jumlah anggota terbanyak pada simpul t .
 - c. Proses penghentian pembentukan pohon klasifikasi (*stop the splitting*).
 - d. Proses pemangkasan pohon klasifikasi (*Pruning*) terjadi jika dua simpul anak dan simpul induk memenuhi persamaan $R(t) = R(t_L) + R(t_R)$.

7. Pendugaan pada data *testing*.
8. Proses *bagging* (*bootstrap aggregating*).
Langkah 1 : Ulangi langkah 5 – 7 dengan data *learning* yang berbeda sehingga diperoleh 25 buah pohon klasifikasi.
Langkah 2 : Melakukan pendugaan gabungan berdasarkan hasil pendugaan 25 buah pohon klasifikasi yang terbentuk menggunakan aturan *majority vote* (suara terbanyak).
9. Menghitung ketepatan klasifikasi hasil dari pendugaan gabungan menggunakan kriteria APER.

Tabel 2. Variabel Prediktor

Variabel	Nama	Keterangan	Tipe Variabel
X1	Jenis Kelamin	1 : Laki-laki 2 : Perempuan	Kategorik
X2	Usia	-	Kontinu
X3	Status Marital	1 : Belum Menikah 2 : Menikah 3 : Cerai	Kategorik
X4	Jumlah Anak	-	Kontinu
X5	Pekerjaan	1 : Pegawai Negri Sipil 2 : Pegawai Swasta 3 : Pegawai BUMN/BUMD 4 : Lainnya	Kategorik
X6	Pendapatan	-	Kontinu
X7	Tenor/Jangka Waktu	-	Kontinu
X8	Status Kepemilikan Rumah	1 : Orang Tua 2 : Milik Sendiri 3 : Kredit 4 : Sewa 5 : Lainnya	Kategorik

4. HASIL DAN PEMBAHASAN

4.1. Deskripsi Data

Data yang digunakan pada penelitian ini merupakan data pribadi nasabah kredit pada salah satu bank di Provinsi Lampung yang terdiri dari delapan variabel prediktor. Dari 974 data nasabah yang diambil sebagai sampel, terdapat 781 (80,1%) nasabah berstatus kredit lancar dan sebanyak 193 (19,9%) nasabah yang berstatus kredit macet.

4.2. Pembagian Data

Dari jumlah keseluruhan data yaitu sebanyak 974 data, diambil sebanyak 877 data (90%) sebagai data *learning* yang digunakan untuk proses pembentukan pohon klasifikasi dan sebanyak 97 data (10%) sebagai data *testing* yang digunakan untuk mengukur kemampuan dari pohon klasifikasi yang terbentuk dalam menduga atau memprediksi status nasabah kredit.

Pada metode *Bootstrap Aggregating* nantinya akan ada dua proses. Pertama, proses *bootstrap* dengan pembentukan pohon klasifikasi sebanyak 25 kali pengulangan dengan jumlah data *learning* pada masing-masing pohon klasifikasi sebanyak 877 yang diambil secara acak dan dengan pemulihan menggunakan bantuan *software* Minitab 14. Kedua, proses *aggregating* yaitu pendugaan gabungan berdasarkan dugaan data *testing* pada

masing-masing pohon klasifikasi yang telah terbentuk dengan aturan *majority vote* (suara terbanyak).

4.3. Pembentukan Pohon Klasifikasi Pertama

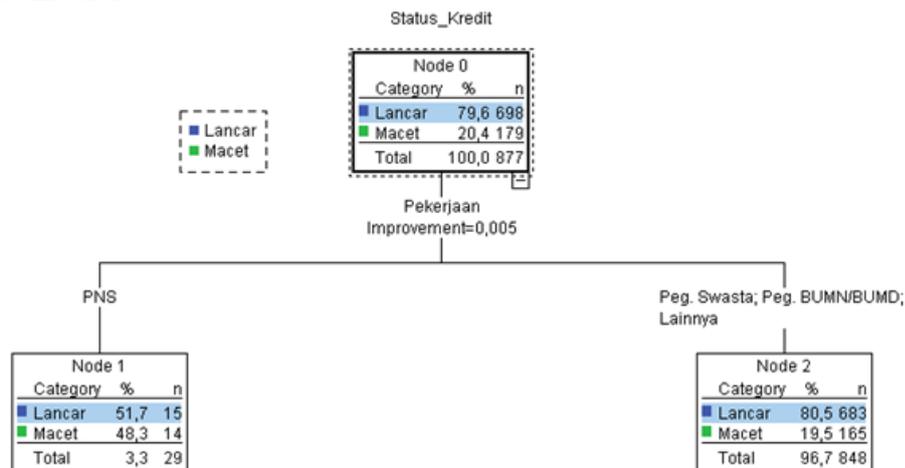
4.3.1. Proses Pemilahan Simpul (*Splitting Node*)

Suatu *split* s akan digunakan untuk memilah simpul t menjadi dua buah simpul yaitu simpul kiri (t_L) dan simpul kanan (t_R) jika s memaksimalkan nilai $\Delta i(s^*, t) = \max_{s \in S} \Delta i(s, t)$. Nilai *goodness of split* pada semua kemungkinan pemilah pada variabel pekerjaan dapat dilihat pada Tabel 3.

Tabel 3. *Goodness of Split* Variabel Pekerjaan

No.	Pemilah		<i>Goodness of Split</i>
	Simpul Kiri	Simpul Kanan	
1.	PNS	Peg. Swasta, Peg. BUMN/BUMD, Lainnya	0,005311
2.	Peg. Swasta	PNS, Peg. BUMN/BUMD, Lainnya	0,000165
3.	Peg. BUMN/BUMD	PNS, Peg. Swasta, Lainnya	0,001462
4.	Lainnya	PNS, Peg. Swasta, Peg. BUMN/BUMD	0,000327
5.	PNS, Peg. Swasta	Peg. BUMN/BUMD, Lainnya	0,003288
6.	PNS, Peg. BUMN/BUMD	Peg. Swasta, Lainnya	0,000003
7.	PNS, Lainnya	Peg. Swasta, Peg. BUMN/BUMD	0,00464

Pemilah terbaik untuk simpul 0 atau simpul akar adalah variabel pekerjaan dengan kriteria pemilahan PNS pada simpul kiri (simpul 1) dan Peg. Swasta, Peg. BUMN/BUMD, Lainnya pada simpul kanan (simpul 2). Variabel tersebut terpilih karena memiliki nilai *goodness of split/improvement* tertinggi dari variabel lainnya. Proses pemilahan dapat dilihat pada Gambar 1.



Gambar 1. Pemilahan Simpul Akar pada Pohon Klasifikasi Pertama

4.3.2. Proses Pelabelan Kelas (*Class Assignment*)

Proses pelabelan kelas pada simpul-simpul yang terbentuk berdasarkan aturan jumlah anggota kelas terbanyak yaitu jika $P(j_0|t) = \max_j P(j|t)$, maka $j_0 = j$ dengan $j = \text{lancar}$ dan $j = \text{macet}$. Sebagai contoh yaitu simpul 1 pada Gambar 1.

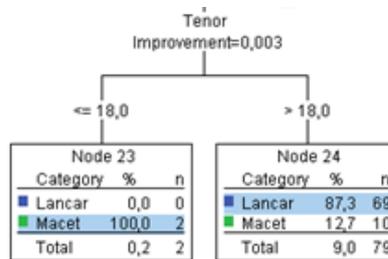
$$P(\text{Lancar}|\text{simpul 1}) = \frac{15}{29} = 0,517$$

$$P(\text{Macet}|\text{simpul 1}) = \frac{14}{29} = 0,483$$

sehingga simpul 1 diberi label kelas lancar, karena proporsi kelas lancar lebih besar dari proporsi kelas macet.

4.3.3. Proses Penghentian Pemilahan (*Stop the Splitting*)

Pohon klasifikasi maksimal yang pertama memiliki 14 simpul dalam dan 15 simpul akhir. Pada pembahasan disini kedalaman maksimal sebesar lima tingkatan. Proses penghentian pemilahan dapat dilihat pada contoh simpul 23 dan simpul 24 pada Gambar 2. Pada simpul 23 terdapat dua amatan pada kelas yang sama (homogen) dan pada simpul 24 terdapat 79 amatan pada kedalaman maksimal yaitu kedalaman lima sehingga proses pemilahan simpul dihentikan.



Gambar 2. Simpul 23 dan 24 pada Pohon Klasifikasi Maksimal Pertama

4.3.4. Proses Pemangkasan Pohon Klasifikasi (*Pruning*)

Proses pemangkasan pohon klasifikasi maksimal dimulai dengan mengambil t_L yang merupakan simpul kiri dan t_R yang merupakan simpul kanan dari T_{max} yang dihasilkan dari simpul induk t . Jika diperoleh dua simpul anak dan simpul induk yang memenuhi persamaan $R(t) = R(t_L) + R(t_R)$, maka simpul anak t_L dan t_R dipangkas. Proses tersebut diulang sampai tidak ada lagi pemangkasan yang mungkin dilakukan. Sebagai contoh simpul yang dipangkas yaitu pada simpul 13.

Pada simpul 13 diperoleh

$$r(\text{simpul 13}) = 1 - \max_j P(j|\text{simpul 13}) = 1 - 0,667 = 0,333$$

$$P(\text{simpul 13}) = \frac{12}{877} = 0,014$$

$$R(\text{simpul 13}) = r(\text{simpul 13}) * P(\text{simpul 13}) \\ = 0,333 * 0,014 = 0,004$$

Selanjutnya dihitung nilai $R(t_L)$ dan $R(t_R)$ pada simpul anak, yaitu simpul 19 dan simpul 20.

Pada simpul 19 diperoleh

$$r(\text{simpul 19}) = 1 - \max_j P(j|\text{simpul 19}) = 1 - 0,5 = 0,5$$

$$P(\text{simpul 19}) = \frac{8}{877} = 0,009$$

$$R(\text{simpul 19}) = r(\text{simpul 19}) * P(\text{simpul 19}) \\ = 0,5 * 0,009 = 0,004$$

Pada simpul 20 diperoleh

$$r(\text{simpul } 20) = 1 - \max_j P(j|\text{simpul } 20) = 1 - 1 = 0$$

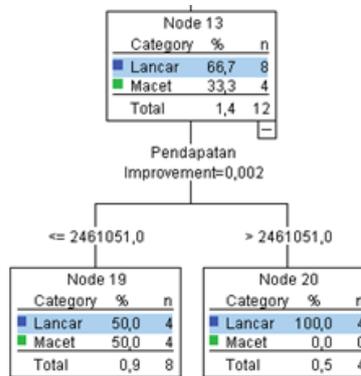
$$P(\text{simpul } 20) = \frac{4}{877} = 0,005$$

$$R(\text{simpul } 20) = r(\text{simpul } 20) * P(\text{simpul } 20) \\ = 0 * 0,005 = 0$$

Dengan demikian persamaan $R(t) = R(t_L) + R(t_R)$ dengan

$$R(\text{simpul } 19) + R(\text{simpul } 20) = 0,004 + 0 \\ = 0,004 \\ = R(\text{simpul } 13)$$

terpenuhi, sehingga dilakukan pemangkasan pada simpul tersebut.



Gambar 3. Simpul 13 pada Pohon Klasifikasi Maksimal Pertama yang Dipangkas

4.4. Prediksi pada Pohon Klasifikasi Pertama

Sebagai contoh dugaan pada data *testing* pertama dengan kriteria jenis kelamin perempuan, usia 31 tahun, status marital belum menikah, jumlah anak sebanyak 0, pekerjaan pegawai swasta, pendapatan 2.055.000, tenor/jangka waktu 36 bulan, dan status kepemilikan rumah orang tua. Kriteria tersebut masuk pada model Pekerjaan = Peg. Swasta, Peg. BUMN/BUMD, Lainnya (*node* 2), Usia > 30,5 (*node* 6), dan Pendapatan ≤ 48.105.150 yang diprediksi masuk ke dalam kelas lancar.

4.5. Prediksi dengan Majority Vote

Setelah dilakukan proses *bootstrap* sebanyak 25 kali pembentukan pohon klasifikasi dan prediksi data *testing* pada masing-masing pohon klasifikasi, langkah selanjutnya adalah prediksi gabungan (*aggregating*) berdasarkan 25 dugaan pada data *testing* dengan aturan *majority vote* (suara terbanyak). Sebagai contoh pada data *testing* yang pertama.

Berdasarkan hasil prediksi 25 pohon klasifikasi, data *testing* pertama yang diprediksi masuk kelas kredit lancar sebanyak 21 sedangkan yang diprediksi masuk kelas kredit macet sebanyak 4. Hasil akhir pendugaan berdasarkan suara terbanyak, sehingga prediksi akhir data *testing* pertama masuk ke dalam kelas kredit lancar.

4.6. Ketepatan Klasifikasi

Hasil prediksi gabungan yang telah diperoleh selanjutnya digunakan untuk menguji ketepatan klasifikasi pada penerapan metode *bagging* CART. Uji ketepatan klasifikasi dilakukan menggunakan matriks konfusi pada Tabel 4.

Tabel 4. Matriks Konfusi Hasil Klasifikasi Menggunakan *Bagging* CART

Observasi	Prediksi	
	Lancar	Macet
Lancar	78	0
Macet	18	1

$$\begin{aligned} APER &= \frac{n_{12} + n_{21}}{n_1 + n_2} \\ &= \frac{0 + 18}{97} \\ &= 0,1856 \end{aligned}$$

Dengan nilai $APER = 0,1844$, maka ketepatan klasifikasinya adalah $1 - 0,1856 = 0,8144$ atau sebesar 81,44%.

5. KESIMPULAN

Berdasarkan hasil dari pembahasan, dapat disimpulkan bahwa metode *bagging* CART dapat diterapkan untuk mengklasifikasikan nasabah kredit. Hasil pengklasifikasian tersebut selanjutnya digunakan untuk memprediksi calon nasabah yang akan menggunakan layanan jasa kredit, sehingga dapat meminimalisir terjadinya kerugian akibat dari pembayaran angsuran kredit yang macet. Dalam penelitian ini, pengklasifikasian nasabah kredit dengan metode *bagging* CART menghasilkan ketepatan klasifikasi sebesar 81,44%. Artinya bahwa metode *bagging* CART dapat mengklasifikasikan nasabah kredit secara tepat sebesar 81,44%.

DAFTAR PUSTAKA

- [1] Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1993. *Classification and Regression Trees*. New York : Chapman and Hall (Wadsworth, Inc.)
- [2] Breiman, L., 1996. *Bagging Predictors*. Machine Learning. Vol. 24. 123-140
- [3] Efron, B. and Tibshirani, R. J., (1993). *An Introduction to the Bootstrap*. New York : Chapman & Hall.
- [4] Johnson, R. A. and Wichern, D. W., 2007. *Applied Multivariate Statistical Analysis*. Sixth Edition. USA : Pearson Education Inc.
- [5] Kasmir. 2002. *Bank dan Lembaga Keuangan Lainnya*. Jakarta : PT. Raja Grafindo Persada.
- [6] Lewis, R. J., 2000. *An Introduction to Classification and Regression Trees (CART) Analysis*. Presented at the 2000 Annual Meeting of Society for Academic Emergency Medicine of Sanfransisco. California.
- [7] Otok, B. W. dan Sumarmi. 2009. *Bagging CART pada Klasifikasi Anak Putus Sekolah*. Seminar Nasional Statistika IX. Surabaya : Institut Teknologi Sepuluh Nopember.
- [8] Sutton, C. D., 2005. *Classification and Regression Trees, Bagging, and Boosting*. Handbook of Statistics. Vol. 24. Hal 303-329.