

Perbandingan CART dan SMOTE CART dalam Mengklasifikasikan Kebutuhan KB Tidak Terpenuhi di Indonesia*

Ulfa Afilia Shofa¹, Muhammad Nur Aidi^{2‡}, and Budi Susetyo³

^{1,2,3}Department of Statistics, IPB University, Indonesia

[‡]corresponding author: muhammadai@apps.ipb.ac.id

Copyright © 2022 Ulfa Afilia Shofa, Muhammad Nur Aidi, and Budi Susetyo. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Indonesia is ranked fourth in the world as the country with the largest population. The high population growth in Indonesia can cause problems in several fields. The government seeks to suppress the rate of population growth through the Family Planning (KB) program. In Indonesia, the number of unmet needs for family planning is still relatively high and has not yet reached the BKKBN target. Therefore, it is necessary to identify the characteristics of unmet need for family planning among married women or living with partner. This study used the *Classification and Regression Trees (CART)* method. This study handling unbalanced data by *Synthetic Minority Oversampling Technique (SMOTE)*. This study aims to compare the performance of the CART and SMOTE CART classification methods in classifying unmet need for family planning and to identify the characteristics of unmet need for family planning among married women or living with partner in Indonesia. The SMOTE CART model has better performance than the CART model, with the percentages of balanced accuracy, sensitivity, and specificity being respectively 54,83%, 34,96%, and 74,70%. In general, the characteristics of unmet need for family planning among married women or living with partner in Indonesia are having 1-4 living children, not getting information from mass media, not accessing the internet in the last month, having a primary or secondary education level, a husband with no education or with a primary or secondary education level, and aged more than 30 years old.

Keywords: CART, SMOTE CART, unmet need for family planning

* Received: Jan 2022; Reviewed: May 2022; Published: May 2022

1. Pendahuluan

Indonesia merupakan negara dengan jumlah penduduk terbanyak keempat di dunia setelah Cina, India, dan Amerika Serikat. Hasil Sensus Penduduk 2020 yang dirilis oleh Badan Pusat Statistik (BPS) menunjukkan bahwa jumlah penduduk Indonesia mencapai 270,20 juta jiwa dengan laju pertumbuhan penduduk sebesar 1,25% (BPS 2021). Tingginya pertumbuhan penduduk di Indonesia dapat berpotensi menyebabkan berbagai masalah di beberapa bidang. Oleh karena itu, Pemerintah berupaya untuk menekan laju pertumbuhan penduduk melalui program Keluarga Berencana (KB). Kebutuhan akan KB di Indonesia pada kenyataannya belum sepenuhnya terpenuhi. Hal ini dapat dilihat dari masih banyak Pasangan Usia Subur (PUS) yang kebutuhan KBnya tidak terpenuhi. Kebutuhan KB tidak terpenuhi merujuk pada wanita subur berusia 15–49 tahun berstatus menikah yang ingin menunda kehamilan atau tidak ingin memiliki anak lagi tetapi tidak menggunakan alat/metode kontrasepsi (Bradley *et al.* 2012). Berdasarkan laporan hasil Survei Indikator Kinerja Program Kependudukan Keluarga Berencana dan Pembangunan Keluarga, Rencana Pembangunan Jangka Menengah Nasional (KKBP RPJMN) Tahun 2017, kebutuhan KB tidak terpenuhi secara nasional masih sebesar 17,5% (Puslitbang KB dan KS BKKBN 2017). Angka kebutuhan KB tidak terpenuhi di Indonesia masih belum mencapai target BKKBN, dimana berdasarkan pada buku Rencana Strategis (Renstra) BKKBN 2015-2019 sasaran BKKBN yaitu menurunkan kebutuhan KB tidak terpenuhi (*unmet need* KB) pada tahun 2017 menjadi sebesar 10,26%. Berdasarkan Rencana Strategis (Renstra) BKKBN 2020–2024, BKKBN menargetkan angka kebutuhan KB tidak terpenuhi di Indonesia tahun 2020 dan 2021 secara berturut-turut sebesar 8,6% dan 8,3%.

Kebutuhan KB tidak terpenuhi mengakibatkan tingginya angka kelahiran total dan berdampak pada pertumbuhan populasi yang cepat. Selain itu, kebutuhan KB tidak terpenuhi bagi wanita yang kehamilannya tidak diinginkan akan membuat wanita tersebut rentan terhadap berbagai masalah seperti aborsi yang tidak aman, kematian ibu, kekurangan gizi, dan penyakit mental (Dejene *et al.* 2021). Mengingat bahwa kebutuhan KB tidak terpenuhi di Indonesia masih perlu mendapat perhatian, maka perlu diidentifikasi karakteristik yang banyak ditemui dari wanita usia subur berstatus menikah yang memiliki kebutuhan KB tidak terpenuhi. Salah satu analisis yang dapat digunakan adalah analisis klasifikasi. Penelitian ini hanya berfokus pada wanita usia subur berstatus menikah atau memiliki pasangan. Peubah yang digunakan pada penelitian ini merupakan peubah yang digunakan pada penelitian sebelumnya oleh Asif dan Pervaiz (2019) yang berjudul *Socio-demographic Determinants of Unmet Need for Family Planning Among Married Women in Pakistan* dengan menambahkan peubah frekuensi penggunaan internet satu bulan terakhir. Metode klasifikasi yang digunakan pada penelitian ini adalah *Classification and Regression Trees* (CART).

Classification and Regression Trees (CART) merupakan metode klasifikasi nonparametrik dari salah satu teknik pohon keputusan. CART memberikan hasil yang mudah diinterpretasikan dan mampu menghasilkan *accuracy* hasil klasifikasi yang baik. Berdasarkan beberapa penelitian sebelumnya yang membandingkan metode CART dengan metode klasifikasi pohon keputusan yang lain, metode CART memberikan hasil klasifikasi yang lebih baik. Hal ini berdasarkan penelitian yang

dilakukan oleh Aritonang *et al.* (2016) yang membandingkan metode CART dan CHAID, Alverina *et al.* (2018) yang membandingkan metode CART dan C4.5, Assiroj *et al.* (2018) yang membandingkan metode CART dan C5.0, dan Shirali *et al.* (2018) yang membandingkan metode CART dan CHAID. Metode CART telah banyak digunakan dalam kasus klasifikasi di beberapa bidang, namun belum pada penelitian terkait kebutuhan KB. Hal ini dilihat dari beberapa penelitian terdahulu terkait kebutuhan KB tidak terpenuhi, diantaranya penelitian oleh Anggraeni dan Budiantara (2016) menggunakan Regresi Nonparametrik *Spline*, penelitian oleh Wulifan *et al.* (2017) menggunakan metode Regresi Logistik Multilevel, dan penelitian oleh Widyatami *et al.* (2021) menggunakan metode Regresi Logistik Biner. Oleh karena itu, pada penelitian ini akan dilakukan analisis klasifikasi kebutuhan KB tidak terpenuhi di Indonesia menggunakan metode CART.

Data yang digunakan dalam penelitian ini termasuk kasus data tidak seimbang. Data dikatakan tidak seimbang apabila suatu kelas data memiliki jumlah amatan yang jauh lebih sedikit dibandingkan kelas lainnya. Menurut Krawczyk (2016), rasio data tidak seimbang berkisar dari 1:4 hingga 1:100. Perbedaan kategori status kebutuhan KB terpenuhi dan tidak terpenuhi pada penelitian ini memiliki perbedaan proporsi yang cukup tinggi, sehingga data pada penelitian ini termasuk data tidak seimbang. Pengklasifikasian pada data tidak seimbang cenderung memiliki bias pada hasil klasifikasi untuk kelas mayor dan menghasilkan kinerja klasifikasi yang buruk pada kelas minor (Torres *et al.* 2016). Salah satu metode yang dapat digunakan untuk menangani masalah data tidak seimbang adalah metode *Synthetic Minority Oversampling Technique* (SMOTE). Menurut Chawla *et al.* (2002), SMOTE dapat meningkatkan *accuracy* pengklasifikasi untuk kelas minor. Berdasarkan penelitian yang dilakukan oleh Wijaya *et al.* (2018), dengan menerapkan SMOTE untuk menangani masalah data tidak seimbang mampu meningkatkan nilai sensitivitas (*sensitivity*) dari 0% menjadi 44,4%. Oleh karena itu, penanganan data tidak seimbang pada penelitian ini dilakukan menggunakan SMOTE. Penelitian ini bertujuan untuk membandingkan kinerja metode klasifikasi CART dan SMOTE CART dalam mengklasifikasikan kebutuhan KB tidak terpenuhi di Indonesia serta mengidentifikasi karakteristik wanita usia subur berstatus menikah atau memiliki pasangan yang kebutuhan KBnya tidak terpenuhi berdasarkan hasil klasifikasi terbaik.

2. Metodologi

2.1 Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari hasil Survei Demografi dan Kesehatan Indonesia (SDKI) tahun 2017 dari Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN). Objek penelitian pada penelitian ini adalah wanita usia subur bersatus menikah atau memiliki pasangan yang tersebar di 34 provinsi di Indonesia. Adapun peubah-peubah yang digunakan pada penelitian ini dapat dilihat pada Tabel 1.

Tabel 1 Peubah yang digunakan

Kode	Peubah	Keterangan
Y	Status kebutuhan KB	0: Terpenuhi 1: Tidak terpenuhi
X ₁	Usia istri (tahun)	1: 15–19 tahun 2: 20–24 tahun 3: 25–29 tahun 4: 30–34 tahun 5: 35–39 tahun 6: 40–44 tahun 7: 45–49 tahun
X ₂	Tingkat pendidikan istri	0: Tidak sekolah 1: Pendidikan dasar 2: Pendidikan menengah 3: Pendidikan tinggi
X ₃	Status bekerja istri	0: Tidak bekerja 1: Bekerja
X ₄	Tingkat pendidikan suami	0: Tidak sekolah 1: Pendidikan dasar 2: Pendidikan menengah 3: Pendidikan tinggi
X ₅	Kuintil kekayaan	1: Terbawah 2: Menengah bawah 3: Menengah 4: Menengah atas 5: Teratas
X ₆	Banyak anak hidup	0: Tidak ada 1: 1–2 anak 2: 3–4 anak 3: 5 anak atau lebih
X ₇	Daerah tempat tinggal	1: Perkotaan 2: Pedesaan
X ₈	Mendapat informasi KB dari media massa	0: Tidak 1: Ya
X ₉	Frekuensi penggunaan internet sebulan terakhir	0: Tidak sama sekali 1: Kurang dari seminggu sekali 2: Setidaknya seminggu sekali 3: Hampir setiap hari

Pengkategorian pada masing-masing peubah mengacu pada penelitian sebelumnya, yaitu peubah yang digunakan pada penelitian Asif dan Pervaiz (2019). Peubah frekuensi penggunaan internet sebulan terakhir ditambahkan atas dasar inisiatif peneliti yang menduga bahwa peubah ini memiliki hubungan dengan kasus kebutuhan KB.

2.2 Prosedur Analisis Data

Analisis data pada penelitian ini dilakukan dengan menggunakan bantuan *software Microsoft Excel* dan RStudio versi 3.6.3. Tahapan analisis yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Melakukan pra-proses data. Tahapan ini dilakukan pelabelan data pada peubah respon atau peubah status kebutuhan KB. Proses ini dilakukan dengan melakukan pelabelan ulang terhadap label yang sudah tersedia pada data SDKI 2017. Apabila status kebutuhan KB pada data SDKI 2017 dikategorikan "*met need for limiting*" atau "*met need for spacing*" maka akan diganti labelnya menjadi 0 untuk wanita yang kebutuhan KBnya terpenuhi dan apabila dikategorikan "*unmet need for limiting*" atau "*unmet need for spacing*" maka akan diganti labelnya menjadi 1 untuk wanita yang kebutuhan KBnya tidak terpenuhi. Selain itu, pada tahapan ini juga dilakukan pengecekan data hilang.
2. Melakukan eksplorasi data untuk melihat persentase dari kebutuhan KB terpenuhi dan tidak terpenuhi serta melihat persebaran kebutuhan KB tidak terpenuhi di Indonesia.
3. Membagi dataset menjadi data latih utama dan data uji utama secara acak dan proporsional dengan perbandingan data latih utama 70% dan data uji utama 30%. Data latih utama kemudian dibagi lagi menjadi data latih dan data uji menggunakan metode validasi silang *5-fold*. Proses pengacakan pada tahapan ini dilakukan menggunakan *set.seed(044)*.
4. Melakukan analisis klasifikasi menggunakan metode CART dengan tahapan sebagai berikut:
 - a. Membangun model klasifikasi CART menggunakan data latih. Model yang dihasilkan berupa pohon klasifikasi. Tahapan pembentukan pohon klasifikasi menurut Breiman *et al.* (1993) meliputi pemilihan pemilah, penentuan simpul terminal, penandaan label kelas, dan pemangkasan pohon. Pembentukan pohon pada penelitian ini dilakukan menggunakan batasan jumlah amatan minimum pada simpul (*minsplit*) sebesar 2 amatan, parameter biaya kompleksitas (*CP*) minimum sebesar 0, dan kedalaman pohon maksimal (*maxdepth*) sebesar 5. Pembentukan pohon klasifikasi pada penelitian ini menggunakan *set.seed(044)*.
 - b. Melakukan evaluasi pohon klasifikasi yang terbentuk dengan melihat nilai *accuracy*, sensitivitas, spesifisitas, dan *balanced accuracy* yang dihasilkan. Tahapan ini dilakukan menggunakan data uji hasil dari proses validasi silang *5-fold*.
 - c. Model dengan nilai sensitivitas atau *balanced accuracy* tertinggi akan dicobakan pada data uji utama dan akan dibandingkan dengan model utama atau model yang dibentuk menggunakan data latih utama.
 - d. Model CART yang akan dipilih dipertimbangkan berdasarkan nilai sensitivitas atau *balanced accuracy* yang dihasilkan.
 - e. Melakukan pemangkasan pohon klasifikasi CART yang dihasilkan dari tahap 4d berdasarkan nilai parameter biaya kompleksitas (*CP*) untuk mendapat pohon klasifikasi yang optimal.
5. Melakukan penanganan ketidakseimbangan data pada data latih dengan menggunakan metode SMOTE. Proses SMOTE dilakukan dengan menggunakan jumlah *k* tetangga terdekat sebanyak 5 dengan persentase

oversampling sebesar 200% dan persentase *undersampling* sebesar 150%. Tahapan ini menggunakan *set.seed(044)*.

6. Melakukan analisis klasifikasi menggunakan metode CART dengan SMOTE (SMOTE CART) dengan tahapan seperti pada tahap 4 menggunakan data baru hasil dari proses SMOTE.
7. Mengevaluasi kinerja CART dan SMOTE CART dengan membandingkan nilai *accuracy*, sensitivitas, spesifisitas, dan *balanced accuracy* dari kedua model yang terpilih. Model dengan nilai sensitivitas tertinggi akan dipilih sebagai model terbaik karena fokus penelitian ini pada kelas minor atau kelas kebutuhan KB tidak terpenuhi sehingga model dengan nilai sensitivitas tertinggi mampu menghasilkan ketepatan hasil klasifikasi untuk kelas kebutuhan KB tidak terpenuhi lebih baik.
8. Menginterpretasikan hasil dari pohon klasifikasi terbaik untuk mengetahui karakteristik wanita usia subur berstatus menikah atau memiliki pasangan yang kebutuhan KBnya tidak terpenuhi.

3. Hasil dan Pembahasan

3.1 Eksplorasi Data

Total amatan pada penelitian ini sebanyak 24.883 wanita. Dari total amatan tersebut, sebanyak 83,8% wanita usia subur berstatus menikah atau memiliki pasangan di Indonesia memiliki kebutuhan KB terpenuhi, namun ada sebanyak 16,2% wanita yang kebutuhan KBnya tidak terpenuhi. Wanita yang kebutuhan KBnya tidak terpenuhi merupakan wanita subur berstatus menikah yang ingin menunda memiliki anak atau tidak ingin memiliki anak lagi tetapi tidak menggunakan alat/metode kontrasepsi. Berdasarkan persentase tersebut menunjukkan bahwa program KB di Indonesia belum sepenuhnya berjalan dengan baik, terbukti dari masih ada sekitar 4.025 wanita yang kebutuhan KB-nya tidak terpenuhi. Selain itu, perbedaan persentase antara kedua kategori tersebut cukup tinggi, sehingga dapat dikatakan bahwa data yang digunakan pada penelitian ini termasuk data tidak seimbang.



Gambar 1 Persebaran kebutuhan KB tidak terpenuhi di Indonesia

Berdasarkan Gambar 1, terlihat bahwa Provinsi Jawa Barat memiliki kasus kebutuhan KB tidak terpenuhi tertinggi di Indonesia yang ditandai dengan warna kuning, sedangkan provinsi dengan kasus kebutuhan KB tidak terpenuhi terendah berada di

Provinsi Yogyakarta yang ditandai dengan warna ungu tua. Secara nasional, kasus kebutuhan KB tidak terpenuhi di Indonesia tertinggi berada di Pulau Jawa dengan persentase sebesar 29,54%, selanjutnya di Pulau Sumatera dengan persentase sebesar 21,39%, Pulau Sulawesi dengan persentase 17,09%, Pulau Papua dengan persentase 12,55%, Pulau Nusa Tenggara dan Bali dengan persentase sebesar 11,23%, dan terendah di Pulau Kalimantan dengan persentase sebesar 8,20%.

3.2 Pembagian Data dengan Validasi Silang K-Fold

Pembagian data menjadi data latih dan data uji pada proses pemodelan dilakukan untuk mengukur tingkat *accuracy* dari model yang dibentuk. Ketika ingin mengevaluasi kinerja suatu model dapat menggunakan metode validasi silang *K-fold* dengan cara membagi data menjadi *K* bagian secara acak. Banyaknya *K* yang digunakan biasanya $K = 5$ atau $K = 10$ (James et al. 2013). Pada penelitian ini *K* yang digunakan sebanyak 5 karena untuk mempercepat proses komputasi. Pembagian data pada penelitian ini dilakukan dengan membagi dataset secara acak menjadi 70% data latih utama dan 30% data uji utama. Data yang termasuk data latih utama dibagi lagi menggunakan metode validasi silang *5-fold*. Data latih utama dibagi secara acak menjadi lima bagian atau fold dengan banyaknya amatan di setiap bagian relatif sama. Iterasi pertama, bagian pertama digunakan sebagai data uji dan empat bagian lainnya digunakan sebagai data latih. Iterasi kedua, bagian kedua digunakan sebagai data uji dan bagian lainnya digunakan sebagai data latih begitupun seterusnya. Data latih, digunakan untuk membangun model klasifikasi dan data uji digunakan untuk mengevaluasi kinerja dari model klasifikasi. Proses validasi silang ini akan menghasilkan lima data latih dan lima data uji.

3.3 Analisis Klasifikasi CART

Model klasifikasi CART yang terbentuk berupa pohon klasifikasi karena peubah respon yang digunakan bertipe kategorik. Pohon klasifikasi CART yang terbentuk pada penelitian ini merupakan pohon klasifikasi tunggal dan dibentuk menggunakan *set.seed(044)*, sehingga apabila pembentukan pohon klasifikasi dilakukan menggunakan *set.seed* yang berbeda kemungkinan akan menghasilkan pohon klasifikasi yang berbeda juga. Pembentukan pohon klasifikasi dilakukan melalui proses pemilahan, yaitu dengan memecah simpul induk menjadi dua simpul anak yang dilakukan secara rekursif. Pohon klasifikasi CART dibentuk dengan batasan jumlah amatan minimum pada setiap simpul (*minsplitt*) sebesar 2, parameter biaya kompleksitas (*CP*) minimum sebesar 0, dan kedalaman pohon maksimal (*maxdepth*) sebesar 5. Batasan *minsplitt* ditentukan berdasarkan hasil uji coba yang menghasilkan *minsplitt* dengan *accuracy* tertinggi, yaitu ketika *minsplitt* sebesar 2. Parameter biaya kompleksitas yang digunakan sebesar 0 karena agar menghasilkan pohon klasifikasi maksimal. Kedalaman pohon maksimal ditentukan sebesar 5 agar pohon klasifikasi yang dihasilkan tidak terlalu kompleks sehingga mudah untuk diinterpretasikan. Berikut merupakan performa hasil klasifikasi dari proses validasi silang *5-fold*:

Tabel 2 Performa model klasifikasi CART dalam setiap *fold*

Fold	Accuracy (%)	Sensitivitas (%)	Spesifisitas (%)	Balanced Accuracy (%)
1	85,36	1,57	99,73	50,65
2	85,28	1,76	99,63	50,69
3	85,27	0,78	99,80	50,29
4	85,33	1,56	99,80	50,68
5	85,30	0,39	99,73	50,06
Rata-rata	85,31	1,21	99,74	50,65

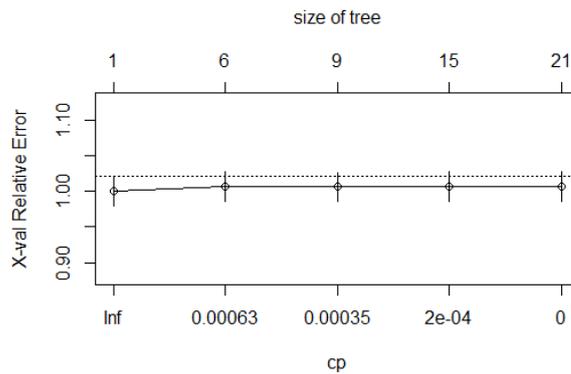
Nilai *accuracy*, sensitivitas, spesifisitas dan *balanced accuracy* dari model klasifikasi CART untuk masing-masing *fold* dapat dilihat pada Tabel 2. Nilai *accuracy* menunjukkan ketepatan hasil klasifikasi secara keseluruhan, nilai sensitivitas menunjukkan ketepatan hasil klasifikasi untuk kelas minor yaitu status kebutuhan KB tidak terpenuhi, nilai spesifisitas menunjukkan ketepatan hasil klasifikasi untuk kelas mayor yaitu status kebutuhan KB terpenuhi, dan nilai *balanced accuracy* menunjukkan rata-rata *accuracy* yang dihasilkan dari kedua kelas mayor dan minor. Menurut Bekkar *et al.* (2013), pada kasus data tidak seimbang *accuracy* memberikan bobot yang lebih besar pada kelas mayor daripada kelas minor. Hal ini dapat menyulitkan pengklasifikasi untuk memprediksi hasil klasifikasi pada kelas minor. Oleh karena itu, *accuracy* dapat menjadi indikator yang menyesatkan dalam menilai kinerja model klasifikasi pada data yang tidak seimbang. Untuk mengatasi hal tersebut salah satu ukuran kinerja pada data tidak seimbang dapat diukur menggunakan *balanced accuracy*, sehingga pada penelitian ini lebih memperhatikan nilai *balanced accuracy* dibandingkan nilai *accuracy*. Perbedaan nilai *balanced accuracy* pada masing-masing *fold* tidak terlalu berbeda jauh, hal ini menunjukkan bahwa secara umum data dalam setiap *fold* memberikan kontribusi yang cenderung sama yaitu dengan rata-rata *balanced accuracy* sebesar 50,65%. Penentuan model klasifikasi CART terbaik dilakukan dengan membandingkan performa model utama (model yang dibangun menggunakan data latih utama) dengan model *fold* yang dicobakan pada data uji utama. Model *fold* yang terpilih untuk dibandingkan dengan model utama adalah model *fold* 2 karena menghasilkan nilai sensitivitas dan *balanced accuracy* tertinggi.

Tabel 3 Perbandingan performa model klasifikasi CART

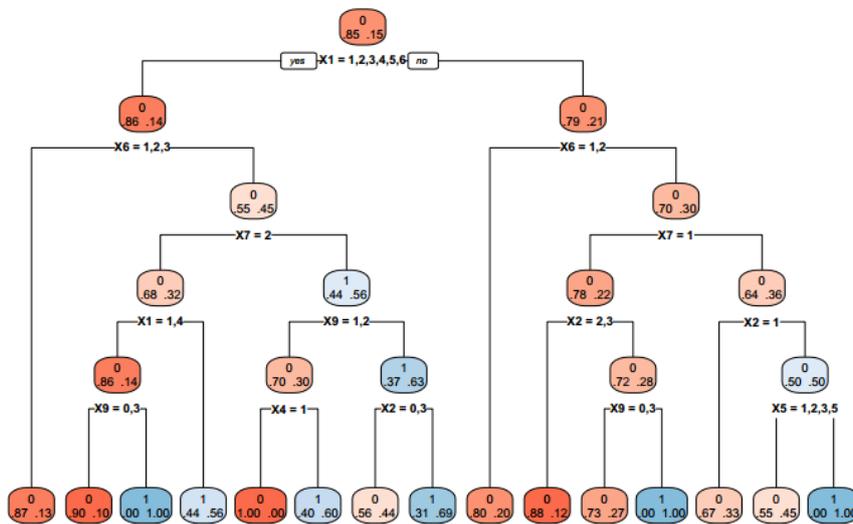
Model	Accuracy (%)	Sensitivitas (%)	Spesifisitas (%)	Balanced Accuracy (%)
Model utama	83,71	0,41	99,77	50,09
Model <i>fold</i> 2	83,83	1,82	99,65	50,74

Setelah dilakukan evaluasi model terhadap data uji utama, diperoleh hasil sesuai pada Tabel 3. Model yang ditetapkan sebagai model CART terbaik adalah model *fold* 2 karena menghasilkan nilai sensitivitas dan *balanced accuracy* yang lebih tinggi dibandingkan model utama. Berdasarkan Tabel 2 dan 3, nilai sensitivitas yang dihasilkan model CART sangat rendah. Hal ini dikarenakan data yang digunakan untuk membangun model tidak seimbang. Kondisi ini sesuai dengan pernyataan Torres *et al.* (2016), dimana pengklasifikasian pada data yang tidak seimbang menghasilkan ketepatan hasil klasifikasi pada data kelas minor yang buruk karena model akan cenderung mengklasifikasikan data pada kelas mayor.

Pohon klasifikasi CART yang terbentuk dari model *fold 2* terdiri dari 21 simpul terminal yang terdiri dari 8 simpul terminal dengan label status kebutuhan KB tidak terpenuhi dan 13 simpul terminal dengan label status kebutuhan KB terpenuhi. Setelah pohon klasifikasi terbentuk, tahap selanjutnya adalah proses pemangkasan pohon klasifikasi maksimal untuk mendapatkan pohon yang lebih optimal. Proses pemangkasan dilakukan dengan menggunakan validasi silang 10-*fold* dan dari proses ini dihasilkan nilai kesalahan relatif. Pemangkasan dilakukan berdasarkan parameter biaya kompleksitas atau *CP*. *CP* yang menghasilkan nilai kesalahan relatif (*xerror*) paling kecil dipilih untuk memangkas pohon klasifikasi maksimal. Berdasarkan Gambar 2, nilai kesalahan relatif paling kecil diperoleh ketika *CP* bernilai 0,0002, sehingga pohon klasifikasi dipangkas dengan nilai *CP* = 0,0002. Hasil pohon klasifikasi optimal menggunakan metode CART dapat dilihat pada Gambar 3.



Gambar 2: Plot *CP* CART



Gambar 3 Pohon klasifikasi CART optimal berdasarkan model terbaik

Pohon klasifikasi CART optimal yang terbentuk terdiri dari 15 simpul terminal yang melibatkan peubah usia istri (X_1), banyak anak hidup (X_6), daerah tempat tinggal (X_7), tingkat pendidikan istri (X_2), frekuensi penggunaan internet sebulan terakhir (X_9), tingkat pendidikan suami (X_4), dan kuintil kekayaan (X_5). Peubah usia istri merupakan pemilah terbaik karena menghasilkan nilai *goodness of split* tertinggi dengan kriteria pemilahan usia istri 15–44 tahun pada simpul anak kiri dan usia istri 45–49 tahun pada simpul anak kanan. Berdasarkan Gambar 3, dapat dilihat bahwa dalam setiap simpul memuat label dari simpul tersebut (nilai yang di atas) beserta nilai peluang untuk

masing-masing kategori (nilai yang di bawah). Nilai peluang di dalam simpul pada bagian kiri menunjukkan peluang dari kategori status kebutuhan KB terpenuhi dan bagian kanan menunjukkan kategori status kebutuhan KB tidak terpenuhi. Menurut Chen *et al.* (2006), default *threshold* pada pohon klasifikasi biner adalah 0,5, sehingga pada penelitian ini menggunakan *threshold* peluang 0,5 untuk menentukan prediksi kelas pada simpul terminal. Simpul terminal akan diberi label 1 (kelas kebutuhan KB tidak terpenuhi) jika peluang simpul tersebut lebih dari 0,5 dan akan diberi label 0 (kelas kebutuhan KB terpenuhi) jika peluang simpul tersebut kurang dari sama dengan 0,5. Dari 15 simpul terminal yang dihasilkan terdapat 6 simpul dengan kategori status kebutuhan KB tidak terpenuhi dan 9 simpul dengan kategori status kebutuhan KB terpenuhi. Simpul yang berwarna merah menunjukkan simpul dengan kategori status kebutuhan KB terpenuhi dan simpul yang berwarna biru menunjukkan simpul dengan kategori status kebutuhan KB tidak terpenuhi.

3.4 Penanganan Data Tidak Seimbang

Implementasi penerapan SMOTE pada penelitian ini dilakukan dengan bantuan fungsi SMOTE yang terdapat pada *package DMwR* pada *software Rstudio*. Fungsi SMOTE memuat beberapa parameter, diantaranya *perc.over*, *perc.under*, dan *k*. Parameter *k* yang digunakan pada penelitian ini sebesar 5. Parameter *perc.over* yang digunakan pada penelitian ini sebesar 200% yang artinya data buatan dibangkitkan sebesar 2 kali dari banyaknya data minor asli, yaitu $2.818 \times 2 = 5.636$. Data buatan tersebut kemudian ditambahkan ke data minor asli sehingga diperoleh total data minor baru sebanyak $2.818 + 5.636 = 8.454$. Parameter *perc.under* yang digunakan pada penelitian ini sebesar 150% yang artinya data pada kelas mayor akan berkurang menjadi sebanyak 1,5 kali dari banyaknya data buatan yang dihasilkan sebelumnya, yaitu $1,5 \times 5.636 = 8.454$. Penentuan persentase tersebut ditentukan berdasarkan dari hasil beberapa kali percobaan dengan menyesuaikan jumlah data kelas minor dan mayor agar mendapatkan hasil yang seimbang. Adapun hasil dari proses SMOTE disajikan pada Tabel 4.

Tabel 4 Hasil SMOTE pada data latih

Keterangan	Tanpa SMOTE		Dengan SMOTE	
	Terpenuhi	Tidak terpenuhi	Terpenuhi	Tidak terpenuhi
Data latih utama	14601	2818	8454	8454
Data latih <i>fold</i> 1	11894	2041	6123	6123
Data latih <i>fold</i> 2	11895	2040	6120	6120
Data latih <i>fold</i> 3	11896	2040	6120	6120
Data latih <i>fold</i> 4	11897	2038	6114	6114
Data latih <i>fold</i> 5	11890	2045	6135	6135

Berdasarkan Tabel 4 terlihat bahwa jumlah data antara kategori kebutuhan KB terpenuhi dan tidak terpenuhi sudah seimbang dengan persentase data pada kelas mayor sebesar 50% dan minor sebesar 50%.

3.5 Analisis Klasifikasi dengan SMOTE (SMOTE CART)

Model klasifikasi SMOTE CART dibentuk menggunakan data latih yang sudah seimbang. Model klasifikasi SMOTE CART yang terbentuk berupa pohon klasifikasi karena peubah respon yang digunakan bertipe kategorik. Pohon klasifikasi SMOTE CART yang terbentuk pada penelitian ini merupakan pohon klasifikasi tunggal dan dibentuk menggunakan *set.seed(044)*. Pohon klasifikasi SMOTE CART juga dibentuk dengan batasan jumlah amatan minimum pada setiap simpul (*minsplit*) sebesar 2, parameter biaya kompleksitas (*CP*) minimum sebesar 0, dan kedalaman pohon maksimal (*maxdepth*) sebesar 5. Berikut merupakan performa hasil klasifikasi SMOTE CART dari proses validasi silang 5-fold:

Tabel 5 Performa model klasifikasi SMOTE CART dalam setiap *fold*

Fold	Accuracy (%)	Sensitivitas (%)	Spesifisitas (%)	Balanced Accuracy (%)
1	74,97	25,49	83,45	54,47
2	73,54	24,85	81,90	53,38
3	72,52	21,92	81,22	51,57
4	75,26	22,61	84,35	53,48
5	72,59	21,54	81,26	51,40
Rata-rata	73,78	23,28	82,44	52,86

Nilai *accuracy*, sensitivitas, dan spesifisitas dari model klasifikasi SMOTE CART untuk masing-masing *fold* dapat dilihat pada Tabel 5. Perbedaan nilai *balanced accuracy*, *accuracy*, sensitivitas, spesifisitas pada masing-masing *fold* tidak terlalu berbeda jauh, hal ini menunjukkan bahwa secara umum data dalam setiap *fold* memberikan kontribusi yang cenderung sama yaitu dengan rata-rata *balanced accuracy* sebesar 52,86%, *accuracy* sebesar 73,78%, sensitivitas sebesar 23,28%, dan spesifisitas sebesar 82,44%. Penentuan model SMOTE CART terbaik dilakukan dengan membandingkan performa model utama (model yang dibangun menggunakan data latih utama yang sudah seimbang) dengan model *fold* hasil SMOTE yang dicobakan pada data uji utama. Model *fold* yang terpilih untuk dibandingkan dengan model utama adalah model *fold* 1 karena menghasilkan nilai sensitivitas dan *balanced accuracy* tertinggi.

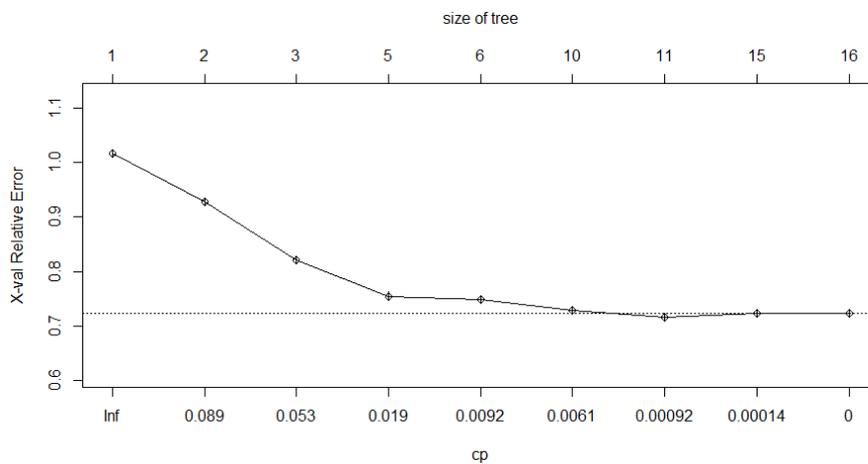
Tabel 6 Perbandingan performa model klasifikasi SMOTE CART

Model	Accuracy (%)	Sensitivitas (%)	Spesifisitas (%)	Balanced Accuracy (%)
Model utama	70,46	26,35	78,97	52,66
Model <i>fold</i> 1	72,01	23,86	81,30	52,58

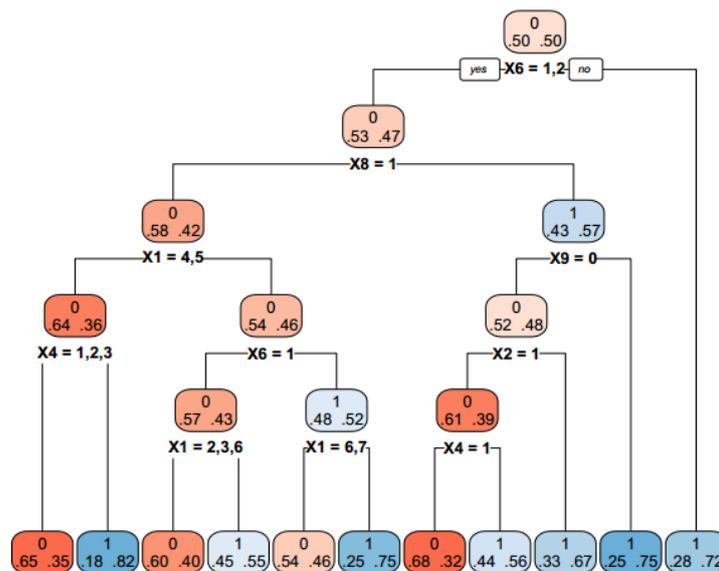
Setelah dilakukan evaluasi model terhadap data uji utama, diperoleh hasil sesuai pada Tabel 6. Model yang ditetapkan sebagai model SMOTE CART terbaik adalah model utama karena menghasilkan nilai sensitivitas dan *balanced accuracy* yang lebih tinggi dibandingkan model *fold* 1. Berdasarkan Tabel 5 dan 6, nilai sensitivitas dan *balanced accuracy* yang dihasilkan dari model SMOTE CART meningkat dibandingkan model CART, sedangkan nilai *accuracy* dan spesifisitas yang dihasilkan model SMOTE CART menurun dibandingkan model CART. Hal ini menunjukkan bahwa klasifikasi

pada data yang seimbang mampu memperbaiki ketepatan hasil klasifikasi untuk kelas minor.

Pohon klasifikasi SMOTE CART yang terbentuk dari model utama terdiri dari 16 simpul terminal yang terdiri dari 6 simpul terminal dengan label status kebutuhan KB terpenuhi dan 10 simpul terminal dengan label status kebutuhan KB tidak terpenuhi. Setelah pohon klasifikasi terbentuk, tahap selanjutnya adalah proses pemangkasan pohon klasifikasi maksimal untuk mendapatkan pohon yang lebih optimal. Proses pemangkasan dilakukan dengan menggunakan validasi silang 10-fold dan dari proses ini dihasilkan nilai kesalahan relatif. Pemangkasan dilakukan berdasarkan parameter biaya kompleksitas atau *CP*. *CP* yang menghasilkan nilai kesalahan relatif (*xerror*) paling kecil dipilih untuk memangkas pohon klasifikasi maksimal. Berdasarkan Gambar 4, nilai kesalahan relatif paling kecil diperoleh ketika *CP* bernilai 0,00092, sehingga pohon klasifikasi dipangkas dengan *CP* = 0,00092. Hasil pohon klasifikasi SMOTE CART optimal dapat dilihat pada Gambar 5.



Gambar 4 Plot *CP* SMOTE CART



Gambar 5 Pohon klasifikasi optimal SMOTE CART berdasarkan model terbaik

Pohon klasifikasi SMOTE CART yang terbentuk terdiri dari 11 simpul terminal yang melibatkan peubah banyak anak hidup (X_6), mendapati informasi KB dari media massa (X_8), usia istri (X_1), frekuensi penggunaan internet sebulan terakhir (X_9), tingkat

pendidikan istri (X_2), dan tingkat pendidikan suami (X_4). Peubah banyak anak hidup merupakan pemilah terbaik karena menghasilkan nilai *goodness of split* tertinggi dengan kriteria pemilahan banyak anak hidup 1–4 anak pada simpul anak kiri dan tidak memiliki anak atau banyak anak hidup 5 atau lebih pada simpul anak kanan. Berdasarkan Gambar 5, dari 11 simpul terminal yang dihasilkan terdapat 7 simpul dengan kategori status kebutuhan KB tidak terpenuhi dan 4 simpul dengan kategori status kebutuhan KB terpenuhi.

3.6 Perbandingan Model Klasifikasi Terbaik

Perbandingan kinerja model klasifikasi CART dan SMOTE CART dilakukan dengan membandingkan hasil *accuracy*, sensitivitas, spesifisitas, dan *balanced accuracy*, dari kedua pohon klasifikasi optimal atau pohon klasifikasi yang sudah dipangkas. Adapun hasil perbandingannya disajikan pada Tabel 7.

Tabel 7 Perbandingan hasil klasifikasi CART dan SMOTE CART optimal

Ketepatan	CART (%)	SMOTE CART (%)
<i>Accuracy</i>	83,95	68,27
Sensitivitas	1,57	34,96
Spesifisitas	99,84	74,70
<i>Balanced accuracy</i>	50,71	54,83

Berdasarkan Tabel 7, penerapan penanganan data tidak seimbang dengan SMOTE mengakibatkan persentase *accuracy* dan spesifisitas model klasifikasi SMOTE CART menurun secara berturut-turut sebesar 15,68% dan 25,14%, namun meningkatkan persentase sensitivitas dan *balanced accuracy* secara berturut-turut sebesar 33,39% dan 4,12%. Hal ini menunjukkan bahwa dengan melakukan penanganan data tidak seimbang mampu menurunkan bias pada kinerja pengklasifikasi dalam mengklasifikasikan data pada kelas kebutuhan KB terpenuhi dan mampu memperbaiki kinerja klasifikasi dalam mengklasifikasikan data pada kelas kebutuhan KB tidak terpenuhi. Penentuan model terbaik antara model CART dan SMOTE CART pada penelitian ini dilakukan berdasarkan nilai sensitivitas tertinggi dikarenakan fokus pada penelitian ini ingin melihat karakteristik wanita usia subur berstatus menikah atau memiliki pasangan yang kebutuhan KBnya tidak terpenuhi. Oleh karena itu, model terbaik yang dipilih yaitu model yang mampu mengklasifikasikan kelas kebutuhan KB tidak terpenuhi lebih baik. Model SMOTE CART dipilih sebagai model terbaik karena menghasilkan nilai sensitivitas yang lebih tinggi dibandingkan model CART. Hal ini menunjukkan bahwa model SMOTE CART mampu menghasilkan kinerja yang lebih baik dalam mengklasifikasikan wanita dengan status kebutuhan KB tidak terpenuhi. Selain itu jika dilihat dari nilai *balanced accuracy* model SMOTE CART juga lebih tinggi dibandingkan model CART yang artinya secara keseluruhan model SMOTE CART cukup baik dalam mengklasifikasikan wanita dengan kebutuhan KB terpenuhi dan tidak terpenuhi.

3.7 Interpretasi Model Klasifikasi Terbaik

Karakteristik kebutuhan KB pada wanita usia subur berstatus menikah atau memiliki pasangan berdasarkan pohon klasifikasi SMOTE CART pada Gambar 5 dapat dilihat pada tabel berikut:

Tabel 8 Karakteristik kebutuhan KB berdasarkan model SMOTE CART

Simpul Terminal	Kategori	Karakteristik
1	Terpenuhi	Memiliki 1–4 anak yang masih hidup, mendapatkan informasi KB dari media massa, berusia 30–39 tahun, suami dengan tingkat pendidikan dasar atau menengah atau atas.
2	Tidak terpenuhi	Memiliki 1–4 anak yang masih hidup, mendapatkan informasi KB dari media massa, wanita berusia 30–39 tahun, suami tidak sekolah
3	Terpenuhi	Memiliki 1–2 anak yang masih hidup, mendapatkan informasi KB dari media massa, wanita berusia 20–29 tahun atau berusia 40–44 tahun.
4	Tidak terpenuhi	Memiliki 1–2 anak yang masih hidup, mendapatkan informasi KB dari media massa, berusia 15–19 tahun atau lebih dari 44 tahun.
5	Terpenuhi	Memiliki 3–4 anak yang masih hidup, mendapatkan informasi KB dari media massa, berusia lebih dari 39 tahun.
6	Tidak terpenuhi	Memiliki 3–4 anak yang masih hidup, mendapatkan informasi KB dari media massa, berusia kurang dari 30 tahun.
7	Terpenuhi	Memiliki 1–4 anak yang masih hidup, tidak mendapatkan informasi KB dari media massa, dalam satu bulan terakhir tidak mengakses internet sama sekali, pendidikan dasar, suami dengan tingkat pendidikan dasar.
8	Tidak terpenuhi	Memiliki 1–4 anak yang masih hidup, tidak mendapatkan informasi KB dari media massa, dalam satu bulan terakhir tidak mengakses internet sama sekali, tingkat pendidikan dasar, suami tidak sekolah atau tingkat pendidikan suami menengah atau tinggi.
9	Tidak terpenuhi	Memiliki 1–4 anak yang masih hidup, tidak mendapatkan informasi KB dari media massa, dalam satu bulan terakhir tidak mengakses internet sama sekali, tingkat pendidikan menengah atau tinggi atau tidak sekolah.
10	Tidak terpenuhi	Memiliki 1–4 anak yang masih hidup, tidak mendapatkan informasi KB dari media massa, dalam sebulan terakhir mengakses internet.
11	Tidak terpenuhi	Tidak memiliki anak atau memiliki lebih dari 4 anak.

4. Simpulan dan Saran

4.1 Simpulan

Data kebutuhan KB yang digunakan dalam penelitian ini merupakan kasus data tidak seimbang. Analisis klasifikasi CART menghasilkan kinerja klasifikasi yang buruk dalam mengklasifikasikan data kelas minor yaitu kategori kebutuhan KB tidak terpenuhi. Hal ini dilihat berdasarkan nilai sensitivitas yang dihasilkan yaitu sebesar 1,57%. Setelah dilakukan penanganan data tidak seimbang menggunakan metode SMOTE, analisis klasifikasi SMOTE CART memberikan hasil yang lebih baik dalam mengklasifikasikan kelas kebutuhan KB tidak terpenuhi. Hal ini dilihat berdasarkan nilai sensitivitas yang meningkat menjadi 34,96%. Oleh karena itu, pada penelitian ini analisis klasifikasi SMOTE CART mampu memberikan kinerja yang lebih baik dibandingkan metode CART sehingga metode SMOTE CART merupakan metode terbaik dalam mengklasifikasikan kebutuhan KB tidak terpenuhi di Indonesia. Berdasarkan model SMOTE CART, secara umum karakteristik wanita usia subur bertatus menikah atau memiliki pasangan yang kebutuhan KBnya tidak terpenuhi, yaitu memiliki 1–4 anak yang masih hidup, tidak mendapatkan informasi KB dari media massa, dalam satu bulan terakhir tidak mengakses internet, berusia lebih dari 30 tahun, memiliki tingkat pendidikan dasar atau menengah, suami tidak sekolah atau memiliki pendidikan dasar atau menengah.

4.2 Saran

Penelitian ini menghasilkan ketepatan hasil klasifikasi untuk kelas kebutuhan KB tidak terpenuhi yang cukup kecil, sehingga penelitian selanjutnya dapat mencobakan beberapa parameter SMOTE dengan persentase yang lain atau dapat menerapkan metode *Ensemble* seperti *Bagging* atau *Boosting* untuk dapat meningkatkan kinerja hasil klasifikasi. Apabila setelah dicobakan persentase SMOTE yang lain atau menerapkan metode *Ensemble* namun masih menghasilkan nilai ketepatan hasil klasifikasi yang rendah, maka dapat diduga bahwa dengan metode pohon klasifikasi CART belum dapat mengidentifikasi karakteristik wanita yang kebutuhan KBnya tidak terpenuhi dengan baik, sehingga penelitian selanjutnya dapat mencobakan analisis dengan menggunakan metode yang lain.

Daftar Pustaka

- Alverina D, Chrismanto AR, Santosa RG. 2018. Perbandingan akurasi algoritma C4.5 dan CART dalam memprediksi kategori indeks prestasi mahasiswa. *Jurnal Teknologi dan Sistem Komputer*. 6(2):76–83. DOI:10.14710/jtsiskom.6.2.2018.76–83.
- Anggraeni AT, Budiantara IN. 2016. Pemodelan faktor-faktor yang memengaruhi unmet need KB di Provinsi Jawa Timur dengan pendekatan regresi nonparametric spline. *Jurnal Sains dan Seni ITS*. 5(2): 163–168.
- Aritonang NE, Rusgiyono A, Rita R. 2016. Klasifikasi status kerja pada angkatan kerja Kota Semarang tahun 2014 menggunakan metode CHAID dan CART. *Jurnal Gaussian*. 5(1):183–192.

- Asif MF, Pervaiz Z. 2019. Socio-demographic determinants of unmet need for family planning among married women in Pakistan. *BMC Public Health*. 19:1226. DOI:10.1186/s12889-019-7487-5.
- Assiroj P, Warnars HLHS, Fauzi A. 2018. Comparing CART and C5.0 algorithm performance of human development index. 2018. Di dalam: 2018 Third International Conference on Informatics and Computing (ICIC); Okt 17–18; Palembang, Indonesia. IEEE. hlm 1–5.
- [BPS] Badan Pusat Statistik. 2021. Hasil Sensus Penduduk 2020. Jakarta: BPS.
- Bekkar M, Djemaa HK, Alitouche TA. 2013. Evaluation measures for models assessment over imbalanced data sets. *Journal of Information Engineering and Applications*. 3(10): 27–38.
- Bradley SEK, Croft TN, Fishel JD, Westoff CF. 2012. Revising Unmet Need for Family Planning. Maryland (USA): ICF International.
- Breiman L, Friedman JH, Olshen RA, Stone CJ. 1993. *Classification and Regression Trees*. New York (US): Chapman and Hall.
- Chawla VN, Bowyer KW, Hall LO, Kegelmeyer WP. 2002. SMOTE: Synthetic Minority Oversampling Technique. *Journal of Artificial Intelligence Research*. 16: 321–357.
- Chen JJ, Tsai CA, Moon H, Ahn H, Young JJ, Chen CH. 2006. Decision threshold adjustment in class prediction. *SAR and QSAR in Environmental Research*. 17(3):337–352.
- Dejene H, Abera M, Tadele A. 2021. Unmet need for family planning and associated factors among married women attending antiretroviral treatment clinics in Dire Dawa City, Eastern Ethiopia. *PLoS ONE*. 16(4): e0250297. DOI:10.1371/journal.pone.0250297.
- James G, Witten D, Hastie T, Tibshirani, R. 2013. *An Introduction to Statistical Learning with Application in R*. New York (US): Springer.
- Krawczyk B. 2016. Learning from imbalanced data: open challenges and future directions. *Prog Artif Intell*. 5:221–232. DOI:10.1007/s13748-016-0094-0.
- [Puslitbang KB dan KS BKKBN] Pusat Penelitian dan Pengembangan Keluarga Berencana dan Keluarga Sejahtera Badan Kependudukan dan Keluarga Berencana. 2017. Laporan Survei Kinerja dan Akuntabilitas Program. Jakarta: BKKBN.
- Shirali GA, Noroozi MV, Malehi AS. 2018. Predicting the outcome of occupational accidents by CART and CHAID methods at a steel factory in Iran. *Journal of Public Health Research*. 7(2):74-80. DOI:10.4081/jphr.2018.1361.
- Torres RF, Carraso-Ochoa JA, Martinez-Trinidad JF. 2016. SMOTE-D a deterministic version of SMOTE. *Lecture Notes in Computer Science*. 9703:177–188. DOI:10.1007/978-3-319-39393-318.
- Widyatami AI, Natungga SG, Damayanti R, Dewi SE, Siagian TH. 2021. Determinan unmet need pada pasangan usia subur di Kawasan Indonesia Timur. *Jurnal Keluarga Berencana*. 6(1):31–41.

- Wijaya J, Soleh AM, Rizki A. 2018. Penanganan data tidak seimbang pada pemodelan Rotation Forest keberhasilan studi mahasiswa Program Magister IPB. *Xplore*. 2(2):32–40.
- Wulifan JK, Jahn A, Hien H, Ilboudo PC, Meda N, Robyn PJ, Hamadou TS, Haidara O, Allegri MD. 2017. Determinants of unmet need for family planning in rural Burkina Faso: a multilevel logistic regression analysis. *BMC Pregnancy and Childbirth*. 17: 426. DOI:10.1186/s12884-017-1614-z.