

The SMOTE Application of CART Methods for Coping Imbalanced Data in Classifying Status Work on Labor Force in the City of Padang

Andini Yulianti, Fadhilah Fitri*, Nonong Amalita, Dodi Vionanda

Departemen Statistika, Universitas Negeri Padang, Kota Padang, Negara Indonesia

*Corresponding author: fadhilahfitri@fmipa.unp.ac.id

Submitted : 16 Agustus 2022

Revised : 04 Maret 2023

Accepted : 04 Mei 2023

ABSTRACT

Unemployment is a complex problem faced by developing countries such as Indonesia. The high unemployment rate in Indonesia has an impact on poverty so that the government seeks to carry out economic development. The city of Padang is the city that has the highest unemployment rate in West Sumatra from 2013 to 2021. One of the efforts to solve this problem is to identify the factors that influence and the characteristics of unemployment, so that the policies made by the government are not misdirected. This study uses the CART method to determine the factors that affect the number of the workforce in the city of Padang. The advantage of the CART method is that it is easy to interpret the results of the analysis, but the accuracy of the classification tree is low due to data imbalance. Therefore, this study uses the SMOTE method to overcome these problems. Based on the analysis that has been done, it can be concluded that SMOTE can increase the accuracy of the CART method applied to unbalanced data by 43.1%. The optimal classification tree is formed from 8 terminal nodes and involves 4 explanatory variables consisting of marital status, education level, gender and age.

Keywords: *CART, SMOTE, Workforce*



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Masalah ketenagakerjaan merupakan salah satu perhatian terpenting di semua negara, terutama di negara berkembang. Sebagai negara berkembang Indonesia tidak lepas dari permasalahan tersebut. Masalah ketenagakerjaan Indonesia meliputi kurangnya pekerjaan, kelebihan tenaga pekerja dan distribusi pekerjaan yang tidak merata. Hal ini dikarenakan pertumbuhan angkatan kerja yang lebih besar dari pada pertumbuhan kesempatan kerja yang ada, sehingga banyak tenaga kerja yang tidak dapat memperoleh pekerjaan sehingga menimbulkan pengangguran.

Ketersediaan lapangan kerja merupakan salah satu faktor keberhasilan ekonomi. Angkatan kerja terdiri dari penduduk yang telah mencapai usia kerja tertentu. Usia minimum di Indonesia adalah 15 tahun dan tidak ada batas usia. Oleh karena itu, mereka yang telah mencapai usia pensiun dan tetap bekerja disebut sebagai tenaga kerja (Simanjuntak, 2001). Angkatan kerja yang terdiri dari angkatan kerja dan pengangguran, disebut sebagai angkatan kerja (Dinas Ketenagakerjaan, 2019).

Kota dengan Tingkat Pengangguran Terbuka (TPT) tertinggi di Provinsi Sumatera Barat dari tahun 2013 sampai dengan tahun 2021 adalah Kota Padang yang sebenarnya memiliki potensi ekonomi yang cukup tinggi (BPS, 2021). TPT adalah jumlah pengangguran sebagai persentase dari total angkatan kerja. Untuk memudahkan pemerintah mengatasi pengangguran di Indonesia, perlu dilakukan pengidentifikasian terhadap faktor-faktor yang mempengaruhi serta karakteristik dari pengangguran. Berdasarkan hasil identifikasi yang diperoleh, ditentukan karakteristik dari pengangguran dan mendeskripsikan karakteristik masing-masing kelompok berdasarkan faktor yang mempengaruhi pengangguran. Analisis statistik yang dapat digunakan adalah analisis non parametrik dengan metode Classification and Regression Tree (CART), yang bertujuan untuk mengetahui faktor-faktor yang berpengaruh terhadap angkatan kerja yang bekerja dan pengangguran, agar kebijakan pemerintah tidak salah sasaran.

Penelitian sebelumnya oleh Pratiwi dan Zain (2014) menunjukkan bahwa usia, jenis kelamin, tingkat pendidikan, status perkawinan, dan status rumah tangga memiliki pengaruh yang signifikan terhadap kemungkinan pengangguran, tetapi tidak berpengaruh signifikan terhadap pengalaman pelatihan kerja dan tempat tinggal. Perbedaan dari penelitian sebelumnya adalah variabel dan tempat penelitian dilakukan. Penelitian ini menggunakan variabel dependen yang terdiri dari dua kategori yaitu angkatan kerja yang bekerja dan

pengangguran, dengan lima variabel independen yaitu usia, jenis kelamin, status perkawinan, tingkat pendidikan, dan kartu prakerja yang dilakukan di Kota Padang.

Metode CART merupakan salah satu metode atau algoritma dari metode pohon keputusan (*decision tree*). Metode CART dapat diterapkan pada data berdimensi tinggi karena dapat memilih variabel yang paling penting dan interaksi antar peubah yang paling penting dalam menentukan peubah responnya. CART dapat mempermudah interpretasi hasil analisis dan memberikan estimasi dengan tingkat kesalahan yang rendah (Sartono dan Syafitri, 2010). Tujuan utama dari CART adalah untuk mendapatkan sekumpulan data yang homogen sebagai fitur klasifikasi. Hasil klasifikasi ini akan digunakan untuk menentukan karakteristik masing-masing kategori angkatan kerja.

Pembentukan pohon klasifikasi dikondisikan oleh ketidakseimbangan jumlah observasi pada setiap kelas respon, yaitu ada kelas yang lebih dominan (*major*) dan kelas yang observasinya jauh lebih sedikit (*minor*). Kondisi ini dapat menyebabkan kesalahan klasifikasi pada golongan minor atau, dalam penelitian ini, golongan pengangguran. Oleh karena itu, salah satu pilihan untuk meningkatkan akurasi pada kelas ini adalah dengan melakukan teknik Synthetic Minority Oversampling Technique (SMOTE) pada pre-data processing. Dengan menggunakan model klasifikasi Decision Tree dan algoritma SMOTE sebagai penyeimbang data kelompok minoritas diharapkan dapat membantu dan menjadi solusi untuk meminimalisir pengangguran di kota Padang.

II. METODE PENELITIAN

Data yang digunakan dalam penelitian ini adalah data Survei Angkatan Kerja Nasional (SAKERNAS) Kota Padang Tahun 2021 dengan jumlah responden sebanyak 1290 orang. Data dari BPS Provinsi Sumatera Barat. Perangkat lunak yang digunakan adalah Rstudio. Variabel yang digunakan dalam penelitian ini disajikan pada Tabel 1.

Tabel 1. Variabel penelitian

Nama Variabel	Skala	Keterangan
Status Kerja (Y)	Nominal	0 : Bekerja 1 : Pengangguran
Usia (X_1)	Nominal	0 : Produktif 1 : Tidak Produktif
Jenis Kelamin (X_2)	Nominal	0 : Laki-laki 1 : Perempuan
Status Perkawinan (X_3)	Nominal	0 : Belum Pernah Kawin 1 : Sudah Pernah Kawin
Tingkat Pendidikan (X_4)	Ordinal	0 : Pendidikan Dasar 1 : Pendidikan Menengah 2 : Pendidikan Tinggi
Kartu Prakerja (X_5)	Nominal	0 : Lolos 1 : Tidak Lolos

Leo Breiman, Jerome H. Friedman, Richard A. Olshen, dan Charles J. Stone adalah pelopor yang pertama kali mengenalkan metode CART yaitu pada tahun 1984 (Breiman, 1984). Metode CART merupakan metode pendekatan regresi non parametrik yang dikembangkan untuk topik analisis klasifikasi dalam bentuk pohon (Ghiasi, 2020). Pohon klasifikasi pada metode CART terbentuk apabila data variabel dependen kategorik. Keunggulan dari metode ini adalah efektif digunakan untuk analisis sejumlah besar variabel dimana dapat memperoleh hasil yang bermanfaat dengan hanya menggunakan beberapa variabel penting. Tujuan dari metode ini adalah untuk memperoleh kelompok data yang akurat (Komalasari, 2007). Proses pembentukan pohon klasifikasi metode CART adalah sebagai berikut.

1. Pembentukan pohon klasifikasi
 - a. Pemilihan pemilah

Pemilihan pemilah adalah bergantung pada nilai yang berasal dari satu variabel independen (Breiman, 1984). Pemilihan pemilah dilakukan berdasarkan indeks Gini yang merupakan salah satu kriteria untuk menentukan titik pemecah terbaik.

$$i(t) = 1 - \sum_j p^2(j|t) \tag{1}$$

) (Breiman, 1984:19)

$p(j|t)$ merupakan proporsi kelas j pada simpul t . Probabilitas pengamatan yang masuk pada simpul kanan dan kiri dihitung dengan rumus sebagai berikut (Mardiani, 2012):

$$P_L = \frac{\text{calon simpul kiri}}{\text{data latih}} \quad (2)$$

$$P_R = \frac{\text{calon simpul kanan}}{\text{data latih}} \quad (3)$$

Untuk menentukan nilai *goodness of split* adalah sebagai berikut:

$$\Delta i(s, t) = i(t) - P_L i(t_L) - P_R i(t_R) \quad (4)$$

Dimana :

$i(t)$ = Fungsi keheterogenan pada simpul

$i(t_L)$ = Fungsi keheterogenan pada simpul anak kiri

$i(t_R)$ = Fungsi keheterogenan pada simpul anak kanan

P_L = Probabilitas pengamatan pada simpul kiri

P_R = Probabilitas pengamatan pada simpul kanan

Nilai *goodness of split* terbesar yang merupakan pemilah utama.

b. Penentuan simpul terminal

Penentuan simpul terminal dapat diketahui apabila pengamatan berjumlah kurang atau sama dengan 5 ($n < 5$) dan tingkat kedalaman pohon sudah maksimal.

c. Penandaan label kelas

Penandaan label kelas dilakukan berdasarkan aturan jumlah terbanyak sebagai berikut:

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

2. Pemangkasan pohon klasifikasi

Pemangkasan pohon klasifikasi digunakan untuk mendapatkan ukuran pohon yang layak, dengan rumus sebagai berikut:

$$g_m(t) = \frac{R(t) - R(T_k)}{|T_k| - 1} \quad (6)$$

Dimana :

$g_m(t)$ = cost complexity

$R(t)$ = Kesalahan pengklasifikasian pada node t

T_k = Subtree ke-k, dengan $k= 1,2,\dots,n$

$R(T_k)$ = Kesalahan pengklasifikasian pada pohon

Pemangkasan dilakukan pada $g_m(t)$ terkecil dengan rumus sebagai berikut:

$$g_m(t_m) = \min_{t \in T_k} g_m(t) \quad (7)$$

3. Penentuan pohon klasifikasi optimal

Penentuan ini dilakukan berdasarkan penduga *cross validation* sebagai berikut:

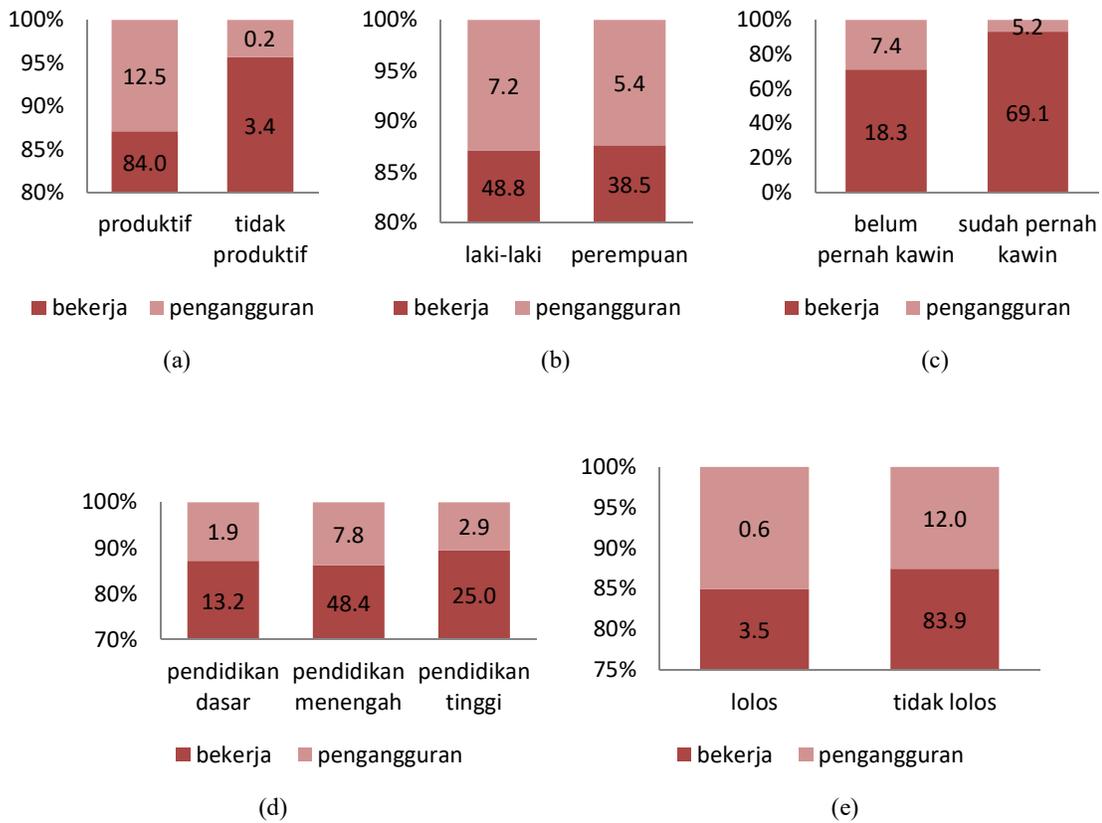
$$R(T_t^v) = \frac{1}{N_v} \sum X(d^{(v)}) \quad (8)$$

Dimana: $X(d^{(v)})$ hasil pengklasifikasian dan N_v jumlah pengamatan dalam L_v .

SMOTE diperkenalkan oleh Nithes V. Chawla (Chawla, 2002). Metode yang dilakukan pada kelas yang memiliki ketidakseimbangan data sampel berlebih (mayoritas) agar didapat data sampel yang seimbang merupakan pengertian dari metode SMOTE. Metode SMOTE bertujuan untuk dapat meningkatkan kinerja dari metode klasifikasi, dengan fokus kerja terhadap kelas negatif (minoritas). Data tidak seimbang (imbalance class) terjadi ketika jumlah objek dalam suatu kelas lebih banyak dari kelas lain. Membuat replikasi dari data minoritas merupakan cara pendekatan ini bekerja, yaitu dengan penambahan jumlah sampel pada kelas minor sehingga menghasilkan jumlah data yang seimbang antara data kelas minor dengan data pada kelas mayor yaitu dengan secara acak melakukan replikasi data pada data minoritas. Menurut Chawla, et al. (2002), permasalahan ketidakseimbangan data jika diabaikan akan menyebabkan dominasi kelas mayor dan mengabaikan kelas minor pada pengolahan data.

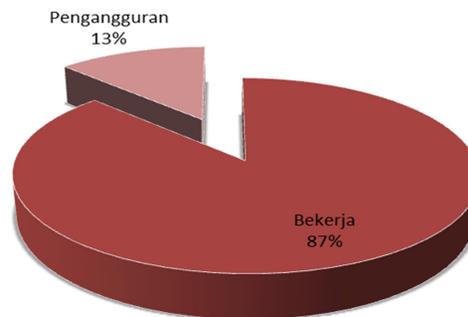
III. HASIL DAN PEMBAHASAN

Eksplorasi data berdasarkan karakteristik angkatan kerja bekerja dan pengangguran berdasarkan peubah independen.



Gambar 1. (a) Persentase berdasarkan peubah usia, (b) Persentase berdasarkan peubah jenis kelamin, (c) Persentase berdasarkan peubah status perkawinan, (d) Persentase berdasarkan peubah tingkat pendidikan, (e) Persentase berdasarkan peubah kartu prakerja

Sebelum menganalisis data, terlebih dahulu dilihat keseimbangan data yang digunakan. Deskripsi data status kerja pada tahun 2021 disajikan seperti pada Gambar 2.



Gambar 2. Deskripsi data status kerja tahun 2021

Berdasarkan Gambar 2, responden yang bekerja berjumlah 1127 orang dan pengangguran berjumlah 163 orang. Ini menunjukkan responden yang bekerja lebih banyak dari pada tidak bekerja. Sehingga terlihat bahwa data tersebut *imbalance* (tidak seimbang). Untuk menyeimbangkan data tersebut, perlu dilakukannya proses penyeimbangan data dengan SMOTE.

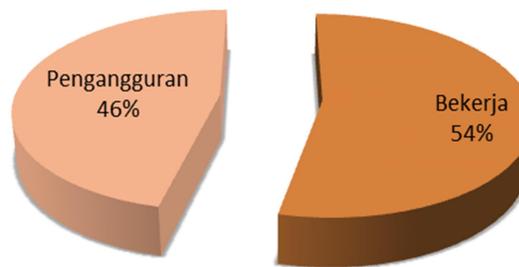
Proses SMOTE dilakukan dengan menyeimbangkan data kelas minor kelompok pengangguran dan kelas mayor kelompok bekerja. Untuk kelas minor kelompok pengangguran dengan jumlah dependen sebanyak 163 orang sehingga perlu dilakukan replikasi sebanyak 6 kali agar didapatkan jumlah dependen yang seimbang dengan

anggota pada kelas mayor. Dengan melakukan replikasi sebanyak 6 kali maka didapat jumlah tetangga terdekat untuk masing-masing data yaitu sebanyak 5 data. Titik koordinat data hasil replikasi yang dihasilkan (x_{syn}) akan berbeda dengan titik koordinat data yang direplikasi (x_i). Setelah dilakukan replikasi dengan metode SMOTE, distribusi data ditunjukkan pada Tabel 2 sebagai berikut:

Tabel 2. Distribusi data dengan menggunakan SMOTE

Sebelum Direplikasi		Setelah Direplikasi		Jumlah Replikasi
Mayor	Minor	Mayor	Minor	
(Bekerja)	(Pengangguran)	(Bekerja)	(Pengangguran)	6 kali
1127 (87%)	163 (13%)	1127 (54%)	978 (46%)	

Seperti pada Tabel 2, jumlah data yang awalnya berjumlah 163 data akan bertambah sampai dapat dikatakan jumlah anggota pada masing-masing kelas seimbang. Setelah dilakukan penyeimbangan jumlah anggota pada variabel dependen di masing-masing tipe, semoga tidak terjadi kasus overfitting atau underfitting sehingga dapat menghasilkan tingkat akurasi yang baik. Jumlah anggota untuk kedua kelas pada variabel dependen setelah dilakukan replikasi ditunjukkan pada Gambar 3 sebagai berikut:



Gambar 3. Persentase banyaknya variabel dependen setelah SMOTE

Seperti pada Gambar 3, dapat dilihat banyaknya anggota pada masing-masing tipe variabel dependen sudah seimbang dengan persentase yang tidak jauh berbeda. Setelah didapatkan jumlah data dengan proporsi kelas yang seimbang, selanjutnya dilakukan analisis classification and regression trees (CART).

Langkah analisis klasifikasi CART ini dimulai dengan menentukan pemilah utama. Pada Tabel 3 disajikan data calon pemilah.

Tabel 3. Data calon simpul kiri dan calon simpul kanan

No	Calon Simpul Kiri	Calon Simpul Kanan
1.	Usia : - Produktif	Usia : - Tidak Produktif
2.	Jenis Kelamin : - Laki-laki	Jenis Kelamin : - Perempuan
3.	Status Perkawinan : - Kawin	Status Perkawinan : - Belum Kawin
4.	Pendidikan : - Pendidikan dasar - Pendidikan menengah	Pendidikan : - Pendidikan tinggi
5.	Pendidikan : - Pendidikan dasar - Pendidikan tinggi	Pendidikan : - Pendidikan menengah
6.	Pendidikan : - Pendidikan menengah - Pendidikan Tinggi	Pendidikan : - Pendidikan dasar
7.	Kartu Prakerja - Lolos	Kartu Prakerja : - Tidak Lolos

Selanjutnya untuk memudahkan dalam menghitung indeks Gini dan kriteria *goodness of split*, maka sebaiknya dilakukan perhitungan nilai probabilitas pada setiap simpul yang disajikan pada Tabel 4.

Tabel 4. Tabel hasil perhitungan probabilitas simpul

Simpul	P_L	P_R	Kelas	$P(j t_L)$	$P(j t_R)$
1	0,973	0,023	Bekerja	0,529	0,897
			Pengangguran	0,471	0,102
2	0,558	0,428	Bekerja	0,536	0,551
			Pengangguran	0,464	0,448
3	0,385	0,61	Bekerja	0,291	0,683
			Pengangguran	0,709	0,306
4	0,733	0,257	Bekerja	0,514	0,615
			Pengangguran	0,485	0,384
5	0,414	0,576	Bekerja	0,566	0,514
			Pengangguran	0,434	0,486
6	0,843	0,144	Bekerja	0,534	0,557
			Pengangguran	0,466	0,442
7	0,044	0,955	Bekerja	0,484	0,538
			Pengangguran	0,516	0,462

Setelah memperoleh probabilitas berdasarkan simpul yang terdapat pada Tabel 4, maka selanjutnya menghitung indeks Gini dan *kriteria goodness of split* untuk menentukan parent node disajikan pada Tabel 5.

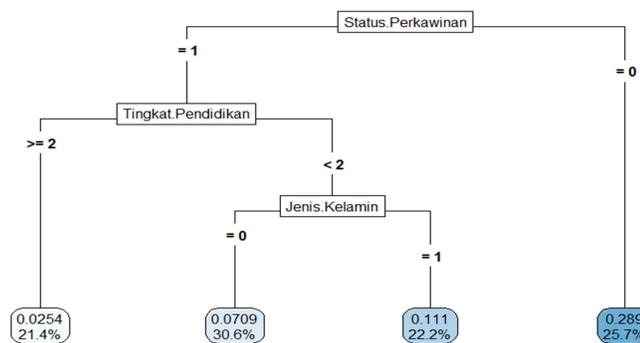
Tabel 5. Tabel hasil perhitungan indeks gini dan goodness of split

Simpul	I(t)	$\Delta i(s, t)$
1	0,052	-0,437
2	0,505	0,015
3	0,478	0,060
4	0,396	-0,092
5	0,496	0,006
6	0,268	-0,223
7	0,084	-0,413

Berdasarkan Tabel 5, diperoleh parent node pada metode CART adalah calon simpul ketiga. Calon simpul ketiga ini merupakan variabel status perkawinan. Pada Gambar 5 disajikan diagram keputusan metode CART setelah SMOTE.

A. Analisis CART tanpa SMOTE

Tahap pertama dalam penerapan metode CART adalah menentukan pemilah terbaik sebagai pemilah utama. Pemilah utama yang didapatkan berdasarkan perhitungan *indeks Gini* adalah status perkawinan. Pohon klasifikasi menghasilkan 4 simpul terminal dengan 3 peubah independen, yaitu status perkawinan (X3), tingkat pendidikan (X4), jenis kelamin (X2) dengan kedalaman pohon sebesar 3.



Gambar 4. Pohon klasifikasi tanpa SMOTE

Tahap berikutnya adalah pemangkasan pohon klasifikasi untuk memperoleh pohon optimal atau pohon yang lebih sederhana. Untuk memangkaskan pohon klasifikasi maksimal yaitu dengan melihat nilai kesalahan

relatif paling kecil dari nilai CP, sehingga mendapatkan pohon optimal. Pada kasus ini nilai kesalahan relatif paling kecil yang dihasilkan sebesar 0,933 dengan CP sebesar 0,00001 sehingga tidak perlu melakukan pemangkasan karena pohon sudah optimal. Gambar 4 adalah pohon klasifikasi optimal yang terbentuk dari 4 simpul terminal yang menjelaskan kategori bekerja. Hasil pohon klasifikasi status kerja di Kota Padang dievaluasi dengan menghitung nilai akurasi, sensitivitas, dan spesifisitas.

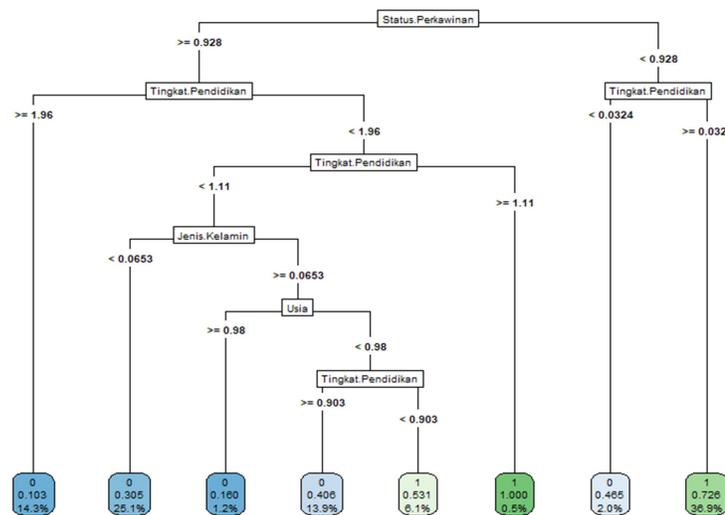
Tabel 6. Ketepatan klasifikasi CART tanpa SMOTE

Kelas Asli (i)	Kelas Hasil Prediksi (j)		Ketepatan
	Bekerja	Pengangguran	
Bekerja	854	61	93,3%
Pengangguran	273	102	27,2%
Akurasi Total			74,1%

Berdasarkan Tabel 6 ketepatan klasifikasi CART tanpa SMOTE menghasilkan nilai akurasi sebesar 74,1%, spesifisitas sebesar 93,3%, dan sensitivitas sebesar 27,2%. Ketepatan klasifikasi pada bekerja sebagai bekerja (spesifisitas) cukup tinggi yaitu sebanyak 854 individu sedangkan ketepatan klasifikasi pengangguran sebagai pengangguran (sensitivitas) cukup kecil yaitu sebanyak 273 individu. Adapun kesalahan klasifikasi bekerja sebagai pengangguran sebanyak 61 individu dan pengangguran sebagai bekerja sebanyak 102 individu.

B. Analisis CART dengan SMOTE

Penerapan CART dengan SMOTE pada penelitian ini menggunakan data asli yang digabungkan dengan data buatan hasil SMOTE. Tahap pertama melakukan pemilihan pemilah. Pemilah terbaik pada analisis CART dengan SMOTE adalah peubah status perkawinan (X3) yang dapat dilihat pada Gambar 5. Peubah status perkawinan memiliki nilai impuritas paling tinggi daripada peubah independen lainnya. Pohon klasifikasi menghasilkan 8 simpul terminal dan melibatkan 4 peubah independen, yaitu status perkawinan (X3), tingkat pendidikan (X4), jenis kelamin (X2), dan usia (X1).



Gambar 5. Pohon klasifikasi dengan SMOTE

Proses pemangkasan dilakukan setelah mendapatkan pohon maksimal. Nilai kesalahan relatif minimum yang didapatkan pada pohon klasifikasi CART dengan SMOTE adalah sebesar 0,72 dengan CP 0,00001. Hal tersebut menunjukkan bahwa tidak perlu dilakukan pemangkasan karena pohon dapat dikatakan sudah optimal. Pohon menghasilkan 5 simpul terminal yang mengklasifikasikan angkatan kerja berkategori bekerja dan 3 simpul terminal yang mengklasifikasikan angkatan kerja berkategori pengangguran.

Tabel 7. Ketepatan klasifikasi CART dengan SMOTE

Kelas Asli (i)	Kelas Hasil Prediksi (j)		Ketepatan
	Bekerja	Pengangguran	
Bekerja	854	331	72,1%
Pengangguran	273	647	70,3%
Akurasi Total			71,3%

Setelah dilakukan penerapan model pohon klasifikasi dengan SMOTE pada data uji menghasilkan nilai akurasi dan sensitivitas yang menurun, namun nilai spesifitas meningkat. Tabel 7 menjelaskan tentang ketepatan klasifikasi dengan nilai akurasi sebesar 71,3%, spesifitas (ketepatan klasifikasi bekerja sebagai bekerja) sebesar 72,1%, dan sensitivitas (ketepatan klasifikasi pengangguran sebagai pengangguran) sebesar 70,3%.

IV. KESIMPULAN

Data angkatan kerja yang digunakan merupakan kasus data tidak seimbang. Analisis klasifikasi CART menghasilkan kinerja yang tidak baik dalam mengklasifikasikan data angkatan kerja. Hal ini dilihat berdasarkan nilai sensitivitas yang dihasilkan yaitu sebesar 27,2%. Setelah dilakukan penanganan data tidak seimbang menggunakan metode SMOTE, analisis klasifikasi SMOTE CART memberikan hasil yang lebih baik dalam mengklasifikasikan kelas angkatan kerja. Hal ini dilihat berdasarkan nilai sensitivitas yang meningkat menjadi 71,8%. Jadi, pada penelitian ini analisis klasifikasi SMOTE CART mampu memberikan kinerja yang lebih baik dari pada metode CART sehingga metode SMOTE CART merupakan metode terbaik dalam mengklasifikasikan angkatan kerja di Kota Padang. Berdasarkan model SMOTE CART.

DAFTAR PUSTAKA

- Breiman L., Friedman J.H Olshen R.A & Stone C.J. 1984. *Classification And Regression Tree*. New York, NY: Chapman And Hall.
- Badan Pusat Statistik (BPS).” Jumlah Penduduk Menurut Kabupaten/Kota di Provinsi Sumatera Barat (Jiwa), 2021”. Publikasi BPS diakses dari <https://sumbar.bps.go.id/>, tanggal 24 Maret 2022.
- Chawla, N. V, Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE : Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357.
- Dinas Tenaga Kerja, ”Masalah Tenaga Kerja dan Angkatan Kerja Di Indonesia 2019”, 2019.
- Pratiwi, F. E dan Zain, I. 2014. Klasifikasi Pengangguran Terbuka Menggunakan CART di Provinsi Sulawesi Utara. *Jurnal Sains dan Seni Pomits*. 3(1): 2337-3520.
- Sartono B. Syafitri UD. 2010. Metode Pohon Gabungan: solusi pilihan untuk mengatasi kelemahan pohon regresi dan klasifikasi tunggal. *From Statistika dan Komputasi*. 15(1): 1-7.
- Simanjutak, P. J. 2001. *Pengantar Ekonomi Sumber Daya Manusia*. Jakarta: Lembaga Penerbit Fakultas Ekonomi Universitas Indonesia.
- WB Kolmasari. 2007. Metode Pohon Regresi untuk Eksploratori Data dengan Peubah yang Banyak dan Kompleks. *Informatika Pertanian*, 16(1): 967- 980.