

## KLASIFIKASI PROVINSI DI INDONESIA DENGAN METODE CART DAN KNN BERDASARKAN PDRB PER KAPITA

Petrus Chanelius Laumay<sup>1</sup>, Yudi Setyawan<sup>2\*</sup>, Noviana Pratiwi<sup>3</sup>

<sup>1,2,3</sup> Universitas AKPRIND Indonesia, \*Penulis Koresponden

e-mail: <sup>1</sup>petruschanel8632@gmail.com, <sup>2</sup>setyawan@akprind.ac.id, <sup>3</sup>novianapратиwi@akprind.ac.id

### ABSTRACT

Economic growth can be measured using Gross Regional Domestic Product (GRDP) per capita by emphasizing the region's ability to increase GRDP beyond the population growth rate. This research was conducted on the classification of Indonesian Provinces based on Gross Regional Domestic Product per capita per year at constant prices using the Classification and Regression Trees (CART) and K-Nearest Neighbors (KNN) methods. The variables used include the open unemployment rate (X1), the provincial minimum wage (X2), the percentage of the poor population (X3), the average length of schooling (X4), and labor force participation (X5). Data was taken from the Indonesia Central Statistics Agency published in 2023. The CART algorithm is a data classification method by categorizing data based on the largest proportion of terminal nodes. The KNN algorithm is an algorithm that classifies data based on its k nearest neighbors. The results of this study indicate that the prediction results of GRDP per capita per year at constant prices using the CART and KNN algorithms obtained good accuracy values. It is also known that the CART algorithm is the best model in classifying GRDP per capita per year at constant prices because it has the highest average accuracy value, namely 90,475%.

**Keywords:** CART, Gross Regional Domestic Product per capita, KNN

### INTISARI

Pertumbuhan ekonomi dapat diukur menggunakan Produk Domestik Regional Bruto (PDRB) per kapita dengan menekankan pada kemampuan daerah untuk meningkatkan PDRB melebihi tingkat pertumbuhan penduduk. Dalam penelitian ini dilakukan pengklasifikasian provinsi di Indonesia berdasar Produk Domestik Regional Bruto per kapita per tahun atas harga konstan menggunakan metode *Classification and Regression Trees* (CART) dan *K-Nearest Neighbors* (KNN). Variabel yang digunakan meliputi tingkat pengangguran terbuka (X1), upah minimum provinsi (X2), persentase penduduk miskin (X3), rata-rata lama sekolah (X4), dan partisipasi Angkatan kerja (X5). Data diambil dari publikasi Badan Pusat Statistik tahun 2023. Algoritma CART adalah metode klasifikasi data dengan mengkategorikan data berdasarkan proporsi terbesar pada simpul terminal. Algoritma KNN adalah algoritma klasifikasi data berdasar k tetangga terdekatnya (*nearest neighbors*). Dari hasil penelitian ini menunjukkan bahwa hasil prediksi PDRB per kapita per tahun atas harga konstan menggunakan algoritma CART dan KNN memperoleh nilai akurasi yang baik. Diketahui juga bahwa algoritma CART merupakan model terbaik dalam mengklasifikasikan PDRB per kapita per tahun atas harga konstan karena memiliki nilai rata-rata akurasi paling tinggi yaitu sebesar 90,475%.

**Kata Kunci:** CART, KNN, PDRB per kapita

### 1. PENDAHULUAN

Indonesia merupakan negara berkembang yang terdiri dari banyak pulau dan wilayah dengan karakteristik ekonomi yang beragam. Informasi yang mendalam tentang Produk Domestik Regional Bruto (PDRB) di tingkat regional sangat penting untuk perencanaan ekonomi yang efektif, pengambilan keputusan, dan alokasi sumber daya. Keadaan ekonomi suatu daerah dapat terlihat dari PDRB daerah tersebut. Pertumbuhan ekonomi memang biasanya diukur dengan menggunakan PDRB, tetapi indikator ini tidak selalu tepat karena tidak menggambarkan pertumbuhan yang sebenarnya. Indikator lain yang dapat digunakan adalah PDRB per kapita per tahun di mana indikator ini lebih komprehensif dalam mengukur pertumbuhan ekonomi karena lebih menekankan pada kemampuan daerah untuk meningkatkan PDRB melebihi tingkat pertumbuhan penduduk (Hartini, 2017).

Berdasarkan data Badan Pusat Statistik, PDB per kapita di Indonesia dari tahun 2013 sampai 2022 mengalami kenaikan (BPS, 2023). Namun, jika dilihat nilai PDRB per kapita (atas harga konstan) provinsi di tahun 2022 dapat diketahui bahwa terdapat perbedaan nilai PDRB per kapita yang cukup signifikan dari masing-masing provinsi. PDRB per kapita per tahun dibagi menjadi dua yaitu atas harga berlaku dan atas harga konstan. Pada penelitian ini digunakan PDRB per kapita per tahun atas harga konstan. PDRB per kapita per tahun dapat dikategorikan tinggi atau rendah. PDRB per kapita per tahun dikategorikan tinggi apabila berada di atas atau sama dengan nilai rata-rata dari keseluruhan PDRB per kapita per tahun di Indonesia. Sedangkan PDRB per kapita per tahun dikategorikan rendah apabila berada di bawah nilai rata-rata dari keseluruhan PDRB per kapita per tahun di

Indonesia.

Klasifikasi merupakan salah satu teknik pada *data mining* yang memetakan data ke dalam kelompok atau kategori yang telah ditentukan (Setio, Saputro, dan Winarno, 2020). Proses klasifikasi tingkat pertumbuhan PDRB per kapita per tahun meliputi proses pengumpulan, analisis dan evaluasi data tingkat pertumbuhan PDRB per kapita per tahun beserta faktor-faktor yang mempengaruhinya. Pengklasifikasian PDRB per kapita per tahun pada penelitian ini dilakukan dengan menggunakan metode *Classification and Regression Trees* (CART) dan *K-nearest Neighbors* (KNN). Metode CART termasuk metode statistika nonparametrik untuk menggambarkan hubungan antara variabel respons dengan satu atau lebih variabel prediktor (Sumartini dan Purnami, 2015). Pengklasifikasian dengan metode CART dilakukan menggunakan algoritma penyekatan rekursif secara biner (*binary recursive partitioning*). Dalam pengklasifikasian ini dilihat kategori yang memiliki proporsi terbesar pada simpul terminal. Sedangkan algoritma KNN adalah algoritma yang berfungsi untuk melakukan klasifikasi suatu data berdasarkan data pembelajaran (*train data sets*), yang diambil dari  $k$  tetangga terdekatnya (*nearest neighbors*). Algoritma ini bekerja dengan berdasarkan pada jarak terpendek dari sampel uji ke sampel latih untuk menentukan tetangga terdekatnya. Setelah sejumlah  $k$  tetangga terdekat dikumpulkan, mayoritas kelas dari tetangga terdekat itu dijadikan prediksi dari sampel uji (Argina, 2020). Hasil klasifikasi dengan kedua metode dibandingkan untuk mengetahui tingkat akurasi yang lebih baik.

## 2. METODE

Penelitian ini bersifat *descriptive and diagnostic research studies* di mana dilakukan penggambaran karakteristik data dan analisis komparatif antara satu metode dengan metode lain. Pada penelitian ini dibandingkan akurasi antara hasil klasifikasi metode CART dan KNN dalam mengklasifikasikan provinsi-provinsi di Indonesia berdasarkan PDRB per kapita per tahun atas harga konstan. Data yang digunakan merupakan data sekunder yang diperoleh dari web BPS (BPS, 2023), yakni PDRB per kapita per tahun per provinsi serta variabel terkait lainnya. Variabel penelitian meliputi PDRB per kapita (Y), tingkat pengangguran terbuka (X1), upah minimum provinsi (X2), persentase penduduk miskin (X3), rata-rata lama sekolah (X4), dan partisipasi Angkatan kerja (X5).

Metode analisis yang digunakan dalam penelitian ini terdiri dari statistika deskriptif dan statistika inferensial. Pengolahan baik analisis deskriptif maupun inferensial menggunakan bantuan *software* Microsoft Excel dan R Studio. Statistika deskriptif dalam penelitian ini bertujuan untuk menggambarkan kondisi PDRB per kapita (Y), tingkat pengangguran terbuka (X1), upah minimum provinsi (X2), persentase penduduk miskin (X3), rata-rata lama sekolah (X4), dan partisipasi angkatan kerja (X5) pada 34 Provinsi di Indonesia tahun 2022.

Selanjutnya dalam statistika inferensial digunakan beberapa tahapan *Knowledge Discovery in Database* (KDD) dengan data mining berupa metode CART dan KNN yang bertujuan untuk mengklasifikasi dan memprediksi PDRB per kapita provinsi-provinsi di Indonesia. Adapun proses KDD yang dilakukan pada penelitian ini adalah: (1) *Feature selection*, yaitu langkah pemilihan variabel yang akan digunakan; (2) Transformasi data atau *values*, yaitu langkah mengkategorisasikan variabel rasio; (3) Menentukan *seed* untuk melakukan pengacakan; (4) Menentukan Proporsi Data *Training* dan Data *Testing*. Penentuan data *training* dan *testing* pada penelitian ini dilakukan secara acak menggunakan *software* R-Studio. Proporsi data *training* dan data *testing* yang digunakan adalah 80%:20%, 70%:30%, dan 60%:40%. Proses berikutnya adalah melakukan klasifikasi dengan algoritma CART dan algoritma KNN, mengukur tingkat akurasi masing-masing klasifikasi serta membandingkan hasilnya. Langkah terakhir adalah melakukan *profiling* hasil klasifikasi terbaik untuk mengetahui karakteristik masing-masing kelas sehingga dapat diberikan rekomendasi yang sesuai.

### 2.1. Algoritma CART dan KNN

#### a. *Classification and Regression Trees* (CART)

CART merupakan salah satu metode atau algoritma dari teknik eksplorasi data yaitu teknik pohon keputusan. CART terbilang sederhana namun merupakan metode yang kuat. CART bertujuan untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu klasifikasi, selain itu CART digunakan untuk menggambarkan hubungan antara variabel respons dengan satu atau lebih variabel prediktor. Algoritma ini diperkenalkan oleh Leo Breiman, Jerom H, Friedman, Richard A. Olshen dan Charles J. Stone (Breiman dkk, 1987). Menurut (Prabawati dkk, 2019), algoritma CART memiliki kelebihan sebagai berikut:

- 1) Sangat baik digunakan untuk data yang berukuran besar.
- 2) Hasilnya berupa pohon keputusan sehingga sangat mudah untuk di interpretasikan.
- 3) Lebih akurat dan perhitungan yang mudah.
- 4) Mampu mengatasi adanya *missing value*.
- 5) Tidak berpengaruh terhadap *outlier*, *kolineritas*, dan *heteroskedastisitas* pada variabel-variabel independennya.
- 6) Dapat digunakan untuk berbagai tipe data baik kontinu maupun kategorik.
- 7) Tidak memiliki asumsi dan memiliki komputasi yang cepat.

b. *K-Nearest Neighbors*

Algoritma KNN termasuk dalam *supervised learning*, di mana hasil *query instance* yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam KNN. Algoritma ini bekerja dengan berdasarkan pada jarak terdekat dari sampel objek yang akan diklasifikasikan ke semua data yang lain untuk menentukan KNN-nya. Setelah mengumpulkan KNN, kemudian diambil mayoritas dari KNN untuk dijadikan prediksi dari sampel uji atau sampel yang ingin diklasifikasikan. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak Euclid (Cahyani dkk, 2020).

**2.2. Klasifikasi Menggunakan Algoritma CART**

Klasifikasi dengan algoritma CART dilakukan melalui tahapan sebagai berikut (Tanjung dan Kartiko, 2017):

- a. Pemilihan pemilah berdasarkan *indeks Gini* menggunakan Persamaan (1), yang kemudian akan dievaluasi dengan menggunakan kriteria *goodness of split* menggunakan Persamaan (2).
  - 1) Rumus menghitung *indeks Gini*

$$i(t) = 1 - \sum_{j=1}^n P(j|t)^2, \quad j = (0,1) \dots\dots\dots (1)$$

dengan,

$i(t)$  : nilai *indeks Gini*

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

- 2) Rumus menghitung *goodness of split*:

$$\Delta i(s, t) = i(t) - P_{Li}(t_L) - P_{Ri}(t_R) \dots\dots\dots (2)$$

dengan,

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

$P_{Li}(t_L)$  : proporsi pengamatan dari simpul t menuju kiri

$P_{Ri}(t_R)$  : proporsi pengamatan dari simpul t menuju kanan

- b. Menentukan simpul terminal menggunakan Persamaan (3):

$$\Delta i(s^*, t) = \max_{s \in S} \Delta i(s, t) \dots\dots\dots (3)$$

- c. Penandaan label kelas menggunakan Persamaan (4):

$$P(j|t) = \max \{j: (N(j|t))/(N(t))\} \dots\dots\dots (4)$$

dengan,

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

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

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

- d. Menentukan *Confusion Matrix*

Langkah-langkah dalam algoritma CART juga dapat dilihat di (Adhitya dkk, 2023) dan (Arrahimi dkk, 2019).

**2.3. Klasifikasi Menggunakan Algoritma KNN**

Langkah-langkah dalam klasifikasi dengan algoritma KNN adalah sebagai berikut:

- a. Menentukan jumlah *k* (tetangga terdekat)
- b. Menghitung jarak Euclid menggunakan Persamaan (5):

$$d_{(x,y)} = \sqrt{\sum_{j=1}^m (x_j - y_j)^2} \dots\dots\dots (5)$$

dengan,

$d_{(x,y)}$  : jarak antar objek ke-x dan objek ke-y

m : jumlah variabel

$x_j$  : nilai variabel ke j dari objek x

$y_j$  : nilai variabel ke j dari objek y

- c. Melakukan prediksi klasifikasi
- d. Menentukan *confusion matrix*

**2.4. Membandingkan Akurasi Metode CART dan KNN**

Evaluasi model dilakukan untuk mengetahui apakah model yang diperoleh telah baik dan layak digunakan atau tidak. Pada umumnya untuk melihat evaluasi kinerja model pada klasifikasi dapat menggunakan *confusion matrix*. *Confusion matrix* adalah alat yang berguna untuk menganalisis seberapa baik atau seberapa akurat model klasifikasi dapat mengenali objek pengamatan dari kelas atau kategori yang berbeda (Sihombing dan Yuliati,

2021). Untuk mempermudah perhitungan tingkat ketepatan klasifikasi digunakan bantuan Tabel 1.

**Tabel 1.** *Confusion Matrix*

Prediksi	Aktual	
	Positif	Negatif
Positif	<i>TP</i>	<i>FP</i>
Negatif	<i>FN</i>	<i>TN</i>

Rumus untuk menghitung akurasi adalah sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} \times 100\% \dots\dots\dots (6)$$

### 3. HASIL DAN PEMBAHASAN

Analisis yang dilakukan meliputi hasil klasifikasi provinsi di Indonesia menggunakan metode CART dan KNN berdasarkan PDRB per kapita per tahun atas harga konstan dengan variabel sebagaimana disebutkan sebelumnya. Analisis klasifikasi meliputi statistik deskriptif, pembagian data *training* dan data *testing*, penentuan *maxdepth* dan nilai *k*, pembentukan model dan prediksi klasifikasinya, tabel *confusion matrix* dan pemilihan model terbaik.

#### 3.1. Statistik Deskriptif

Nilai statistika deskriptif ditunjukkan pada Tabel 2 berikut.

**Tabel 2.** Statistika Deskriptif Variabel-variabel Penelitian

	X <sub>1</sub> (%)	X <sub>2</sub> (Rp)	X <sub>3</sub> (%)	X <sub>4</sub> (tahun)	X <sub>5</sub> (%)	Y (Rp)
N	34	34	34	34	34	34
Nilai Minimum	2,340	1.812.935	4,530	7,310	63,080	13.298.850
Nilai Maksimum	8,310	4.641.854	26,800	11,300	77,750	182.908.690
Mean	4,966	2.723.613	10,299	9,247	68,640	45.360.824
Standar Deviasi	1,600	581.082	5,286	0,817	3,641	34.003.223

Berdasarkan Tabel 2 diketahui bahwa jumlah data dari setiap variabel sebanyak 34 data. Variabel persentase tingkat pengangguran terbuka (X<sub>1</sub>) di Indonesia memiliki persentase terendah 2,340% dan tertinggi 8,310%. Persentase tingkat pengangguran terbuka di Indonesia memiliki rata-rata 4,966% dengan standar deviasinya 1,600%. Variabel upah minimum provinsi (X<sub>2</sub>) di Indonesia terendah sebesar Rp 1.812.935 dan tertinggi sebesar Rp 4.641.854 dengan rata-rata sebesar Rp 2.723.613 dan standar deviasinya sebesar Rp 581.012. Variabel persentase penduduk miskin (X<sub>3</sub>) terendah sebesar 4,530% dan tertinggi sebesar 26,800% dengan rata-rata 10,299% dan standar deviasi 5,286%. Variabel rata-rata lama sekolah (X<sub>4</sub>) terendah sebesar 7,310 tahun dan tertinggi sebesar 11,300 tahun dengan rata-rata sebesar 9,247 tahun dan standar deviasi 0,817 tahun. Variabel tingkat partisipasi angkatan kerja (X<sub>5</sub>) terendah sebesar 63,080% dan tertinggi sebesar 77,75% dengan rata-rata sebesar 68,64% dan standar deviasi sebesar 3,641%. Variabel PDRB per kapita per tahun atas harga konstan (Y) terendah sebesar Rp 13.298.850 per jiwa dan tertinggi sebesar Rp 182.908.690 per jiwa dengan rata-rata sebesar Rp 45.360.824 per jiwa dan standar deviasi sebesar Rp 34.003.223 per jiwa. Analisis deskriptif ini dilakukan dengan menggunakan *software* R-Studio.

#### 3.2. Proporsi Data Training Dan Data Testing

Pembagian data *training* dan data *testing* dilakukan secara acak menggunakan *software* R-Studio dengan tujuan memastikan keadilan dalam evaluasi model yang dikembangkan. Pada penelitian ini menggunakan 3 proporsi data *training* dan *testing* yaitu 80% : 20%, 70% : 30%, dan 60% : 40%. Pembagian data *training* dan *testing* dilakukan secara acak menggunakan fungsi *set.seed*. Fungsi *set.seed* bertujuan untuk membuat hasil berulang saat penulisan kode pengacakan dalam pembagian data *training* dan *testing*. Dengan menggunakan fungsi *set.seed* dapat dipastikan bahwa akan dihasilkan nilai acak yang sama setiap kali menjalankan kode yang sama. Pengacakan dilakukan sebanyak 3 kali yaitu menggunakan *set.seed* 867, 433, dan 367.

#### 3.3. Analisis Klasifikasi Menggunakan Metode CART

Analisis dan pembentukan pohon keputusan algoritma CART dilakukan dengan menggunakan 3 proporsi data yaitu 80% : 20%; 70% : 30%; 60% : 40% untuk data *training* dan *testing*. *Maxdepth* atau kedalaman yang digunakan yaitu *maxdepth* 1, 2, 3, dan 4. Setiap proporsi dilakukan pengacakan menggunakan fungsi *set.seed*. Fungsi *set.seed* yang digunakan yaitu *set.seed* 867, *set.seed* 433, *set.seed* 362. Berdasarkan hasil klasifikasi

menggunakan metode CART maka diperoleh nilai akurasi seperti pada Tabel 3. Berdasarkan nilai akurasi dari masing-masing proporsi di atas diketahui bahwa pada proporsi 60% : 40% yang memiliki rata-rata akurasi tertinggi berada pada *maxdepth* 3 dan 4 dengan rata-rata akurasi untuk data *training* sebesar 100% dan rata-rata akurasi untuk data *testing* sebesar 80,952%. Pada proporsi 70% : 30% diketahui bahwa nilai rata-rata akurasi tertinggi berada pada *maxdepth* 3 dan 4 dengan rata-rata akurasi untuk data *training* sebesar 98,611% dan rata-rata akurasi untuk data *testing* sebesar 73,333%. Pada proporsi 80% : 20% diketahui bahwa nilai rata-rata akurasi terbesar berada pada *maxdepth* 4 yaitu rata-rata akurasi data *training* sebesar 97,531% dan rata-rata akurasi data *testing* sebesar 80,952%. Berdasarkan uraian di atas diketahui bahwa proporsi 60% : 40% memiliki rata-rata akurasi terbesar yaitu 90,476%, sehingga dapat dikatakan proporsi 60% : 40% merupakan proporsi terbaik dan model terbaik pada proporsi tersebut berada pada *maxdepth* 3 dan 4.

**Tabel 3.** Nilai Akurasi CART Setiap Proporsi

Proporsi	Maxdepth	ser.seed	Akurasi (%) Training	Akurasi (%) Testing
60%:40%	1	867	90,000	85,714
		433	80,000	78,571
		362	90,000	78,571
		Rata - rata	86,667	80,952
	2	867	95,000	85,714
		433	90,000	92,857
		362	90,000	78,571
		Rata - rata	91,667	85,714
	3	867	100,000	85,714
		433	100,000	100,000
		362	100,000	57,143
		Rata - rata	100,000	80,952
	4	867	100,000	85,714
		433	100,000	100,000
		362	100,000	57,143
		Rata - rata	100,000	80,952
				90,476
70%:30%	1	867	87,500	90,000
		433	91,667	80,000
		362	91,667	50,000
		Rata - rata	90,278	73,333
	2	867	87,500	90,000
		433	100,000	60,000
		362	100,000	70,000
		Rata - rata	95,833	73,333
	3	867	95,833	90,000
		433	100,000	60,000
		362	100,000	70,000
		Rata - rata	98,611	73,333
	4	867	95,833	90,000
		433	100,000	60,000
		362	100,000	70,000
		Rata - rata	98,611	73,333
				85,972
80%:20%	1	867	88,889	85,714
		433	85,185	57,143
		362	88,889	85,714
		Rata - rata	87,654	76,190
	2	867	92,593	85,714
		433	92,593	85,714
		362	88,889	85,714
		Rata - rata	91,358	85,714
	3	867	96,296	57,143
		433	100,000	100,000
		362	92,590	71,429
		Rata - rata	96,295	76,191
	4	867	96,296	57,143
		433	100,000	100,000
		362	96,296	85,714
		Rata - rata	97,531	80,952
				89,242

### 3.4. Analisis klasifikasi menggunakan metode KNN

Analisis dan pembentukan pohon keputusan algoritma KNN dilakukan dengan menggunakan 3 proporsi data yaitu 80% : 20% ; 70% : 30% ; 60% : 40%. Setiap proporsi dilakukan pengacakan menggunakan fungsi *set.seed*. Fungsi *set.seed* yang digunakan yaitu *set.seed* 867, *set.seed* 433, *set.seed* 362. Berikut ini akan ditunjukkan proses pengklasifikasian PDRB per kapita per tahun atas harga konstan menggunakan proporsi 70% : 30% pada *set.seed* 867, *set.seed* 433 dan *set.seed* 362. Untuk nilai akurasi dari setiap proporsi ditunjukkan pada Tabel 4, berikut.

**Tabel 4.** Nilai Akurasi dari Setiap Proporsi Metode KNN

Proporsi	<i>set.seed</i>	Akurasi Training(%)	Akurasi ( Testing)
60%:40%	867	85.000	71.430
	433	75.000	85.710
	362	80.000	78.570
	Rata - rata	80.000	78.570
		79.285	
70%:30%	867	79.170	90.000
	433	100.000	70.000
	362	100.000	50.000
	Rata - rata	93.057	70.000
		81.528	
80%:20%	867	81.480	71.430
	433	88.890	57.140
	362	81.480	71.430
	Rata - rata	83.950	66.667
		75.308	

Berdasarkan rata-rata nilai akurasi yang diperoleh dari ketiga proporsi pada *set.seed* 867, *set.seed* 433, *set.seed* 362 diketahui bahwa nilai rata-rata akurasi tertinggi berada pada proporsi 70% : 30% dengan nilai rata-rata akurasi data *training* sebesar 93,057%, nilai rata-rata akurasi data *testing* sebesar 70%, dan nilai rata-rata akurasi keseluruhannya sebesar 81,528%. Berdasarkan uraian di atas dapat dikatakan model terbaik untuk algoritma KNN yaitu model dengan proporsi 70% : 30% pada *set.seed* 867, *set.seed* 433, *set.seed* 362.

### 3.5. Pemilihan model terbaik

Pemilihan model terbaik dilihat berdasarkan nilai akurasi yang diperoleh dari kedua algoritma klasifikasi yaitu CART dan KNN. Nilai akurasi dari masing-masing algoritma dapat dilihat berdasarkan tabel *confusion matrix* CART maupun tabel *confusion matrix* KNN. Pada klasifikasi dengan algoritma CART menggunakan *maxdepth* atau kedalaman 4. Berikut disajikan nilai akurasi dari model algoritma terbaik metode CART dan KNN seperti pada Tabel 5.

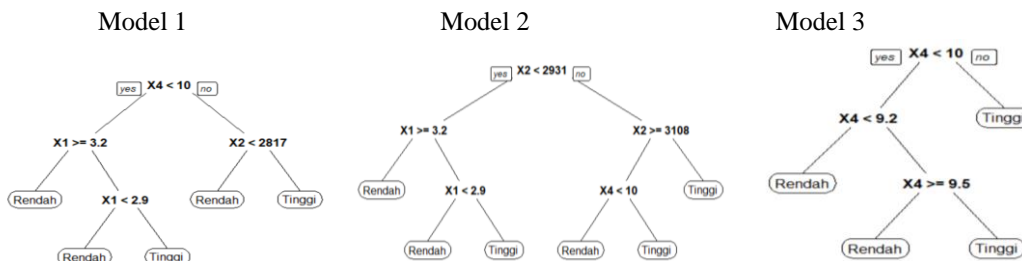
Berdasarkan Tabel 5 diketahui rata-rata nilai akurasi setiap metode dan diperoleh hasil bahwa metode CART dengan proporsi 60% : 40% merupakan metode terbaik dalam mengklasifikasikan PDRB per kapita per tahun atas harga konstan provinsi-provinsi di Indonesia dengan nilai rata-rata nilai akurasi sebesar 90,475%.

**Tabel 5.** Perbandingan Nilai Akurasi Metode CART dan KNN

NO	Metode	Propor si	<i>set.seed</i>	Akurasi (%)	Akurasi Testing(%)
1	CART	60%:40 %	867	100	85.714
			433	100	100
			362	100	57.143
	Rata - Rata		100	80.952	
		90.475			
2	KNN	70%:30 %	867	79.170	90
			433	100	70
			362	100	50
	Rata - Rata		93.057	70	
		81.528			

### 3.6. Interpretasi hasil model terbaik

Berdasarkan pemilihan model terbaik diketahui bahwa model terbaik adalah model klasifikasi menggunakan metode CART dengan proporsi 60%:40% karena memiliki akurasi tertinggi. Pohon keputusan model terbaik dengan menggunakan metode CART ditunjukkan pada Gambar 1.



Gambar 1. Pohon Keputusan Model Terbaik

Berdasarkan pohon keputusan model terbaik di atas diketahui bahwa faktor-faktor yang paling berpengaruh dalam proses klasifikasi PDRB per kapita per tahun atas harga konstan adalah sebagaimana yang ditunjukkan pada Tabel 6.

Tabel 6. Faktor-Faktor yang Paling Berpengaruh

Model	Faktor yang Paling Berpengaruh
1	Rata - Rata Lama Sekolah (X4)
	Tingkat Pengangguran Terbuka (X1)
	Upah Minimum Provinsi (X2)
2	Upah Minimum Provinsi (X2)
	Tingkat Pengangguran Terbuka (X1)
	Rata - Rata Lama Sekolah (X4)
3	Rata - Rata Lama Sekolah (X4)

Berdasarkan ketiga model terbaik dari metode CART diketahui bahwa variabel bebas yang menjadi faktor-faktor yang paling berpengaruh yaitu rata-rata lama sekolah (X4), tingkat pengangguran terbuka (X1), dan upah minimum provinsi (X2). Selanjutnya akan ditunjukkan data aktual dan hasil klasifikasi PDRB per kapita per tahun provinsi-provinsi di Indonesia pada model terbaik, seperti pada Tabel 7.

Berdasarkan Tabel 7 dapat diketahui daerah-daerah yang memiliki hasil klasifikasi PDRB per kapita per tahun atas harga konstan dengan benar sesuai dengan hasil kategori PDRB per kapita per tahun atas harga konstan berdasarkan nilai rata-ratanya maupun yang terklasifikasi salah. Dilihat dari faktor-faktor yang paling mempengaruhi klasifikasi PDRB per kapita per tahun atas harga konstan yaitu rata-rata lama sekolah (X4), tingkat pengangguran terbuka (X1), dan upah minimum provinsi (X2) maka daerah-daerah yang perlu dilakukan perbaikan atau penekanan faktor-faktor tersebut yaitu daerah-daerah yang terklasifikasi dengan benar sebagai daerah yang memiliki PDRB per kapita per tahun atas harga konstan kategori rendah pada model pohon keputusan terbaik. Daerah-daerah tersebut yaitu Provinsi Aceh, Sumatera Utara, Sumatera Selatan, Lampung, Kepulauan Bangka Belitung, Jawa Barat, Jawa Tengah, Jawa Timur, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Maluku Utara, dan Papua. Selain itu, daerah-daerah yang terklasifikasi dengan benar sebagai daerah yang memiliki PDRB per kapita per tahun atas harga konstan kategori tinggi perlu dilakukan upaya untuk tetap mempertahankan kategori tinggi tersebut. Daerah-daerah dengan PDRB per kapita per tahun atas harga konstan kategori tinggi yaitu Provinsi Kepulauan Riau, Daerah Khusus Ibu Kota Jakarta, Kalimantan Timur, Sulawesi Tengah, dan Papua Barat. Sedangkan untuk daerah-daerah yang terklasifikasi tidak tepat pada model terbaik ini yaitu Provinsi Sumatera Barat, Riau, Bengkulu, DI Yogyakarta, Banten, Kalimantan Utara, dan Maluku.

**Tabel 7.** Hasil Prediksi PDRB Per Kapita Model Terbaik

Provinsi	Kategori PDRB per kapita	Prediksi		
		Model 1	Model 2	Model 3
Aceh	Rendah	Rendah	Rendah	Rendah
Sumatera Utara	Rendah	Rendah	Rendah	Rendah
Sumatera Barat	Rendah	Rendah	Rendah	Tinggi
Riau	Tinggi	Rendah	Tinggi	Rendah
Jambi	Rendah	Rendah	Rendah	Rendah
Sumatera Selatan	Rendah	Rendah	Rendah	Rendah
Bengkulu	Rendah	Rendah	Rendah	Tinggi
Lampung	Rendah	Rendah	Rendah	Rendah
Kepulauan Bangka Belitung	Rendah	Rendah	Rendah	Rendah
Kepulauan Riau	Tinggi	Tinggi	Tinggi	Tinggi
Daerah Khusus Ibu Kota Jakarta	Tinggi	Tinggi	Tinggi	Tinggi
Jawa Barat	Rendah	Rendah	Rendah	Rendah
Jawa Tengah	Rendah	Rendah	Rendah	Rendah
Daerah Istimewah Yogyakarta	Rendah	Rendah	Rendah	Tinggi
Jawa Timur	Rendah	Rendah	Rendah	Rendah
Banten	Rendah	Rendah	Rendah	Tinggi
Bali	Rendah	Rendah	Rendah	Rendah
Nusa Tenggara Barat	Rendah	Rendah	Rendah	Rendah
Nusa Tenggara Timur	Rendah	Rendah	Rendah	Rendah
Kalimantan Barat	Rendah	Rendah	Rendah	Rendah
Kalimantan Tengah	Rendah	Rendah	Rendah	Rendah
Kalimantan Selatan	Rendah	Rendah	Rendah	Rendah
Kalimantan Timur	Tinggi	Tinggi	Tinggi	Tinggi
Kalimantan Utara	Tinggi	Rendah	Tinggi	Tinggi
Sulawesi Utara	Rendah	Rendah	Rendah	Rendah
Sulawesi Tengah	Tinggi	Tinggi	Tinggi	Tinggi
Sulawesi Selatan	Rendah	Rendah	Rendah	Rendah
Sulawesi Tenggara	Rendah	Rendah	Rendah	Rendah
Gorontalo	Rendah	Rendah	Rendah	Rendah
Sulawesi Barat	Rendah	Rendah	Rendah	Rendah
Maluku	Rendah	Rendah	Rendah	Tinggi
Maluku Utara	Rendah	Rendah	Rendah	Rendah
Papua Barat	Tinggi	Tinggi	Tinggi	Tinggi
Papua	Rendah	Rendah	Rendah	Rendah

#### 4. KESIMPULAN

Klasifikasi PDRB per kapita per tahun atas harga konstan terbagi menjadi dua, yaitu kategori rendah dan tinggi. Jumlah provinsi yang ada di Indonesia tergolong ke dalam PDRB per kapita per tahun atas harga konstan kategori rendah sebanyak 27 provinsi dan sebanyak 7 provinsi tergolong ke dalam PDRB per kapita per tahun atas harga konstan kategori tinggi. Berdasarkan rata-rata nilai akurasi setiap metode diperoleh hasil bahwa metode CART dengan proporsi 60% : 40% merupakan metode terbaik dalam mengklasifikasikan PDRB per kapita per tahun atas harga konstan provinsi-provinsi di Indonesia. Hal ini dikarenakan memiliki rata-rata nilai akurasi terbesar yaitu 90,475%.

#### DAFTAR PUSTAKA

- Adhitya R.R., Witanti W., Yuniarti R. (2023). Perbandingan Metode CART dan Naïve Bayes untuk Klasifikasi Customer CHURN. *Infotech Journal* Vol. 9 No. 2, pp. 307-318.
- Argina A.M. (2020). Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes. *Indonesian Journal of Data and Science*. Vol 1, No 2, Juli 2020, pp. 29-33
- Arrahimi A., Ihsan M.K., Kartini D., Faisal M.R., Indriani F. (2019) Teknik Bagging dan Boosting pada Algoritma CART untuk Klasifikasi Masa Studi Mahasiswa. *Jurnal Sains dan Informatika* Volume 5, Nomor 1, Juni 2019
- BPS. (2023). *Produk Domestik Regional Bruto Provinsi-provinsi di Indonesia*. Diakses tanggal 20 April 2024, dari <https://www.bps.go.id/id/publication/2023/04/13/d5ce181590472cea3abbb80c/produk-domestik-regional-bruto-provinsi-provinsi-di-indonesia-menurut-lapangan-usaha-2018-2022.html>
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. 1987. *Classification and Regression Trees*, Chapman & Hall/CRC, Boca Raton, Florida.
- Hartini, N.T., 2017, Pengaruh PDRB Per Kapita, Investasi dan IPM Terhadap Ketimpangan Pendapatan antar Daerah di Provinsi DIY Tahun 2011-2015. *Jurnal Pendidikan dan Ekonomi*, Volume 6, Nomor 6, Tahun



2017

- Prabawati, N., Widodo, dan Duskarnaen, M. (2019). Kinerja Algoritma Classification and Regression Trees (CART) dalam Mengklasifikasikan Lama Masa Studi Mahasiswa yang Mengikuti Organisasi di Universitas Negeri Jakarta. *Jurnal Pinter* 3(2), 139-145.
- Cahyani, D., Rahmayani, A., dan Husniar, S. A. (2020). Analisis Performa Metode KNN pada Dataset Pasien Pengidap Kanker Payudara. *Indonesian Journal of Data and Science*, 1(2), 39-43.
- Setio, P. B., Saputro, D. R., dan Winarno, B. (2020). Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika* (pp. 64-71). Surakarta: UNS Press.
- Sihombing, P., dan Yuliati, I. (2021). Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia. *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 20(2), 417-426.
- Sumartini S.H., dan Purnami S.W.( 2015). Penggunaan Metode Classification and Regression Trees (CART) untuk Klasifikasi Rekurensi Pasien Kanker Serviks di RSUD Dr. Soetomo Surabaya. *Jurnal Sains dan Seni ITS* Vol. 4, No.2, D211-D216.
- Tanjung, R. H., dan Kartiko. (2017). Penerapan Metode CART (Classification And Regression Trees) Untuk Menentukan Faktor-faktor yang Mempengaruhi Pembayaran Kredit oleh Nasabah (Studi Kasus Bank BRI Unit AEK TARUM-Sumatera Utara). *Jurnal Statistika Industri dan Komputasi*, 2(2), 78-83.