

KLASIFIKASI TINGKAT PENGANGGURAN TERBUKA DI INDONESIA DENGAN ALGORITMA CLASSIFICATION AND REGRESSION TREE (CART) DAN C4.5

Inggit Fatika¹, Kris Suryowati^{2*}, Noviana Pratiwi^{3*}, Muhammad Sholeh⁴

Jurusan Statistika, Fakultas Sains Terapan, Institut Sains & Teknologi AKPRIND Yogyakarta

Email: fatikainggit@gmail.com, suryowati@akprind.ac.id,
novianapratiwi@akprind.ac.id, muhash@akprind.ac.id

*corresponding author

Abstract

The open unemployment rate in Indonesia in February 2021 was 6.26%, this number increased by 1.32% compared to the situation in February 2020 of 4.94%. A high open unemployment rate in a country can lead to low public participation in economic growth which will decrease the level of national development. In addition, it also can cause high social costs so that various crimes occur in society. So it is necessary to know several factors that can affecting the increase of the open unemployment rate in Indonesia, one of it is by classifying. The classifications used were Classification and Regression Tree (CART) and C4.5. CART and C4.5 are one of the decision tree algorithms that can be used to perform classifications using hierarchical structures. The CART algorithm can be used for classification with large amount of data with many factors and can perform classification analysis on nominal, ordinal, and continuous data scales. The C4.5 algorithm has advantages in handling missing values, overcoming overfitting decision trees, and can be used for continuous data scale which are following the form of research data to be used. The results show that the decision tree model formed from both algorithms is the same, with one maximal depth and the level of employment is the most influential variable. The accuracy, sensitivity and specificity rate of the two models formed based on the algorithm is the same, which is 100%. So that both models are equally good to be used in classifying the open unemployment rate in Indonesia.

Keywords: Classification, CART, C4.5, and open unemployment rate

Abstrak

Tingkat pengangguran terbuka di Indonesia bulan Februari 2021 sebesar 6,26%, jumlah ini mengalami kenaikan sebesar 1,32% jika dibandingkan dengan keadaan bulan Februari 2020 yaitu sebesar 4,94%. Tingkat pengangguran terbuka yang tinggi pada suatu negara dapat menyebabkan rendahnya partisipasi masyarakat dalam pertumbuhan perekonomian yang akan berakibat pada penurunan tingkat pembangunan nasional. Selain itu, tingginya tingkat pengangguran terbuka juga dapat menyebabkan tingginya biaya sosial sehingga terjadi berbagai tindak kriminalitas dalam masyarakat. Agar permasalahan ini dapat diatasi perlunya mengetahui berbagai faktor yang dapat mempengaruhi peningkatan tingkat pengangguran terbuka pada berbagai Provinsi di Indonesia salah satunya dengan melakukan klasifikasi. Pada penelitian ini klasifikasi yang digunakan adalah CART (*Classification and Regression Tree*) dan C4.5. CART dan C4.5 merupakan salah satu algoritma *decision tree* yang dapat digunakan untuk melakukan klasifikasi menggunakan struktur hirarki. Algoritma CART dapat digunakan untuk klasifikasi dengan jumlah data yang cukup besar dengan banyak faktor serta dapat melakukan analisis klasifikasi pada peubah respon baik nominal, ordinal, maupun kontinu. Algoritma C4.5 memiliki kelebihan dalam menangani *missing value*, mengatasi pohon keputusan yang *overfitting*, serta dapat digunakan untuk jenis data kontinu yang mana sesuai dengan bentuk data penelitian yang akan digunakan. Hasil dari penelitian ini menunjukkan bahwa model pohon keputusan yang terbentuk dari kedua algoritma adalah sama, yaitu dengan kedalaman satu dan variabel tingkat kesempatan kerja

merupakan variabel yang paling berpengaruh. Tingkat akurasi, *sensitivity*, dan *Specificity* dari kedua model yang terbentuk berdasarkan algoritma tersebut adalah sama, yaitu sebesar 100%, sehingga kedua model sama baiknya untuk digunakan dalam melakukan klasifikasi tingkat pengangguran terbuka di Indonesia.

Kata Kunci: Klasifikasi, CART, C4.5, dan tingkat pengangguran terbuka

1. Pendahuluan

Tingkat pengangguran terbuka merupakan persentase jumlah pengangguran terhadap jumlah angkatan kerja. Survei Angkatan Kerja Nasional (SAKERNAS) BPS 2021 mencatat bahwa tingkat pengangguran terbuka di Indonesia pada bulan Februari 2021 sebesar 6,26%. Jumlah ini mengalami kenaikan sebesar 1,32% jika dibandingkan dengan keadaan pada bulan Februari 2020 yaitu sebesar 4,94%. Tingkat pengangguran terbuka yang tinggi pada suatu negara dapat menyebabkan rendahnya partisipasi masyarakat dalam pertumbuhan perekonomian yang akan berakibat juga pada penurunan tingkat pembangunan nasional. Selain itu tingkat pengangguran terbuka yang tinggi juga dapat mengakibatkan tingginya biaya sosial sehingga terjadinya berbagai tindak kriminalitas dalam masyarakat.

Berdasarkan dampak dari tingginya tingkat pengangguran terbuka di Indonesia maka perlunya melakukan klasifikasi untuk mengetahui secara terperinci Provinsi mana yang memiliki kategori tingkat pengangguran terbuka yang tinggi maupun rendah serta faktor apa saja yang dapat mempengaruhinya. Klasifikasi merupakan salah satu teknik data mining untuk memetakan data kedalam kedalam kelompok atau kategori tertentu [6]. Adapun teknik klasifikasi yang dapat digunakan antara lain *decision tree*, *neural network*, *naïve Bayes classifier*, dan *support vector machine*.

Metode yang digunakan untuk melakukan klasifikasi pada penelitian ini adalah *decision tree*. Menurut [5] metode *decision tree* adalah salah satu model prediksi terhadap suatu keputusan menggunakan struktur hirarki atau pohon. Terdapat 4 algoritma yang biasa digunakan dalam membangun pohon keputusan, antara lain CART, ID3, C.45 dan CHAID. Pada penelitian ini digunakan algoritma CART dan C4.5. Penggunaan algoritma *classification and regression tree* (CART) dan C4.5 memiliki kelebihan yang sesuai dengan data penelitian yang digunakan antara lain pada algoritma CART dapat digunakan untuk klasifikasi jumlah data yang cukup besar, dengan banyak faktor serta dapat dikembangkan untuk melakukan analisis klasifikasi pada peubah respon baik yang nominal, ordinal, maupun kontinu [3]. Menurut [6] algoritma C4.5 memiliki kelebihan dalam menangani *missing value*, mengatasi pohon keputusan yang *overfitting*, serta dapat digunakan untuk jenis data kontinu yang mana sesuai dengan bentuk data penelitian yang akan digunakan.

2. Metode

Penelitian ini menggunakan jenis desain penelitian *hypothesis-testing research studies* yang mana jenis desain penelitian ini menguji hubungan antar variabel. Desain penelitian ini tidak fleksibel dan fokus memperhatikan tujuan penelitian, analisis data, dan kesimpulan. Metode pengumpulan data yang dilakukan pada penelitian ini adalah menggunakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Indonesia (<https://bps.go.id>) pada tahun 2021. Data yang digunakan merupakan data variabel-variabel pada tahun 2021.

A. Variabel

Variabel yang digunakan dalam penelitian ini antara lain tingkat pengangguran terbuka (TPT), persentase penduduk miskin (PPM), besaran upah minimum provinsi (UMP), angka partisipasi sekolah (APS), indeks pembangunan manusia (IPM), laju pertumbuhan PDRB per kapita (PDRB), tingkat partisipasi angkatan kerja (TPAK), rata-rata lama sekolah (RLS), dan tingkat kesempatan kerja (TKK).

B. Metode Analisis

Tahapan analisis data pada penelitian ini, yaitu:

1. Mengumpulkan data faktor pengaruh tingkat pengangguran terbuka di Indonesia.
2. Melakukan analisis deskriptif.
3. Melakukan data *mining*.
4. Melakukan klasifikasi dengan algoritma *classification and regression tree* (CART):

a) Pemilihan pemilah

Pemilihan pemilah bertujuan untuk memperoleh pemilah yang dapat menghasilkan simpul dengan tingkat kehomogenan nilai peubah respon tertinggi. Aturan pemilihan pemilah menggunakan kriteria *goodness of split* dengan fungsi indeks Gini menggunakan persamaan berikut:

$$i(t) = 1 - \sum_{j=1}^n p^2(j|t) \quad (1)$$

$i(t)$: indeks Gini dari simpul t

$p(j|t)$: proporsi kelas j pada simpul t dimana $j = 1, 2, 3, \dots, n$ dengan

$p(j|t) = \frac{n(j|t)}{n(t)}$ dimana $n(j|t)$ adalah banyaknya pengamatan kelas j pada simpul t

$n(t)$: banyak pengamatan pada simpul t

$$\phi(s, t) = \Delta i(s, t) = i(t) - p_L i(t_L) - p_R i(t_R) \quad (2)$$

Keterangan:

$\Delta i(s, t)$: *goodness of split*

$i(t)$: indeks Gini simpul t

P_L : proporsi pengamatan dari simpul t menuju simpul kiri

P_R : proporsi pengamatan dari simpul t menuju simpul kanan

$i(t_L)$: indeks Gini pada simpul anak kiri

$i(t_R)$: indeks Gini pada simpul anak kanan

b) Penentuan simpul terminal

Menurut [1] pohon akan berhenti dikembangkan apabila pada simpul terdapat pengamatan berjumlah $n \leq 5$. Proses pembentukan pohon juga berhenti apabila telah mencapai batasan level yang telah ditentukan atau tingkat kedalaman (*depth*) pohon mencapai maksimal.

c) Penandaan label kelas

Penentuan label kelas merupakan proses identifikasi tiap node pada kelas tertentu. Penentuan label kelas pada simpul terminal dilakukan berdasarkan aturan jumlah terbanyak, yaitu:

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

Keterangan:

$P(j|t)$: proporsi kelas j pada simpul t

$N(j|t)$: jumlah pengamatan kelas j pada simpul t

$N(t)$: jumlah pengamatan pada simpul t

5. Melakukan klasifikasi dengan algoritma C4.5

a) Perhitungan *entropy*

Entropy adalah suatu parameter untuk mengukur keberagaman dalam himpunan data. Secara matematis *entropy* dihitung dengan rumus sebagai berikut [6] dan [3]:

$$Entropy(C, A) = \sum_{i=1}^n -\pi_i \times \log_2 \pi_i \quad (4)$$

Keterangan:

C : himpunan kasus

A : atribut

n : jumlah partisi atribut C

π : proporsi dari j_i (kelas ke- i) terhadap C

b) Perhitungan *information Gain*

Information Gain merupakan perubahan *entropy* setelah dilakukan pembagian berdasarkan sebuah atribut kedalam subset yang lebih kecil. Nilai *information Gain* digunakan untuk

mengukur efektivitas suatu atribut dengan melakukan klasifikasi. Persamaan untuk menghitung nilai *information Gain* adalah sebagai berikut:

$$Gain(C, A) = Entropy(C) - \sum_{i=1}^n \left| \frac{C_i}{C} \right| \times Entropy(S_i) \quad (5)$$

Keterangan:

- C : himpunan kasus
 A : atribut
 N : jumlah partisi atribut A
 $|C_i|$: Jumlah kasus pada partisi ke-i
 $|C|$: jumlah kasus dalam C

c) Perhitungan *split info*

Split info merupakan *entropy* dari seluruh distribusi probabilitas subset atribut setelah dilakukan partisi. Rumus *split info* adalah sebagai berikut (Romli & Zy, 2020):

$$Split\ info(C, A) = - \sum_{i=1}^i \frac{C_i}{C} \log_2 \frac{C_i}{C} \quad (6)$$

Keterangan:

- C : himpunan kasus
 A : atribut
 C_i : jumlah sampel untuk atribut i

d) *Gain ratio*

Gain ratio merupakan normalisasi *information gain* yang memperhitungkan nilai *entropy* dari distribusi probabilitas subset atribut setelah dilakukan partisi (Romli & Zy, 2020). Rumus *Gain ratio* adalah sebagai berikut:

$$Gain\ ratio(C, A) = \frac{Information\ gain(C,A)}{Split\ info(C,A)} \quad (7)$$

Keterangan:

- C : himpunan kasus
 A : atribut
 $Information\ gain(C, A)$: *information gain* pada atribut A
 $Split\ info(C, A)$: *split info* pada atribut A

6. Menghitung tingkat kesalahan klasifikasi model

Ukuran ketepatan klasifikasi merupakan pengukuran yang dilakukan untuk mengetahui ketepatan klasifikasi data. Pengukuran ini dapat dilakukan dengan perhitungan *sensitivity*, *Specificity*, dan akurasi, yaitu:

- 1) *Sensitivity* digunakan untuk mengukur tingkat ketepatan klasifikasi kelas I,
- 2) *Specificity* digunakan untuk mengukur ketepatan klasifikasi sampel kelas j.
- 3) Akurasi digunakan untuk mengukur tingkat ketepatan klasifikasi data.

Tabel 1. Ketepatan klasifikasi

Aktual	Prediksi		Total
	0	1	
0	n_{11}	n_{12}	N_1
1	n_{21}	n_{22}	N_2
Total	N_1	N_2	N

Berdasarkan tabel tersebut dapat dilakukan perhitungan nilai *sensitivity*, *Specificity*, akurasi, dan tingkat kesalahan klasifikasi dengan rumus sebagai berikut:

$$Sensitivity = \frac{n_{11}}{N_1} \quad (8)$$

$$Specificity = \frac{n_{22}}{N_2} \quad (9)$$

$$Akurasi = \frac{n_{11} + n_{22}}{N} \quad (10)$$

Keterangan:

- n_{11} : jumlah observasi dari kelas 1 yang tepat diprediksi sebagai kelas 1
- n_{12} : jumlah observasi dari kelas 1 yang diprediksi sebagai kelas 2
- n_{21} : jumlah observasi dari kelas 2 yang diprediksi sebagai kelas 1
- n_{22} : jumlah observasi dari kelas 2 yang tepat diprediksi sebagai kelas 2
- N_1 : jumlah observasi kelas 1
- N_2 : jumlah observasi kelas

C. Metode Analisis

1) Algoritma *classification and regression tree* (CART)

Classification and regression tree (CART) merupakan salah satu algoritma dari salah satu teknik eksplorasi data yaitu teknik pohon keputusan (Andarama, dkk, 2020). Algoritma klasifikasi CART terdiri dari dua metode yaitu metode pohon regresi dan pohon klasifikasi. Pohon regresi akan terbentuk apabila variabel dependen yang digunakan bersifat kontinu atau numerik dan pohon klasifikasi akan terbentuk apabila variabel dependen yang digunakan bersifat kategorik (Prabawati, dkk, 2019). Igoritma CART menggunakan *Binary Recursive Partitioning* (BRP). Algoritma BRP merupakan proses pembagian *dataset* menjadi dua kelompok secara berulang kali. Struktur pohon yang dihasilkan algoritma ini sangat besar, namun akan kembali optimal berdasarkan hasil *classification error* yang diperoleh.

2) Algoritma C4.5

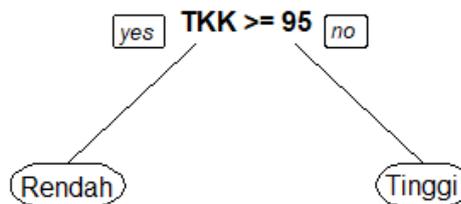
Algoritma C4.5 adalah pengembangan dari algoritma ID3 yang digunakan untuk membentuk *decision tree* (pohon keputusan) yang diciptakan oleh J. Ross Quinlan pada tahun 1970 hingga 1980. Algoritma C4.5 membentuk pohon keputusan berdasarkan pada pemilihan atribut yang memiliki nilai *gain* (*gain ratio*) tertinggi berdasarkan *entropy* atribut sebagai proros klasifikasi (Sukma, dkk, 2019).

3. Hasil dan Pembahasan

Klasifikasi tingkat pengangguran terbuka di Indonesia dilakukan menggunakan dua lagoritma, yaitu CART dan C4.5 seperti berikut:

1. Algoritma *classification and regression tree* (CART)

Berikut adalah pohon keputusan yang terbentuk menggunakan algoritma CART:



Gambar 1. Pohon keputusan CART

Hasil dari model pohon keputusan algoritma CART dengan berbagai kedalaman, yaitu 1, 3, dan 5 dengan pembagian data 80%:20% adalah sama sehingga interpretasi pohon keputusan di atas adalah sebagai berikut:

- 1) Terdapat satu variabel yang berpengaruh dalam klasifikasi tingkat pengangguran terbuka di Indonesia, yaitu tingkat kesempatan kerja. Ketujuh variabel lainnya kurang berpengaruh, yaitu persentase penduduk miskin, besaran upah minimum Provinsi, persentase angka partisipasi sekolah (rata-rata usia 19-24 tahun), persentase indeks pembangunan manusia, laju pertumbuhan PDRB per kapita, tingkat partisipasi angkatan kerja, dan rata-rata lama sekolah.
- 2) Variabel tingkat kesempatan kerja menjadi akar atau merupakan variabel paling penting dalam pembentukan pohon.
- 3) Terdapat 2 *leaf nodes* yang mengidentifikasi:
 - a) Jika tingkat kesempatan kerja ≥ 95 maka termasuk kedalam Provinsi dengan tingkat pengangguran terbuka yang rendah. Terdapat 17 Provinsi yang termasuk kedalam tingkat pengangguran terbuka yang tinggi, antara lain:

Riau, Jambi, Bengkulu, Lampung, Kep. Bangka Belitung, DI Yogyakarta, Jawa Timur, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku Utara, dan Papua.

- b) Jika tingkat kesempatan kerja < 95 maka termasuk kedalam Provinsi dengan tingkat pengangguran terbuka yang tinggi. Terdapat 10 Provinsi yang termasuk kedalam tingkat pengangguran terbuka yang tinggi, antara lain: Aceh, Sumatera Barat, Kep. Riau, DKI Jakarta, Banten, Kalimantan Barat, Sulawesi Utara, Sulawesi Selatan, Maluku, dan Papua Barat.

Tabel 2. Hasil prediksi data *testing* algoritma CART data 34 Provinsi

No	Provinsi	TPT	Prediksi
2	Sumatera Utara	Tinggi	Tinggi
6	Sumatera Selatan	Rendah	Rendah
12	Jawa Barat	Tinggi	Tinggi
13	Jawa Tengah	Tinggi	Tinggi
17	Bali	Rendah	Rendah
23	Kalimantan Timur	Tinggi	Tinggi
26	Sulawesi Tengah	Rendah	Rendah

Hasil prediksi ini menunjukkan bahwa terdapat tiga Provinsi yang diprediksi memiliki tingkat pengangguran terbuka rendah antara lain Sumatera Selatan, Bali, dan Sulawesi Tengah. Empat Provinsi yang diprediksi memiliki tingkat pengangguran terbuka tinggi antara lain Sumatera Utara, Jawa Barat, Jawa Tengah, dan Kalimantan Timur.

Tabel 3. Kesalahan klasifikasi algoritma CART data 34 Provinsi

Aktual	Prediksi		Total
	Rendah	Tinggi	
Rendah	3	0	3
Tinggi	0	4	4
Total	3	4	7

$$\text{Sensitivity} = \frac{3}{3} \times 100\% = 100\%$$

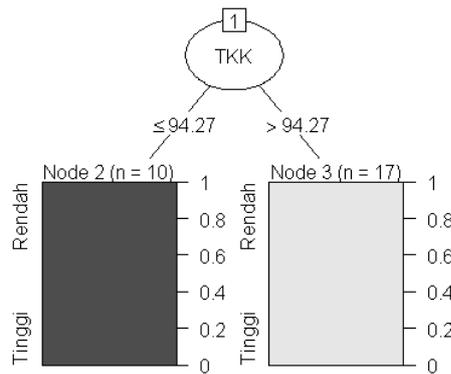
$$\text{Specivity} = \frac{4}{4} \times 100\% = 100\%$$

$$\text{Akurasi} = \frac{3+4}{7} \times 100\% = 100\%$$

Ukuran tingkat ketepatan klasifikasi model pohon keputusan sebesar 100% menunjukkan bahwa model dapat memprediksi klasifikasi dengan akurasi tinggi. Nilai *sensitivity* yaitu nilai yang digunakan untuk mengukur klasifikasi pada kelas tingkat pengangguran terbuka yang rendah adalah sebesar 100%. Nilai *specificity* yaitu nilai yang digunakan untuk mengukur klasifikasi pada kelas tingkat pengangguran terbuka yang tinggi adalah sebesar 100%.

2. Algoritma C4.5

Berikut adalah pohon keputusan yang terbentuk menggunakan algoritma C4.5:



Gambar 2. Pohon keputusan C4.5

Hasil dari model pohon keputusan algoritma C4.5 dengan pembagian data 80%:20% adalah sebagai berikut:

- 1) Terdapat satu variabel yang berpengaruh dalam klasifikasi tingkat pengangguran terbuka di Indonesia, yaitu tingkat kesempatan kerja. Ketujuh variabel lainnya kurang berpengaruh, yaitu persentase penduduk miskin, besaran upah minimum Provinsi, persentase angka partisipasi sekolah (rata-rata usia 19-24 tahun), persentase indeks pembangunan manusia, laju pertumbuhan PDRB per kapita, tingkat partisipasi angkatan kerja, dan rata-rata lama sekolah.
- 2) Variabel tingkat kesempatan kerja menjadi akar atau merupakan variabel paling penting dalam pembentukan pohon.
- 3) Terdapat 2 *leaf nodes* yang mengidentifikasi:
 - a) Jika tingkat kesempatan kerja ≥ 95 maka termasuk kedalam Provinsi dengan tingkat pengangguran terbuka yang rendah. Terdapat 17 Provinsi yang termasuk kedalam tingkat pengangguran terbuka yang tinggi, antara lain: Riau, Jambi, Bengkulu, Lampung, Kep. Bangka Belitung, DI Yogyakarta, Jawa Timur, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku Utara, dan Papua.
 - b) Jika tingkat kesempatan kerja < 95 maka termasuk kedalam Provinsi dengan tingkat pengangguran terbuka yang tinggi. Terdapat 10 Provinsi yang termasuk kedalam tingkat pengangguran terbuka yang tinggi, antara lain: Aceh, Sumatera Barat, Kep. Riau, DKI Jakarta, Banten, Kalimantan Barat, Sulawesi Utara, Sulawesi Selatan, Maluku, dan Papua Barat.

Tabel 4. Hasil prediksi data *testing* algoritma C4.5 data 34 Provinsi

No	Provinsi	TPT	Prediksi
2	Sumatera Utara	Tinggi	Tinggi
6	Sumatera Selatan	Rendah	Rendah
12	Jawa Barat	Tinggi	Tinggi
13	Jawa Tengah	Tinggi	Tinggi
17	Bali	Rendah	Rendah
23	Kalimantan Timur	Tinggi	Tinggi
26	Sulawesi Tengah	Rendah	Rendah

Hasil prediksi ini menunjukkan bahwa terdapat tiga Provinsi yang diprediksi memiliki tingkat pengangguran terbuka rendah antara lain Sumatera Selatan, Bali, dan Sulawesi Tengah. Empat Provinsi yang diprediksi memiliki tingkat pengangguran terbuka tinggi antara lain Sumatera Utara, Jawa Barat, Jawa Tengah, dan Kalimantan Timur.

Tabel 5. Kesalahan klasifikasi algoritma C4.5 data 34 Provinsi

Aktual	Prediksi		Total
	Rendah	Tinggi	
Rendah	3	0	3
Tinggi	0	4	4
Total	3	4	7

$$Sensitivity = \frac{3}{3} \times 100\% = 100\%$$

$$Specivity = \frac{4}{4} \times 100\% = 100\%$$

$$Akurasi = \frac{3+4}{7} \times 100\% = 100\%$$

Ukuran tingkat ketepatan klasifikasi model pohon keputusan sebesar 100% menunjukkan bahwa model dapat memprediksi klasifikasi dengan akurasi tinggi. Nilai *sensitivity* yaitu nilai yang digunakan untuk mengukur klasifikasi pada kelas tingkat pengangguran terbuka yang rendah adalah sebesar 100%. Nilai *specificity* yaitu nilai yang digunakan untuk mengukur klasifikasi pada kelas tingkat pengangguran terbuka yang tinggi adalah sebesar 100%.

3. Pemilihan Model Terbaik

Pemilihan model terbaik dari kedua algoritma yang digunakan berdasarkan pada tingkat akurasi yang dihasilkan. Hasil akurasi ini dapat dilihat pada Tabel 3 kesalahan klasifikasi algoritma CART dan Tabel 5 kesalahan klasifikasi algoritma C4.5. Berdasarkan hasil tersebut diketahui bahwa kedua algoritma memiliki tingkat akurasi yang sama yaitu sebesar 100% dan tidak terdapat perbedaan dalam hasil klasifikasinya. Sehingga kedua model dari algoritma CART maupun C4.5 dapat digunakan untuk melakukan klasifikasi tingkat pengangguran terbuka di Indonesia.

4. Kesimpulan

Berdasarkan hasil analisis yang dilakukan diperoleh model terbaik untuk melakukan klasifikasi tingkat pengangguran terbuka di Indonesia yaitu dapat menggunakan salah satu dari algoritma CART dan C4.5 karena kedua model tersebut menunjukkan hasil yang sama baik, yaitu 1 kedalaman dengan variabel tingkat kesempatan kerja menjadi akar atau merupakan variabel paling penting dalam pembentukan pohon. Bentuk klasifikasinya yaitu jika tingkat kesempatan kerja $< 94,27$ maka termasuk kedalam Provinsi dengan tingkat pengangguran terbuka yang tinggi dan sebaliknya jika $\geq 94,27$ maka termasuk kedalam Provinsi dengan tingkat pengangguran terbuka yang rendah. Tingkat akurasi, *sensitivity*, *specificity* model algoritma CART dan C4.5 sebesar 100%.

Daftar Pustaka

- [1] Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1993). *Classification and Regression Tree*. New York: Chapman and Hall.
- [2] Jatmiko, Y. A., Padmadisastra, S., & Chadidijah, A. (2019). Analisis Perbandingan Kinerja CART Konvensional, Bagging dan Random Forest pada Klasifikasi Objek: Hasil dari Dua Simulasi. *Media Statistika*, 1-12.
- [3] Mujahidin, A., & Pribadi, D. (2017). Penerapan C4.5 untuk Diagnosa Penyakit Pneumonia pada Anak Balita Berbasis Mobile. *Jurnal Swabumi*, 155-161.
- [4] Romli, I., & Zy, A. (2020, September). Penentuan Jadwal Overtime dengan Klasifikasi Data Karyawan menggunakan Algoritma C4.5. *Jurnal Sains Komputer & Informatika (J-*

SAKTI, IV(2), 694-702. Retrieved Maret 28, 2022, from <https://tunasbangsa.ac.id/ejurnal/index.php/jsakti>

- [5] Sartika, D., & Sensuse, D. I. (2017). Perbandingan Algoritma Klasifikasi Naive bayes, Nearest Neighbour, dan decision Tree pada studi kasus Pengambilan Keputusan Pemilihan Pola Pakaian. *Jurnal Teknik Informatika dan Sistem Informasi*, 151-161.
- [6] Setio, P., Saputro, D., & Winarno, B. (2020). Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika. III*, pp. 64-71. Surakarta: Universitas Sebelas Maret. Retrieved from <https://journal.unnes.ac.id/sju/index.php/prisma/article/view/37650>