

BAGGING CLASSIFICATION TREES UNTUK KLASIFIKASI TINGKAT KESEJAHTERAAN RUMAH TANGGA DI KALIMANTAN BARAT

Desi Ayu Wulandari, Dadan Kusnandar, Yudhi

INTISARI

Metode classification and regression trees (CART) merupakan teknik pohon keputusan untuk analisis klasifikasi variabel respon kategorik maupun kontinu yang dapat diterapkan pada data jumlah besar dan variabel yang banyak. Stabilitas dan kekuatan prediksi pohon klasifikasi diperbaiki dengan metode Bootstrap Aggregating (Bagging) classification trees. Tujuan penelitian ini adalah untuk menentukan nilai ketepatan hasil pengklasifikasian tingkat kesejahteraan rumah tangga di Kalimantan Barat dengan menggunakan bagging classification trees. Data yang digunakan adalah data sekunder dari hasil Survei Sosial Ekonomi Nasional (SUSENAS) Provinsi Kalimantan Barat Tahun 2017 dengan 1400 sampel yang terdiri dari tujuh variabel bebas dan satu variabel terikat. Penelitian ini menghasilkan ketepatan klasifikasi sebesar 57,5% dengan menggunakan metode bagging classification trees. Metode bagging classification trees mampu meningkatkan ketepatan klasifikasi dari 50% pada pohon klasifikasi awal menjadi 57,5% pada bagging classification trees. Dapat disimpulkan bahwa penerapan bagging classification trees lebih baik daripada pohon klasifikasi tanpa bagging karena mampu meningkatkan ketepatan klasifikasi sebesar 7,5%.

Kata Kunci: Bagging CART, Pohon Keputusan, Goodness of Split, Ketepatan Klasifikasi

PENDAHULUAN

Kemiskinan merupakan suatu masalah yang menjadi pusat perhatian di berbagai negara. Kemiskinan adalah penyebab utama dari berbagai masalah baik dari segi politik, sosial dan ekonomi, khususnya di negara-negara berkembang. Indonesia adalah salah satu negara berkembang yang selalu berusaha untuk mengatasi permasalahan kemiskinan ini. Garis kemiskinan dipergunakan sebagai suatu batas untuk mengelompokkan penduduk menjadi miskin atau tidak miskin. Pada bulan Maret 2018, jumlah penduduk miskin di Kalimantan Barat mencapai 387,08 ribu orang (7,77 persen). Jumlah ini berkurang sebesar 1,7 ribu orang dibandingkan dengan kondisi September 2017 yaitu 7,86 persen atau sebesar 388,81 ribu orang [1].

Miskin atau tidaknya suatu penduduk sangat penting diketahui agar pemerintah dapat memberikan solusi dan mengatasi permasalahan kemiskinan ini. Banyak program kerja yang dilakukan pemerintah untuk mengatasi permasalahan kemiskinan yang terjadi pada masyarakat. Program-program kerja tersebut dapat tepat sasaran dengan cara mengidentifikasi apakah suatu rumah tangga tersebut miskin atau tidak miskin. Selama ini pengelompokkan rumah tangga miskin hanya berdasarkan pendapatan atau pengeluaran. Namun, pendapatan atau pengeluaran belum optimal untuk menggambarkan kondisi sesungguhnya dari suatu rumah tangga. Oleh sebab itu, maka diperlukan suatu cara agar dapat mengetahui bagaimana pengklasifikasian rumah tangga miskin dan rumah tangga tidak miskin dari berbagai faktor.

Classification and Regression Trees (CART) merupakan sebuah metode yang termasuk dalam anggota analisis klasifikasi pohon keputusan. Metode ini dapat diterapkan pada data dengan variabel dan ukuran obyek yang besar. Oleh karena itu, metode CART digambarkan dalam bentuk yang menyerupai sebuah pohon. Jika variabel terikatnya bersifat kontinu maka pohon yang dihasilkan disebut pohon regresi (regression tree). Sedangkan jika variabel terikatnya bersifat kategorik maka pohon yang dihasilkan disebut pohon klasifikasi (classification tree) [2]. Namun, hasil dari analisis

pohon klasifikasi (*classification tree*) cenderung tidak stabil. Hal ini disebabkan adanya perubahan-perubahan kecil pada data *learning* yang akan mempengaruhi hasil akurasi prediksi. Oleh karena itu, untuk memperbaiki stabilitas dan kekuatan prediksi pohon klasifikasi dapat digunakan metode *Bootstrap Aggregating (Bagging) Classification Trees* [4].

Tujuan penelitian ini adalah untuk menentukan nilai ketepatan hasil pengklasifikasian tingkat kesejahteraan rumah tangga di Kalimantan Barat dengan menggunakan *Bagging Classification Trees*. Data yang dianalisis pada penelitian ini merupakan data sekunder yang berasal dari hasil Survei Sosial Ekonomi Nasional (SUSENAS) Provinsi Kalimantan Barat Tahun 2017. Pada penelitian ini, variabel terikat (Y) yang dianalisis adalah tingkat kesejahteraan rumah tangga yang dikategorikan sebagai berikut:

$Y = 1$, untuk rumah tangga miskin

$Y = 2$, untuk rumah tangga tidak miskin.

Pada penelitian ini digunakan tujuh variabel bebas yaitu tingkat pendidikan (X_1), status bekerja (X_2), status kepemilikan telepon seluler (X_3), status kepemilikan bangunan tempat tinggal (X_4), cara memperoleh air minum (X_5), sumber penerangan (X_6) dan pengalaman membeli/menerima raskin (X_7). Tahap-tahap dalam penelitian ini, yang pertama dilakukan adalah menganalisis deskriptif data. Kemudian peneliti menentukan proporsi data *learning* dan data *testing* dengan perbandingan 9:1 dan dilakukan resampling secara acak [3]. Selanjutnya dilakukan pembentukan pohon klasifikasi yang terdiri dari proses pemilahan simpul dengan kriteria pemilahan terbaik, proses pelabelan kelas dengan aturan jumlah anggota kelas terbanyak pada simpul t , proses penghentian pembentukan pohon klasifikasi dan pemangkasan pohon klasifikasi. Tahap selanjutnya yang dilakukan adalah prediksi data *testing* pada masing-masing pohon klasifikasi dan proses *bagging*. Pada metode *bagging* terdapat dua proses dalam pengerjaannya, yaitu pembentukan pohon klasifikasi dengan data *learning* yang berbeda sebanyak 51 kali. Selanjutnya dilakukan pendugaan gabungan berdasarkan hasil pendugaan 51 buah pohon klasifikasi yang telah terbentuk menggunakan aturan *majority vote* (suara terbanyak). Tahap terakhir dalam penelitian ini adalah menghitung ketepatan klasifikasi hasil dari pendugaan gabungan menggunakan kriteria *apparent error rate* (APER).

CLASSIFICATION AND REGRESSION TREES (CART)

Classification and regression trees (CART) merupakan satu diantara metode nonparametrik yang termasuk dalam teknik pohon keputusan yang digunakan untuk analisis klasifikasi variabel respon kategorik maupun kontinu. Metode ini merupakan metode yang bisa diterapkan untuk himpunan data yang mempunyai jumlah besar dan variabel yang sangat banyak dengan skala variabel campuran melalui prosedur pemilahan biner [5]. Pohon keputusan dibentuk dengan menggunakan algoritma penyekatan rekursif secara biner. Pemilahan dilakukan untuk memilah data menjadi dua kelompok yaitu kelompok yang masuk simpul kiri dan yang masuk simpul kanan. Pemilahan dilakukan pada tiap simpul hingga diperoleh sebuah simpul akhir [6].

GOODNESS OF SPLIT

Pada proses pemilahan simpul, akan dicari simpul terbaik yang memiliki kriteria *goodness of split*. *goodness of split* atau pemilahan terbaik adalah suatu evaluasi pemilahan oleh pemilah s pada simpul t . Persamaan *goodness of split* ($\phi_s(t)$) adalah sebagai berikut:

$$\phi_s(t) = i(t) - P_L \cdot i(t_L) - P_R \cdot i(t_R) \quad (1)$$

dengan $i(t)$ adalah indeks gini, P_L dan P_R adalah proporsi amatan yang masuk ke dalam simpul kiri dan simpul kanan serta t_L dan t_R adalah simpul kiri dan simpul kanan.

Pemilah yang menghasilkan nilai $\phi_s(t)$ lebih tinggi adalah pemilah yang terbaik karena pemilah yang terbaik memungkinkan dapat mereduksi keheterogenan secara lebih signifikan. Metode

pemilahan yang sering digunakan adalah indeks gini, karena lebih mudah dan sesuai untuk diterapkan dalam berbagai kasus dan mempunyai perhitungan yang sederhana dan cepat [5].

PROSES PEMILAHAN SIMPUL

Proses pemilahan simpul adalah tahap awal CART dalam membuat sebuah pohon klasifikasi, dimulai dari simpul utama yang terdiri dari data yang akan dipilah. Proses pemilahan dilakukan untuk memilah data menjadi dua kelompok yaitu kelompok simpul kiri dan simpul kanan. Pemilahan dilakukan pada setiap simpul hingga diperoleh suatu simpul akhir. Proses pemilahan pada setiap simpul induk berdasarkan pada *goodness of split criterion* (kriteria pemilahan terbaik). Kriteria pemilahan terbaik dibentuk berdasarkan fungsi *impurity* (fungsi keheterogenan). Fungsi keheterogenan yang digunakan adalah Indeks Gini. Indeks gini adalah formula yang menghitung probabilitas bahwa sebuah objek adalah anggota kelas j yang berada dalam node t . Persamaan indeks gini yaitu sebagai berikut

$$i(t) = \sum_{\substack{j=1 \\ j \neq k}}^J P(j|t)P(k|t) \quad (2)$$

dengan $P(j|t) = \frac{N_j(t)}{N(t)}$ merupakan proporsi kelas j pada simpul t dengan $N_j(t)$ merupakan jumlah pengamatan kelas j pada simpul t , $N(t)$ merupakan jumlah pengamatan pada simpul t dan $P(k|t)$ merupakan proporsi kelas k pada simpul t [5]. Pada proses pemilahan simpul, akan dicari simpul terbaik yang memiliki kriteria *goodness of split*

PROSES PELABELAN KELAS

Tahap kedua dalam pembentukan pohon klasifikasi CART adalah proses pelabelan kelas (*class assignment*). Proses ini dimulai dari awal pemilahan simpul hingga simpul akhir. Pelabelan setiap simpul akhir berdasarkan aturan jumlah anggota kelas terbanyak, jika:

$$P(j_0|t) = \max_j P(j|t) = \max_j \frac{N_j(t)}{N(t)} \quad (3)$$

dengan j_0 adalah label kelas untuk simpul akhir t [3].

PROSES PENGHENTIAN PEMILAHAN

Proses pemilahan dilakukan hingga jumlah amatan setiap simpul akhir menjadi satu amatan. Jika hal tersebut sudah terpenuhi, maka pengembangan pohon dihentikan dan diperoleh pohon klasifikasi maksimal [5]. Suatu simpul t akan menjadi simpul akhir jika hanya terdapat satu pengamatan dalam setiap simpul anak, semua pengamatan dalam setiap simpul anak memiliki distribusi variabel respon yang identik atau terdapat suatu batasan jumlah kedalaman pohon maksimal yang ditentukan oleh peneliti [7].

PROSES PEMANGKASAN POHON KLASIFIKASI

Pohon klasifikasi maksimal yang telah dibentuk memiliki kemungkinan ukuran yang sangat besar. Jika semakin banyak pemilahan yang dilakukan maka tingkat akurasi semakin tinggi. Oleh karena itu, masalah tersebut dapat diatasi dengan pemangkasan pada pohon klasifikasi maksimal untuk mendapatkan pohon klasifikasi yang optimal [7]. Ukuran pemangkasan yang digunakan untuk memperoleh ukuran pohon yang optimal [5] yaitu:

$$R(T) = \sum_{t \in \bar{T}} r(t) P(t) = \sum_{t \in \bar{T}} R(t) \quad (4)$$

dengan $R(T)$ adalah probabilitas terjadinya kesalahan klasifikasi yang disebabkan oleh pohon klasifikasi yang terbentuk, $P(t)$ adalah proporsi amatan yang masuk ke dalam simpul t , \tilde{T} adalah himpunan simpul akhir dan $r(t)$ adalah probabilitas terjadinya kesalahan klasifikasi di dalam sebuah simpul t dengan persamaan sebagai berikut:

$$r(t) = 1 - \max_j P(j|t) \quad (5)$$

Proses awal pemangkasan dilakukan pada T_t yang merupakan subpohon dari T_{\max} (pohon maksimal) dengan mengambil simpul anak kiri t_L dan simpul anak kanan t_R hasil dari pemilahan simpul induk t . Jika didapatkan dua simpul anak dan simpul induk yang memenuhi persamaan sebagai berikut:

$$R(t) = R(t_L) + R(t_R) \quad (6)$$

maka simpul anak kiri dan kanan harus dipangkas dan diulangi hingga tidak ada lagi kemungkinan pemangkasan.

BOOTSTRAP AGGREGATING (BAGGING)

Metode *bagging* terdapat dua tahapan utama, yaitu *bootstrap* yang merupakan pengambilan sampel dari data *learning* yang dimiliki dan *aggregating* yaitu menggabungkan nilai-nilai dugaan menjadi satu nilai dugaan. Proses pembuatan dugaan secara *bagging* menggunakan pohon adalah sebagai berikut [8]:

1. Tahapan *bootstrap*
 - a. Tarik sampel acak dengan pemulihan berukuran n dari data *learning*.
 - b. Susun pohon terbaik berdasarkan data tersebut.
 - c. Ulangi langkah *a-b* sebanyak B sehingga diperoleh B pohon klasifikasi
2. Tahapan *aggregating*

Lakukan pendugaan gabungan berdasarkan B buah pohon klasifikasi tersebut menggunakan aturan *majority vote* (suara terbanyak), yaitu kategori atau kelas yang paling sering muncul sebagai hasil prediksi dari k pohon klasifikasi.

Penggunaan *bagging* sangat membantu saat mengatasi sifat ketidakstabilan pohon klasifikasi dan pohon regresi. *Bagging* dapat mengurangi tingkat kesalahan klasifikasi pada kasus klasifikasi dengan menggunakan 50 kali untuk kasus klasifikasi dan 25 kali untuk kasus regresi dapat memberikan hasil yang memuaskan ketika melakukan pengulangan [2].

KETEPATAN KLASIFIKASI

Apparent error rate (APER) adalah suatu ukuran yang digunakan untuk menyatakan nilai proporsi sampel yang salah diklasifikasi. APER bisa dihitung dengan menggunakan matriks konfusi yang menyajikan anggota observasi versus anggota prediksi. Jika n_1 adalah jumlah anggota observasi yang diklasifikasikan $Y = 1$ dan n_2 adalah jumlah anggota observasi yang diklasifikasikan $Y = 2$, maka bentuk matriks konfusinya sebagai adalah berikut:

Tabel 1 Matriks Konfusi

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

dengan n_{11} adalah jumlah amatan kelas 1 yang tepat diklasifikasikan sebagai kelas 1; n_{12} adalah jumlah amatan kelas 1 yang salah diklasifikasikan sebagai kelas 2; n_{22} adalah jumlah amatan kelas 2 yang tepat diklasifikasikan sebagai kelas 2 dan n_{21} adalah jumlah amatan kelas 2 yang salah diklasifikasikan sebagai kelas 1. APER dihitung dengan rumus sebagai berikut:

$$APER = \frac{n_{12} + n_{21}}{n_1 + n_2}$$

Nilai ketepatan klasifikasi adalah $1 - APER$

HASIL DAN PEMBAHASAN

Data yang digunakan pada penelitian ini merupakan data sekunder yang berasal dari hasil Survei Sosial Ekonomi Nasional (SUSENAS) Provinsi Kalimantan Barat Tahun 2017. Data yang digunakan dalam penerapan *bagging classification trees* pada klasifikasi tingkat kesejahteraan rumah tangga adalah 1400 sampel data dengan 711 (50,8%) rumah tangga miskin dan 689 (49,2%) rumah tangga tidak miskin. Setelah membagi data ke dalam dua variabel terikat selanjutnya dilakukan pembagian data menjadi dua bagian yaitu data *learning* dan data *testing*.

Dari 1400 data, diambil sebanyak 1260 data (90%) sebagai data *learning* yang digunakan dalam proses pohon klasifikasi. Sedangkan untuk data *testing* sebanyak 140 data (10%) yang digunakan untuk mengukur kemampuan dari pohon klasifikasi yang terbentuk dalam memprediksi status tingkat kesejahteraan rumah tangga. Data *learning* dan data *testing* diambil secara acak. Selanjutnya dilakukan proses pemilahan simpul. Proses pemilahan dimulai dari simpul utama yang terdiri dari data yang akan dipilah. Proses pemilahan diawali dengan menentukan variabel pemilah terbaik dari tujuh variabel bebas yang digunakan berdasarkan kriteria *goodness of split*. Tabel 2 merupakan nilai *goodness of split* tertinggi dari tiap variabel.

Tabel 2 *Goodness of Split*

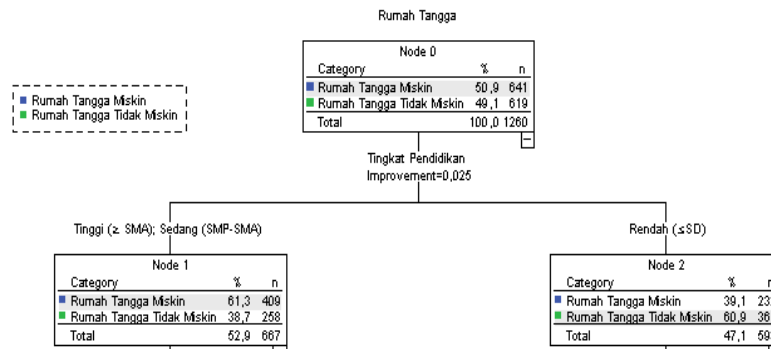
No	Variabel	Pemilah		<i>Goodness of Split</i>
		Simpul Kiri	Simpul Kanan	
1	Tingkat Pendidikan	Tinggi, Sedang	Rendah	0,025
2	Cara Memperoleh Air Minum	Membeli	Tidak membeli	0,023
3	Pengalaman Membeli/ Menerima raskin	Tidak Pernah	Pernah	0,014
4	Status Kepemilikan Telepon Seluler	Ya memiliki	Tidak Memiliki	0,012
5	Sumber Penerangan	Listrik PLN	Listrik Non PLN, Bukan listrik	0,011
6	Status Kepemilikan Bangunan	Dinas	Milik Sendiri, Bebas Sewa, Kontrak	0,005
7	Status Bekerja	Bekerja	Tidak Bekerja	0,002

Pemilah terbaik untuk simpul akar adalah variabel tingkat pendidikan. Variabel tingkat pendidikan terpilih karena memiliki nilai *goodness of Split* tertinggi dari variabel lainnya yaitu sebesar 0,025. Proses pemilahan dapat dilihat pada Gambar 1. Simpul akar terbagi menjadi dua kelompok, yaitu simpul kiri dan simpul kanan. Simpul kiri merupakan tingkat pendidikan tinggi; sedang dengan jumlah rumah tangga miskin lebih banyak yaitu sebesar 409 dan simpul kanan merupakan tingkat pendidikan rendah dengan jumlah rumah tangga tidak miskin lebih banyak yaitu sebesar 361.

Langkah selanjutnya adalah melakukan pemilahan simpul pada variabel bebas lainnya. Proses selanjutnya adalah pelabelan kelas. Proses pelabelan kelas pada simpul-simpul yang terbentuk berdasarkan aturan jumlah anggota kelas terbanyak. Jika $P(j_0|t) = \max_j P(j|t)$, maka $j_0 = j$ dengan j adalah miskin dan tidak miskin. Sebagai contoh yaitu simpul 1 pada Gambar 1.

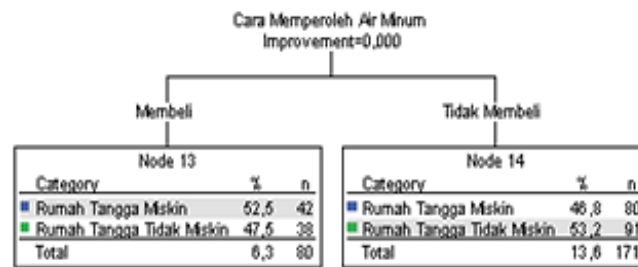
$$P(\text{Rumah Tangga Miskin}|\text{Simpul 1}) = \frac{409}{667} = 0,613$$

$$P(\text{Rumah Tangga Tidak Miskin}|\text{Simpul 1}) = \frac{258}{667} = 0,386$$



Gambar 1. Pemilahan Simpul Akar pada Pohon Klasifikasi Pertama

Simpul 1 proporsi kelas rumah tangga miskin lebih besar dari proporsi kelas rumah tangga tidak miskin. Sedangkan pada simpul 2, proporsi kelas rumah tangga tidak miskin lebih besar dari proporsi kelas rumah tangga miskin. Proses pelabelan kelas dilakukan pada semua simpul khususnya simpul akhir karena simpul akhir adalah simpul yang sangat penting dalam proses pendugaan suatu objek pada kelas tertentu jika objek tersebut berada pada simpul akhir. Proses pemilahan simpul berhenti jika tidak terjadi pemilahan lagi. Kedalaman maksimal yang diperoleh sebesar lima tingkatan yang menghasilkan pohon maksimal. Pohon klasifikasi maksimal yang pertama memiliki enam simpul dalam dan delapan simpul akhir.



Gambar 2. Pohon Klasifikasi Maksimal Pertama

Proses penghentian terdapat pada simpul 13 dan simpul 14 dengan lima tingkatan sebagai kedalaman maksimal. Jadi, proses pemilahan simpul dihentikan. Selanjutnya adalah proses pemangkasan pohon klasifikasi yaitu untuk mendapatkan pohon klasifikasi yang sederhana dan menghasilkan tingkat kesalahan prediksi yang kecil. Proses pemangkasan diawali dengan mengambil t_L (simpul kiri) dan t_R (simpul kanan) dari pohon maksimal yang dihasilkan dari simpul induk t . Jika terdapat dua simpul anak dan simpul induk yang memenuhi persamaan $R(t) = R(t_L) + R(t_R)$, maka simpul t_L dan t_R dipangkas. Sebagai contoh simpul yang dipangkas yaitu pada simpul 9, diperoleh

$$r(\text{simpul } 9) = 1 - \max_j P(j|\text{simpul } 9) = 1 - 0,541 = 0,458$$

$$P(\text{simpul } 9) = \frac{325}{1260} = 0,258$$

$$R(\text{simpul } 9) = r(\text{simpul } 9) \times P(\text{simpul } 9) = 0,458 \times 0,258 = 0,118$$

Selanjutnya dihitung nilai $R(t_L)$ dan $R(t_R)$ pada simpul 11 dan 12. Pada simpul 11 diperoleh

$$r(\text{simpul } 11) = 0,487 ; P(\text{simpul } 11) = \frac{251}{1260} = 0,199 ; R(\text{simpul } 11) = 0,097$$

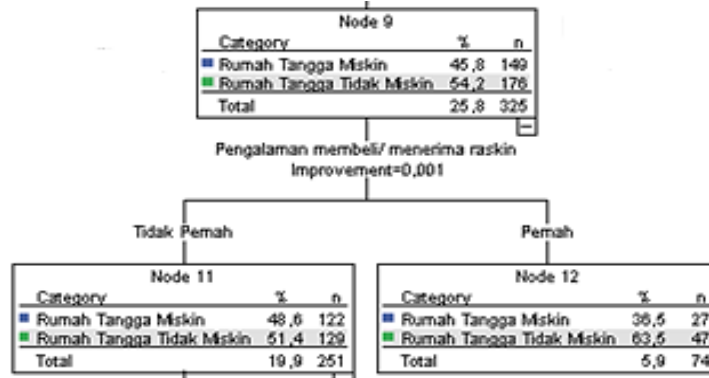
Pada simpul 12 diperoleh

$$r(\text{simpul } 12) = 0,365 ; P(\text{simpul } 12) = \frac{74}{1260} = 0,058 ; R(\text{simpul } 12) = 0,021$$

Berdasarkan persamaan (6) diperoleh

$$R(\text{simpul } 11) + R(\text{simpul } 12) = 0,097 + 0,021 = 0,118 = R(\text{simpul } 9).$$

Karena hasil perhitungan memenuhi persamaan (6), maka dilakukan pemangkasan pada simpul tersebut. Gambar 2 merupakan pemangkasan simpul 9 pada pohon klasifikasi pertama.



Gambar 3. Simpul 9 pada Pohon Klasifikasi Pertama yang Dipangkas

Proses pemangkasan dilakukan hingga tidak ada lagi kemungkinan adanya pemangkasan. Setelah proses pemangkasan berhenti diperoleh pohon klasifikasi optimal pertama. Data *testing* yang digunakan sebanyak 140 data yang digunakan untuk menduga atau memprediksi tingkat kesejahteraan rumah tangga. Sebagai contoh prediksi pada data *testing* kelima dengan kriteria pendidikan terakhir kepala rumah tangga adalah SD, bekerja, tidak memiliki telepon seluler, memiliki rumah sendiri, memperoleh air minum tidak membeli, sumber penerangan listrik PLN dan tidak pernah membeli atau menerima beras miskin. Kriteria tersebut masuk pada model Tingkat Pendidikan = Rendah ($\leq SD$) yang diprediksi masuk ke dalam kelas rumah tangga miskin. Setelah melakukan prediksi pada data pertama, maka akan dilanjutkan dengan proses *bootstrap* yaitu pembentukan pohon klasifikasi sebanyak 51 kali dengan langkah yang sama seperti pohon klasifikasi pertama dengan data *learning* yang berbeda. Kemudian diperoleh data *testing* yang akan digunakan untuk memprediksi tingkat kesejahteraan rumah tangga. Selanjutnya, prediksi gabungan (*aggregating*) berdasarkan 51 dugaan pada data *testing* dengan aturan suara terbanyak.

Hasil prediksi gabungan yang telah diperoleh akan digunakan untuk menguji ketepatan klasifikasi dengan penerapan metode *bagging classification trees*. Hasil uji ketepatan klasifikasi pada observasi pohon klasifikasi pertama dapat dilihat pada Tabel 3.

Tabel 3 Hasil Klasifikasi Pohon Pertama *Bagging Classification Trees*

Observasi	Prediksi	
	Rumah Tangga Miskin	Rumah Tangga Tidak Miskin
Rumah Tangga Miskin	74	2
Rumah Tangga Tidak Miskin	60	4

$$APER = \frac{n_{12} + n_{21}}{n_1 + n_2}$$

$$\frac{60 + 2}{140} = 0,442$$

Berdasarkan nilai $APER = 0,442$, maka ketepatan klasifikasi pertama yaitu 0,558. Selanjutnya sebanyak 51 observasi pada tiap-tiap pohon klasifikasi diuji ketepatan klasifikasi dengan hasil prediksi gabungan dan dihitung nilai rata-rata ketepatan klasifikasi. Diperoleh nilai rata-rata dari $APER$ 51 pohon klasifikasi yaitu sebesar 0,425, maka nilai ketepatan klasifikasinya yaitu $(1 - 0,425)$ atau 57,5%.

Selanjutnya nilai ketepatan klasifikasi dengan *bagging classification trees* dibandingkan dengan nilai ketepatan klasifikasi dengan pohon klasifikasi awal tanpa *bagging*. Hasil perbandingan nilai ketepatan klasifikasi pada pohon awal (tanpa *bagging*) dengan nilai ketepatan klasifikasi setelah menggunakan *bagging classification trees* dapat dilihat pada Tabel 4.

Tabel 4 Hasil Ketepatan Klasifikasi Awal dan Hasil *Bagging*

	Nilai Ketepatan Klasifikasi
Pohon Klasifikasi Pertama (tanpa <i>bagging</i>)	50%
Pohon Klasifikasi dengan <i>bagging</i> (51 kali)	57,5%
Peningkatan	7,5%

KESIMPULAN

Berdasarkan hasil dari pembahasan yang telah diuraikan, penentuan nilai ketepatan hasil pengklasifikasian tingkat kesejahteraan rumah tangga di Kalimantan Barat diperoleh nilai sebesar 57,5% dengan menggunakan metode *bagging classification trees*. Metode *bagging classification trees* mampu meningkatkan ketepatan klasifikasi dari 50% pada pohon klasifikasi awal menjadi 57,5% pada *bagging classification trees*. Dapat disimpulkan bahwa penerapan *bagging classification trees* pada penelitian ini lebih baik daripada pohon klasifikasi tanpa *bagging* karena mampu meningkatkan ketepatan klasifikasi sebesar 7,5%.

DAFTAR PUSTAKA

- [1]. Badan Pusat Statistik Provinsi Kalimantan Barat. *Profil Kemiskinan di Kalimantan Barat 2018*. Badan Pusat Statistik Provinsi Kalimantan Barat; 2018.
- [2]. Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. New York: Chapman and Hall (Wadsworth, Inc.); 1993.
- [3]. Ratnaningrum, D., Mukid, M. A., dan Wuryandari, T. Analisis Klasifikasi Nasabah Kredit Menggunakan *Bootstrap Aggregating Classification And Regression Trees (Bagging Cart)*. Semarang: *Jurnal Gaussian*, 5.1: 81-90; 2016.
- [4]. Breiman, L. Bagging Predictors. *Machine Learning*, 24: 123-140; 1996.
- [5]. Otok, B. W. dan Sumarmi. Bagging CART pada Klasifikasi Anak Putus Sekolah. Surabaya. *Seminar Nasional Statistika IX*; 2009.
- [6]. Pratiwi, E. F., dan Zain, I. Klasifikasi Pengangguran Terbuka Menggunakan CART (*Classification and Regression Tree*) di Provinsi Sulawesi Utara. Surabaya: *Jurnal Sains dan Seni Pomits*, Vol. 3 No. 1; 2014.
- [7]. Lewis, R. J. *An Introduction to Classification and Regression Trees (CART) Analysis*. Presented at the 2000 Annual Meeting of Society for Academic Emergency Medicine of Sanfransisco. California; 2000
- [8]. Sutton, C. D. Classification and Regression Trees, Bagging, and Boosting: *Handbook of Statistics*, 24: 303-329; 2005

DESI AYU WULANDARI : Jurusan Matematika FMIPA UNTAN, Pontianak
desiwulan65@gmail.com

DADAN KUSNANDAR : Jurusan Matematika FMIPA UNTAN, Pontianak
dkusnand@untan.ac.id

YUDHI : Jurusan Matematika FMIPA UNTAN, Pontianak
yudhi@math.untan.ac.id