

PENERAPAN *SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE* DALAM MENGATASI DATA TIDAK SEIMBANG PADA METODE *CLASSIFICATION AND REGRESSION TREE*

Rizky Dwi Permatasari, Setyo Wira Rizki, Naomi Nessyana Debataraja

INTISARI

Metode Classification and Regression Trees (CART) merupakan teknik klasifikasi berbentuk pohon klasifikasi, yang menggambarkan hubungan antara variabel independen dan dependennya. Ketidakseimbangan data dapat menyebabkan rendahnya nilai sensitivitas dan nilai Area Under Curve (AUC). Salah satu metode yang dapat mengatasi data tidak seimbang adalah dengan melakukan Synthetic Minority Oversampling Technique (SMOTE). SMOTE merupakan teknik dengan penambahan data buatan pada kelas minoritas pada tahap sebelum menganalisis data. Tujuan dari penelitian ini untuk membandingkan model tanpa dan dengan SMOTE pada metode CART. Adapun untuk langkah-langkah dalam melakukan SMOTE yaitu dengan menentukan nilai oversampling sebanyak data yang akan dibangkitkan pada data kelas minoritasnya. Sedangkan untuk membentuk pohon klasifikasi yaitu dengan cara pemilihan pemilah, penentuan simpul terminal, penandaan label kelas, dan penentuan pohon klasifikasi optimal. Data yang digunakan dalam penelitian ini adalah data sekunder hasil survey sosial ekonomi nasional (SUSENAS) tahun 2017 di Kabupaten Sambas dengan 494 sampel yang terdiri dari 8 variabel independen dan satu variabel dependen. Hasil ketepatan klasifikasi penelitian ini, menunjukkan bahwa model dengan SMOTE menghasilkan nilai lebih akurat dibandingkan dengan model tanpa SMOTE. Hal ini dapat dilihat model dengan SMOTE menghasilkan nilai sensitivitas lebih tinggi yaitu 67,05% dari sebelumnya yaitu 36,36%, dan nilai AUC 94,35% lebih tinggi dari sebelum dilakukan SMOTE.

Kata kunci: Pohon Klasifikasi, Ketepatan Klasifikasi, data tidak seimbang, Oversampling

PENDAHULUAN

Analisis klasifikasi adalah proses menemukan model terbaik dari *classifier* untuk memprediksi kelas dari suatu data yang label kelasnya tidak diketahui [1]. Metode klasifikasi dapat dilakukan dengan dua pendekatan yaitu pendekatan parametrik dan pendekatan nonparametrik. Salah satu metode klasifikasi nonparametrik adalah metode *Classification and Regression Trees (CART)*. Proses klasifikasi pada metode CART bersifat penyekatan rekursif biner. Metode CART akan menghasilkan suatu pohon klasifikasi apabila variabel dependen berupa kategorik dan menghasilkan pohon regresi apabila variabel dependen berupa data kontinu. Metode ini dapat menyeleksi variabel-variabel dependen yang paling penting dalam menentukan hasil klasifikasi variabel independen. Tujuan utama CART adalah untuk mendapatkan suatu kelompok data yang akurat sebagai ciri dari suatu pengklasifikasian [2]. CART memiliki beberapa kelebihan yaitu mampu bekerja pada dimensi data yang besar dan struktur data yang kompleks, tidak terikat oleh asumsi kenormalan, dapat mengetahui interaksi antar variabel independen serta hasil klasifikasi lebih mudah dipahami dan diinterpretasi karena struktur datanya dapat dilihat secara visual [3].

Suatu pengklasifikasian memiliki tingkat akurasi yang rendah dikarenakan jumlah data masing-masing kelas variabel dependen tidak seimbang, yakni terdapat kelas yang memiliki jumlah lebih besar (mayoritas) atau kelas negatif dan kelas lainnya memiliki jumlah jauh lebih sedikit (minoritas) atau kelas positif. Klasifikasi pada data tidak seimbang cenderung mengklasifikasikan pada data kelas mayoritas dan akan mengabaikan kelas minoritas, sehingga akan menghasilkan nilai akurasi yang kurang baik pada kelas minoritas [4]. Oleh karena itu, untuk meningkatkan nilai akurasi dalam memprediksi pohon klasifikasi yaitu dengan cara membangkitkan data buatan pada kelas minor dengan metode Synthetic Minority Oversampling Technique sebelum dilakukan analisis CART.

Tujuan penelitian ini adalah untuk membandingkan ketepatan nilai akurasi pengklasifikasian tingkat status kelangsungan pendidikan anak di Kabupaten Sambas tahun 2017 sebelum dan sesudah dilakukan SMOTE pada metode CART. Data yang dianalisis pada penelitian ini merupakan data sekunder yang berasal dari hasil Survey Sosial Ekonomi Nasional (SUSENAS) Provinsi Kalimantan Barat tahun 2017 di Kabupaten Sambas. Penelitian ini menggunakan variabel dependen (Y) yang dianalisis adalah status kelangsungan pendidikan anak, dikategorikan sebagai: (1) anak putus sekolah dan (2) anak tidak putus sekolah. Adapun untuk delapan variabel independen (X) adalah jenis kelamin (X_1), umur kepala rumah tangga (X_2), jumlah anggota rumah tangga yang belum atau tidak bekerja (X_3), tingkat pendidikan ibu (X_4), status pekerjaan kepala rumah tangga (X_5), status ekonomi (X_6), klasifikasi tempat tinggal (X_7), dan kepemilikan saudara yang putus sekolah dan tinggal serumah (X_8).

Tahap-tahap dalam penelitian ini, pertama dilakukan yaitu analisis statistik deskriptif pada data dependen (Y). Menentukan proporsi data *training* dan data *testing* dengan perbandingan 80% dan 20% yang dilakukan secara acak. Selanjutnya dilakukan pembentukan pohon klasifikasi yang terdiri dari proses pemilihan pemilah terbaik, penentuan simpul terminal, penandaan label kelas dengan aturan jumlah terbanyak pada simpul t , dan penentuan pohon klasifikasi optimal. Tahap selanjutnya adalah prediksi nilai ketepatan akurasi pada masing-masing data *training* dan data *testing*. Adapun untuk penerapan SMOTE adalah menentukan 16 kali data buatan pada kelas minoritas. Setelah dapat data baru dengan SMOTE, dilakukan kembali pembentukan pohon klasifikasi analisis CART sama seperti cara pembentukan pohon klasifikasi CART sebelum SMOTE. Selanjutnya menentukan nilai ketepatan klasifikasi, terakhir yaitu melihat hasil kebaikan model dari kurva *Receiver Operating Characteristic*. [5].

SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE (SMOTE)

Synthetic Minority Oversampling Technique (SMOTE) adalah satu diantara turunan dari metode *oversampling*. Metode SMOTE merupakan satu diantara metode untuk mengatasi data tidak seimbang dengan replikasi data buatan untuk kelas data minoritas. Pengaruh penggunaan data tidak seimbang dalam pengklasifikasian cenderung diliputi oleh kelas mayor dan akan mengabaikan kelas minor, sehingga kesalahan klasifikasi pada kelas minor dalam kasus tertentu akan berakibat fatal dibandingkan dengan kesalahan pada kelas mayor [4]. Data buatan tersebut dibuat berdasarkan *K-Nearest Neighbor*. Seluruh variabel yang digunakan pada penelitian ini adalah variabel kategorik, perhitungan jarak antar kelas minornya dilakukan dengan rumus *Value Difference Metric* (VDM) [6].

$$\Delta(A, B) = \sum_{i=1}^N \delta(V_{1i}, V_{2i})$$

(1)

dimana:

$\Delta(A, B)$: jarak antara amatan A dengan amatan B

N : banyaknya variabel independen

$\delta(V_{1i}, V_{2i})$: jarak antara amatan A dan B untuk setiap variabel yang dihitung

Untuk menentukan jarak antar amatan A dan B untuk setiap variabel maka digunakan persamaan:

$$\delta(V_1, V_2) = \sum_{i=1}^n \left| \frac{C_{1i} - C_{2i}}{C_1 - C_2} \right|$$

(2)

dimana:

n : banyaknya kategori pada variabel ke- i

C_{1i} : banyaknya kategori ke-1 yang termasuk pada variabel ke- i

C_{2i} : banyaknya kategori ke-2 yang termasuk pada variabel ke- i

C_1 : banyaknya kategori ke-1 terjadi

C_2 : banyaknya kategori ke-2 terjadi

CLASSIFICATION AND REGRESSION TREES (CART)

CART merupakan metode statistik nonparametrik yang dapat menggambarkan hubungan antara variabel independen terhadap variabel dependennya [7]. CART terdiri dari dua analisis, yaitu *Classification Trees* dan *Regression Trees*. Jika variabel dependen yang dimiliki bertipe kategorik maka akan dihasilkan *Classification trees* (pohon klasifikasi) dan jika variabel dependennya bertipe kontinu maka akan dihasilkan *regression Trees*. Variabel yang digunakan dalam suatu pengklasifikasian adalah variabel dependen, variabel independen, data *training*, dan data *testing*. Data *training* digunakan untuk membentuk pohon klasifikasi sedangkan data *testing* digunakan untuk mengetahui seberapa tepat model yang sudah diklasifikasi yang telah dihasilkan dari data *training*.

CART merupakan metode klasifikasi yang memiliki sifat *binary recursive partitioning* (penyekatan berulang secara biner) [8].

TAHAPAN PEMBENTUKAN POHON KLASIFIKASI

Tahapan pembentukan pohon klasifikasi pada Metode *Classification and Regression Trees* sebagai berikut:

1. Pemilihan Pemilah

Pembentukan pohon klasifikasi dilakukan dengan cara memilih pemilah dari setiap simpul yang menghasilkan penurunan tingkat keheterogenan tertinggi. Untuk mengukur tingkat keheterogenan dari suatu simpul tertentu dalam pohon klasifikasi dikenal dengan istilah impuritas. Impuritas merupakan tingkat keacakan suatu simpul. Nilai impuritas dapat diperoleh sebagai berikut:

$$i(t) = \sum_{i \neq j} p(j|t)p(i|t)$$

(3)

dimana:

$i(t)$: nilai impuritas pada simpul ke t

$p(j|t)$: peluang kelas j pada simpul t

$p(i|t)$: peluang kelas i pada simpul t .

Tahapan berikutnya menentukan kriteria *goodness of split* untuk mengevaluasi pemilah dari s pada simpul t . *Goodness of split* menunjukkan ukuran penurunan keheterogenan suatu kelas. Pemilah yang menghasilkan nilai *goodness of split* tertinggi merupakan pemilah terbaik karena memungkinkan mereduksi keheterogenan secara signifikan dengan Persamaan (4) berikut:

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

(4)

dimana:

$\Delta i(s, t)$: penurunan nilai impuritas kelas ke s simpul ke t

P_L : peluang pengamatan pada simpul kiri

P_R : peluang pengamatan pada simpul kanan

$i(t_L)$: nilai impuritas simpul ke t kiri

$i(t_R)$: nilai impuritas simpul ke t kanan

2. Penentuan Simpul Terminal

suatu simpul t akan menjadi simpul terminal apabila pada simpul terminal hanya terdapat satu pengamatan pada tiap simpul anak yang dihasilkan. Jika hal ini sudah terpenuhi maka pertumbuhan pohon dapat dihentikan.

3. Penandaan Label Kelas

Penandaan label kelas pada simpul terminal didasarkan pada aturan jumlah terbanyak. Misalnya suatu variabel dependen terdiri atas keputusan ya dan tidak maka keputusan yang paling banyak dalam suatu simpul terminal tersebutlah dijadikan sebagai label kelas, yaitu dengan Persamaan (5):

$$P(j_0|t) = \max_j \frac{N_j(t)}{N(t)} \quad (5)$$

dimana:

$P(j_0|t)$: peluang kelas pada simpul ke t

$N_j(t)$: banyaknya amatan kelas ke j pada simpul terminal t

$N(t)$: jumlah total amatan dalam simpul terminal t

4. Penentuan Pohon Klasifikasi Optimal

Pohon klasifikasi yang berukuran besar akan memberikan nilai penaksir pengganti yang kecil, sehingga pohon ini cenderung dipilih untuk menaksir nilai dari variabel respon. Ukuran pohon yang besar akan menyebabkan nilai kompleksitas yang tinggi karena struktur data yang digambarkan cenderung kompleks, oleh karena itu untuk mendapatkan pohon optimal adalah dengan melakukan pemangkasan (*pruning*) secara berturut-turut pada bagian yang kurang penting. Tingkat kepentingan pohon bagian diukur berdasarkan ukuran biaya kompleksitas (*cost-complexity*) dengan persamaan (6) berikut:

$$R_\alpha(T) = R(T) + \alpha|\tilde{T}| \quad (6)$$

dimana:

$R_\alpha(T)$: nilai kompleksitas suatu pohon T

$R(T)$: tingkat kesalahan klasifikasi pada pohon T

\tilde{T} : banyaknya simpul terminal pada pohon T

α : parameter kompleksitas

UKURAN KETEPATAN KLASIFIKASI

Model klasifikasi dapat diketahui tingkat ketepatannya dengan menggunakan tiga perhitungan statistik yaitu akurasi, sensitivitas dan spesifisitas. Akurasi merupakan ukuran dari seberapa bagus model bisa mengklasifikasikan data dengan benar yang dihitung dari prediksi yang bernilai positif dan negatif. Sensitivitas merupakan tingkat ketepatan yang memperhitungkan nilai proporsi *true positive* dalam memprediksi nilai yang benar ditunjukkan sedangkan spesifisitas memperhitungkan proporsi *true negative* dalam memprediksi nilai yang benar untuk keadaan yang berlawanan dengan keinginan. Berikut adalah rumus akurasi, sensitivitas, dan spesifisitas. Tabel 1 menyajikan tabel yang merupakan alat ukur berbentuk matriks yang digunakan untuk memperoleh jumlah ketepatan klasifikasi terhadap kelas dengan algoritma yang digunakan [7]. Persamaan berikut ini digunakan untuk memperoleh nilai ketepatan klasifikasi:

$$Akurasi = \frac{TP+TN}{TP+FP+TN+FN} \quad (7)$$

$$Sensitivitas = \frac{TP}{TP+FN} \quad (8)$$

$$\text{Spesifisitas} = \frac{TN}{TN+FN}$$

(9)

Tabel 1 Confusion Matrix

	Kelas	Positif	Negatif
Nilai Aktual	Positif	TP	FN
Negatif	FP	TN	

RECEIVER OPERATING CHARACTERISTIC (ROC)

Receiver Operating Characteristic (ROC) merupakan kurva yang menggambarkan performa pengklasifikasian dengan dua dimensi [8]. Kurva ROC berisikan plot peluang salah positif (1-spesifisitas) pada sumbu X dengan prediksi benar positif (sensitivitas) pada sumbu Y. ROC dapat diubah kedalam bentuk skalar apabila digunakan untuk membandingkan beberapa performa, salah satunya menjadi *Area Under Curve* (AUC). AUC digunakan untuk mengukur klasifikasi apabila data tidak seimbang. Hal ini karena AUC menggunakan sensitivitas atau spesifisitas sebagai dasar pengukuran.

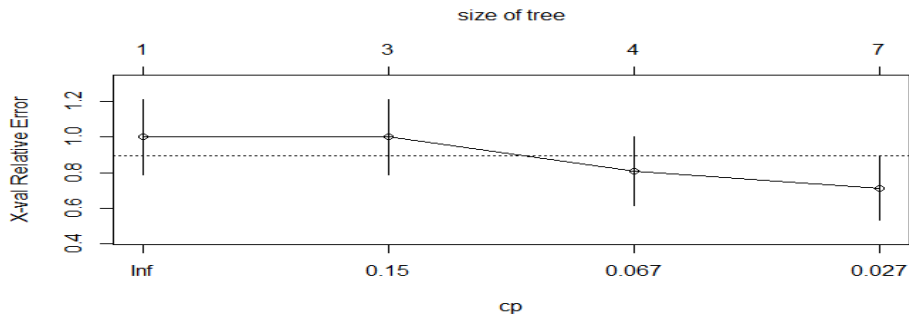
STUDI KASUS

Data yang digunakan pada penelitian ini adalah data sekunder yaitu data individu hasil Survey Sosial Ekonomi Nasional (SUSENAS) di Kabupaten Sambas tahun 2017, yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Kalimantan Barat. Sampel yang digunakan dalam penelitian ini sebanyak 494 individu adalah anak usia sekolah yang berumur 7-17 tahun. Analisis data dilakukan dengan bantuan *software* R. Variabel yang digunakan sebanyak 9 variabel terdiri dari 1 variabel dependen dan 8 variabel independen. Variabel dependen adalah status kelangsungan pendidikan anak. Adapun variabel independen adalah jenis kelamin, umur kepala rumah tangga, jumlah anggota rumah tangga yang belum atau tidak bekerja, tingkat pendidikan ibu, status pekerjaan kepala rumah tangga, status ekonomi, klasifikasi tempat tinggal, dan kepemilikan saudara yang putus sekolah dan tinggal serumah.

Analisis CART tanpa SMOTE

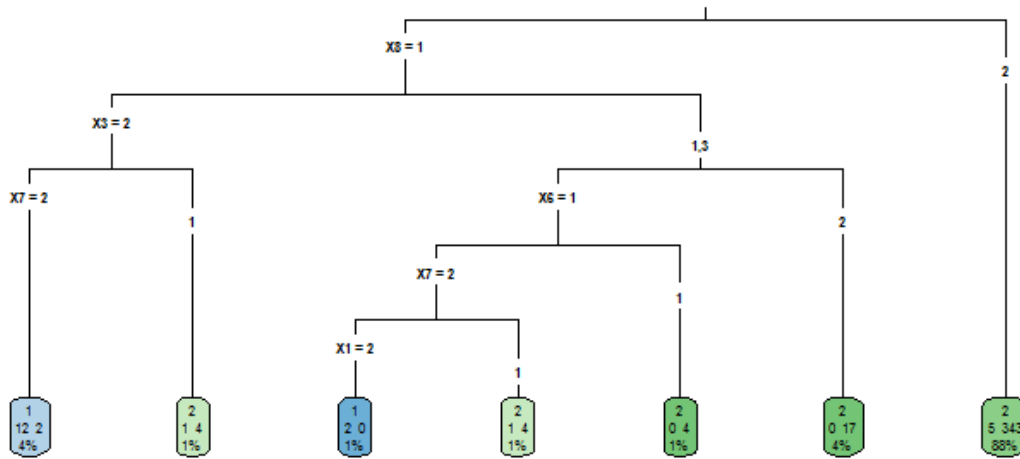
Pembentukan pohon klasifikasi dengan metode CART diawali dengan membagi menjadi data *training* dan data *testing*. Data *training* digunakan untuk membentuk pohon klasifikasi sedangkan data *testing* digunakan untuk validasi model. Pada penelitian ini proporsi masing-masing adalah 80% dan 20% atau masing-masing sebanyak 395 sampel dan 99 sampel. Pohon klasifikasi menampilkan hasil berupa pohon keputusan. Variabel yang memiliki impuritas tertinggi akan menjadi pemilah utama.

Hasil pohon klasifikasi yang besar akan menyebabkan nilai kompleksitas yang tinggi, sehingga perlu dilakukan proses pemangkasan untuk mendapatkan hasil pohon yang sederhana. Hasil dari proses pemangkasan ditentukan dengan menggunakan nilai parameter *cost-complexity* (cp) dan *cross-validation* (cv). Nilai cp yang meminimumkan nilai cv adalah pada ukuran pohon 6 simpul pemilah dan 7 simpul terminal.



Gambar 1 Plot Antara nilai cp dan cv

Berdasarkan Gambar 1 terlihat bahwa pohon klasifikasi terpankaskan memiliki 6 simpul pemilah dan 7 simpul terminal. Pada plot diatas terlihat bahwa nilai cp semakin menurun seiring bertambahnya jumlah simpul. Penambahan jumlah simpul terminal cenderung meningkatkan nilai sensitivitas namun menurunkan nilai akurasi secara keseluruhan.



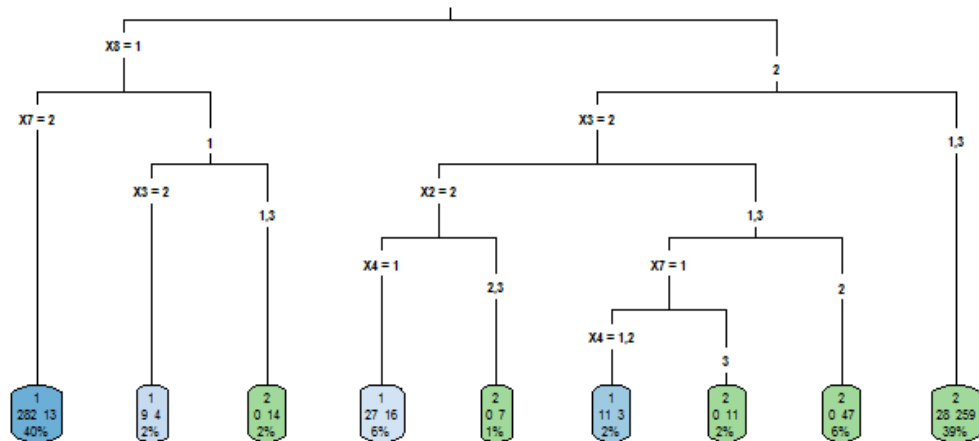
Gambar 2 Pohon Klasifikasi Optimal tanpa SMOTE

Hasil pohon klasifikasi yang terbentuk pada Gambar 2 menunjukkan bahwa klasifikasi status kelangsungan pendidikan anak, yang dikategorikan sebagai anak putus sekolah, terbentuk pada 2 simpul terminal. Sedangkan klasifikasi status kelangsungan pendidikan anak yang dikategorikan sebagai anak tidak putus sekolah terbentuk pada 5 simpul terminal. Pada semua variabel, untuk pemilihan pemilah utama ditentukan dengan menggunakan nilai impuritas tertinggi. Variabel kepemilikan saudara yang putus sekolah dan tinggal serumah (X_8) menghasilkan nilai impuritas tertinggi, dibandingkan dengan variabel lainnya yaitu sebesar 0,027. Selanjutnya variabel yang menentukan klasifikasi status kelangsungan pendidikan anak adalah jumlah anggota rumah tangga yang belum atau tidak bekerja (X_3), klasifikasi tempat tinggal (X_7), dan kepemilikan saudara yang putus sekolah dan tinggal serumah (X_8).

Analisis CART dengan penerapan SMOTE

Persentase awal jumlah amatan pada kategori minoritas adalah sebesar 6,48% sedangkan untuk kategori mayoritas sebesar 93,52%. Kemudian dilakukan SMOTE pada data training dengan beberapa kemungkinan persentase oversampling. Data sintesis dibangkitkan sebesar 16 kali data minor, yaitu $21 \times 16 = 336$ dan data baru ini ditambahkan pada data asli sehingga berjumlah $336 + 21 = 357$. Sehingga didapatkan data baru pada data minor sebanyak 357 amatan. Hasil presentase pada kategori minor adalah sebesar 48,8% dan data mayor sebesar 51,2%. Data yang telah melalui tahapan SMOTE

menjadi lebih seimbang dibandingkan sebelum diterapkan *SMOTE*. Berikut adalah pohon klasifikasi optimal yang terbentuk:



Gambar 3 Pohon Klasifikasi Optimal Dengan SMOTE

Gambar 3 menunjukkan pohon klasifikasi optimal menghasilkan 9 simpul terminal. Nilai impuritas tertinggi sebesar 0,272 pada variabel memiliki saudara yang putus sekolah dan tinggal serumah (X_8). Hal ini menunjukkan bahwa variabel memiliki saudara yang putus sekolah dan tinggal serumah (X_8) merupakan variabel yang paling dominan dalam pembentukan pohon klasifikasi. Status kelangsungan pendidikan anak yang dikategorikan sebagai anak putus sekolah terbentuk pada 4 simpul terminal, sedangkan status kelangsungan pendidikan anak yang dikategorikan sebagai anak tidak putus sekolah terbentuk pada 5 simpul terminal. Selanjutnya variabel yang menentukan klasifikasi status kelangsungan pendidikan anak adalah umur kepala rumah tangga (X_2), jumlah anggota rumah tangga yang belum atau tidak bekerja (X_3), tingkat pendidikan ibu (X_4), klasifikasi tempat tinggal (X_7), dan kepemilikan saudara yang putus sekolah dan tinggal serumah (X_8). Adapun untuk nilai ketepatan klasifikasi model dengan *SMOTE* dapat dilihat pada tabel berikut:

PERBANDINGAN TINGKAT KETEPATAN KLASIFIKASI

Perbandingan dua pohon klasifikasi yang didapatkan sebelumnya dilakukan dengan membandingkan nilai akurasi, sensitivitas, spesifisitas, dan nilai *AUC*. Tabel 2 menjelaskan Perbandingan *CART* tanpa *SMOTE* dan dengan *SMOTE* ini akan dilakukan pada data *testing*:

Tabel 2 Perbandingan *CART* tanpa *SMOTE* dan dengan *SMOTE*

Kriteria	Tanpa <i>SMOTE</i> (%)	<i>SMOTE</i> (%)
Akurasi	91,92	80,68
Sensitivitas	36,36	67,05
Spesifisitas	98,86	94,31
<i>AUC</i>	88,14	94,35

Tabel 2 menunjukkan bahwa tingkat akurasi total pohon klasifikasi tanpa *SMOTE* lebih tinggi dibandingkan pohon klasifikasi dengan *SMOTE*, namun pohon klasifikasi tanpa *SMOTE* tidak dapat melakukan klasifikasi pada kelas anak putus sekolah dengan baik. Hal tersebut dapat dilihat dari nilai sensitivitasnya yang sangat kecil, yaitu 36,36%. Kesalahan dalam mengklasifikasikan seseorang anak yang putus sekolah sebagai anak yang tidak putus sekolah tentunya akan berakibat fatal, karena kebijakan yang dilakukan oleh pemerintah untuk menanggulangi tingginya anak putus sekolah pada suatu kabupaten dapat menjadi salah sasaran. Setelah dilakukan *SMOTE* pada tahap sebelum analisis data, nilai sensitivitas yang diperoleh naik menjadi 67,05%. Hal ini dapat dilihat juga bahwa nilai *AUC*

dengan SMOTE lebih tinggi dibandingkan dengan nilai *AUC* tanpa *SMOTE* yaitu sebesar 88,14% dan 94,35%.

KESIMPULAN

Data tidak seimbang pada kategori kelas minor menyebabkan kesalahan klasifikasi pada kategori kelas minor sangat besar. Penggunaan *SMOTE* diterapkan dapat menyeimbangkan banyaknya data pada setiap anak putus sekolah. Nilai akurasi metode tanpa *SMOTE* sebesar 91,92% sedangkan pada metode dengan *SMOTE* sebesar 80,68%. Namun pada nilai sensitifitas pada kelas minor mengalami kenaikan sebesar 30,69%. Sedangkan untuk nilai *AUC* pada metode *CART* mengalami kenaikan sebesar 10,57%. Dalam kasus ini dapat disimpulkan bahwa metode *SMOTE* dapat meningkatkan nilai akurasi pada metode *CART* yang diterapkan pada data tidak seimbang.

DAFTAR PUSTAKA

- [1] Han, J., Kamber, M. *Data Mining Concept and Techniques*. USA: Academic Press; 2001.
- [2] Breiman, L., Friedman, JH., Olshen, RA., Stone, CJ. *Classification and Regression Trees*. Chapman and Hall, New York: 1993.
- [3] Lewis, JR. An Introduction to Classification and Regression Tree (*CART*). *Analysis Presented at the 2000 Annual Meeting of the Society for Academic Emergency Medicine*. San Francisco California; 2000.
- [4] Chawla, VN., Bowyer, KW., Hall, LO., Kegelmeyer, WP. Synthetic Minority Oversampling Techniques. *Journal of Artificial Inteligence Research*; 16:321-357; 2002.
- [5] Gorunescu., F. *Data Mining Concept Models and Techniques*. Berlin (GER) Springer. Verleg Berlin Heidelberg; 2011.
- [6] Cost, S., Salzberg, S. *A Weighted Nearest Neighbor Algorithm for Learning with Symbolic features*. Machine Learning. Boston (US) Kluwer Academic Publisher, 10:57-58; 1993.
- [7] Pratiwi, EF., Zain, I. Klasifikasi Pengangguran Terbuka Menggunakan *CART* (Classification and Regression Tree). *Sulawesi Utara: Jurnal sains pomits*, 3(1); 2014.
- [8] Fawcett T. *An introduction to ROC analysis*. Pattern recognition letter, 27: 861-874; 2006.

RIZKY DWI PERMATASARI : Jurusan Matematika FMIPA Untan Pontianak,
Rizkydwip@student.untan.ac.id

SETYO WIRA RIZKI : Jurusan Matematika FMIPA Untan Pontianak,
setyo.wirarizki@math.untan.ac.id

NAOMI NESSYANA DEBATARAJA : Jurusan Matematika FMIPA Untan Pontianak,
naominessyana@math.untan.ac.id
