

## KLASIFIKASI STATUS KEMISKINAN KABUPATEN/KOTA DI INDONESIA MENGUNAKAN *NAÏVE BAYES CLASSIFIER* DAN *CLASSIFICATION AND REGRESSION TREE*

Kristina Millenia Seran<sup>1</sup>, Yudi Setyawan<sup>2\*</sup>, Maria Titah Jatipaningrum<sup>3</sup>, Febriani Astuti<sup>4</sup>  
<sup>1,2,3,4</sup> Jurusan Statistika, Fakultas Sains Terapan, Institut Sains & Teknologi AKPRIND Yogyakarta  
Email: [setyawan@akprind.ac.id](mailto:setyawan@akprind.ac.id)

\*corresponding author

**Abstrak :** Indonesia masuk dalam 100 besar negara miskin dengan menempati urutan ke-73 di dunia. Diketahui persentase penduduk miskin dilihat pada Maret 2022 sebesar 5,94% atau 26,16 juta jiwa, meningkat pada September 2022 menjadi 9,57% atau 26,36 juta jiwa. Status kemiskinan adalah kondisi dimana seorang atau sekelompok orang tidak mampu memenuhi kebutuhan hidup. Kemiskinan juga dapat dipengaruhi oleh beberapa faktor seperti pengeluaran perkapita, rata-rata lama sekolah, tingkat pengangguran terbuka serta indeks pembangunan manusia. Data yang digunakan dalam penelitian ini yaitu data kabupaten/kota di Indonesia tahun 2022. Pada penelitian ini klasifikasi dilakukan dengan menggunakan metode *Naïve Bayes Classifier* dan *Classification and Regression Tree*. Kedua metode dilakukan dengan membagi data *training* dan data *testing* sebesar 90%:10%, 80%:20%, dan 70%:30%. Selain itu, dilakukan pengacakan sampel sebanyak 30 kali dengan fungsi *set.seed()*. Hasil analisis dengan kedua metode diperoleh metode *Classification and Regression Tree* lebih baik digunakan dengan rata-rata akurasi untuk proporsi data 90%:10% sebesar 77,81%, untuk proporsi data 80%:20% sebesar 78,31%, dan untuk proporsi data 70%:30% sebesar 79,67%. Sedangkan metode *Naïve Bayes Classifier* menghasilkan rata-rata akurasi untuk proporsi data 90%:10% sebesar 72,57%, untuk proporsi data 80%:20% sebesar 73,00%, dan untuk proporsi data 70%:30% sebesar 73,13%. Kedua metode ini masing-masing menghasilkan nilai akurasi yang tinggi pada proporsi data 70%:30%.

**Kata Kunci:** Klasifikasi, Status kemiskinan, *Naïve Bayes Classifier*, *Classification and Regression Tree*

**Abstract :** Indonesia is included in the top 100 poor countries by ranking 73rd in the world. It is known that the percentage of poor people seen in March 2022 was 5.94% or 26.16 million people, increasing in September 2022 to 9.57% or 26.36 million people. Poverty status is a condition where a person or group of people is unable to meet the needs of life. Poverty can also be influenced by several factors such as per capita expenditure, average length of schooling, open unemployment rate, and human development index. The data used in this study is district / city data in Indonesia. In this study, classification was carried out using the *Naïve Bayes Classifier* and *Classification and Regression Tree* methods. . Both methods are carried out by dividing training and testing data by 90%:10%, 80%:20%, and 70%:30%. In addition, sample randomization is carried out 30 times with the *set.seed()* function. The results of the analysis with both methods obtained the *Classification and Regression Tree* method are better used with an average accuracy for the proportion of data 90%: 10% of 77.81%, for the proportion of data 80%: 20% of 78.31%, and for the proportion of data 70% : 30% of 79.67%. While the *Naïve Bayes Classifier* method produces an average accuracy for the proportion of 90%:10% data of 72.57%, for the proportion of 80%:20% data of 73.00%, and for the proportion of 70%:30% data of 73.13%. These two methods each produce a high accuracy value at a data proportion of 70%:30%.

**Keywords:** *Classification, Poverty status, Naïve Bayes Classifier, Classification and Regression Tree*

### 1. Pendahuluan

Kemiskinan merupakan salah satu persoalan mendasar dan masalah yang sulit untuk diatasi di seluruh negara yang ada di dunia termasuk di Indonesia. Hal ini karena, kemiskinan memiliki sifat yang kompleks artinya kemiskinan yang ada tidak muncul sendiri secara tiba-tiba akan tetapi dipengaruhi oleh beberapa hal yaitu tidak terpenuhinya kebutuhan dasar yang mencakup aspek primer dan sekunder. Salah satu indikator dari kemiskinan adalah garis kemiskinan yang

merupakan tingkat minimum pendapatan yang dianggap perlu dipenuhi untuk memperoleh standar hidup yang mencukupi di suatu daerah atau pengeluaran minimum yang diperlukan seseorang untuk memenuhi kebutuhan pokok dalam sebulan.

Status kemiskinan dapat dilihat dari tingkat kemiskinan yaitu persentase penduduk miskin suatu daerah. Menurut BPS persentase penduduk miskin adalah penduduk yang memiliki rata-rata pengeluaran perkapita per bulan di bawah garis kemiskinan. Indonesia memiliki 34 provinsi dan 514 kabupaten/kota dengan jumlah penduduk miskin di kabupaten/kota di Indonesia tidak merata sehingga perlu dilakukan klasifikasi untuk mengetahui status kemiskinan setiap kabupaten/kota di Indonesia. Kategori dalam pengklasifikasian ini adalah kategori tinggi jika nilai persentase penduduk miskin suatu kabupaten/kota di atas atau sama dengan rata-rata dan kategori rendah jika nilai persentase penduduk miskin suatu kabupaten/kota di bawah rata-rata. Klasifikasi merupakan salah satu teknik data mining untuk memetakan data ke dalam beberapa kelompok atau kategori tertentu (Setio, Saputro, & Winarno, 2020). Pengklasifikasian ini dilakukan dengan menggunakan metode klasifikasi yaitu *Naïve Bayes Classifier* dan *Classification and Regression Tree*. Untuk menentukan metode terbaik dapat dilihat dari rata-rata nilai akurasi tertinggi. Beberapa penelitian terdahulu terkait *Naïve Bayes Classifier* dan *Classification and Regression Tree* antara lain:

- a. Nurul Insan, Mustika Hadijati, dan Irwansyah pada tahun 2020 melakukan penelitian untuk membandingkan hasil klasifikasi status gizi balita di Kelurahan Pagesangan Barat menggunakan metode *Classification and Regression Tree* (CART) dan *Naïve Bayes Classifier* (NBC).
- b. Erma Shofi Utami, Yudi Setyawan dan Noeryanti pada tahun 2022 melakukan penelitian terhadap tingkat kemiskinan kabupaten/kota di Indonesia berdasarkan tingkat kedalaman dan keparahan kemiskinan menggunakan *Naïve Bayes Classifier* (NBC) dan *K-Nearest Neighbor* (KNN).
- c. Inggit Fatika, Kris Suryowati dan Noviana Pratiwi pada tahun 2022 dalam penelitian mengklasifikasikan tingkat pengangguran terbuka di Indonesia dengan algoritma *Classification and Regression Tree* (CART) dan C4.5
- d. Naisah Marito Erfin Nur Rohma Khakim pada tahun 2022 melakukan penelitian dengan membandingkan algoritma klasifikasi data kesejahteraan sosial kabupaten Bantul menggunakan *Naïve Bayes Classifier* dan *Decision Tree*

Pada penelitian ini menggunakan 30 kali pengacakan data dengan fungsi *set.seed()*. Penelitian ini diharapkan dapat membantu pemerintah dalam mengambil kebijakan untuk mengatasi kemiskinan.

## 2. Metode

Penelitian ini menggunakan data sekunder yang bersumber dari Badan Pusat Statistik (BPS) tahun 2022 yang diakses melalui link pada website <https://bps.go.id/>. Data yang digunakan adalah data tingkat kemiskinan (Y), Pengeluaran perkapita ( $X_1$ ) rata-rata lama sekolah ( $X_2$ ), tingkat pengangguran terbuka ( $X_3$ ), indeks pembangunan manusia ( $X_4$ ), kabupaten/kota di Indonesia tahun 2022 yang terdiri dari 479 kabupaten/kota. Untuk melihat status kemiskinan kabupaten/kota di Indonesia maka variabel dependen (Y) yang digunakan yaitu persentase penduduk miskin kemudian dihitung nilai rata-ratanya setelah itu dilakukan kategorisasi data menjadi dua kategori berdasarkan nilai rata-rata yaitu kategori tinggi ( $x > \text{rata-rata}$ ) dan kategori rendah ( $x \leq \text{rata-rata}$ ). Jika suatu kabupaten/kota termasuk dalam kategori tinggi berarti kabupaten/kota tersebut termasuk miskin dan jika suatu kabupaten/kota termasuk dalam kategori rendah berarti kabupaten/kota tersebut termasuk tidak miskin.

### a. Naïve Bayes Classifier

Klasifikasi *Naïve Bayes* merupakan salah satu tipe pembelajaran *supervised* yang termasuk kedalam metode klasifikasi. Klasifikasi *Naïve Bayes* adalah klasifikasi berdasarkan teorema Bayes yang digunakan untuk menghitung probabilitas tiap kelas dengan asumsi bahwa antar satu kelas dengan kelas yang lain tidak saling tergantung (*independen*). Metode klasifikasi *Naïve Bayes* memiliki beberapa kelebihan, sebagai berikut :

- Hanya membutuhkan jumlah data training yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses klasifikasi.
- Bisa digunakan untuk data kualitatif maupun kuantitatif.
- Jika terdapat nilai yang hilang, maka dapat diabaikan dalam perhitungan.
- Bisa digunakan untuk klasifikasi kasus biner maupun *multiclass*.
- Perhitungannya cepat, efisien, mudah dipahami, serta bahasa pemrograman yang sederhana.

Pada metode ini didasari dengan teorema Bayes dengan persamaan sebagai berikut:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

Dimana:

$P(Y)$  : probabilitas kelas awal Y (*prior probability*)

$P(X)$  : probabilitas X

$P(Y|X)$ : probabilitas akhir bersyarat (*posterior probability*) suatu kelas Y terjadi jika diberikan petunjuk (atribut) X

$P(X|Y)$ : probabilitas sebuah petunjuk (atribut) X terjadi akan mempengaruhi kelas Y

Klasifikasi *Naïve Bayes* pada penelitian ini dengan data bertipe kontinu menggunakan rumus distribusi *Gauss* seperti berikut:

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi}\sigma_{i,j}} \exp \frac{-(x_i - \mu_j)^2}{2\sigma_{i,j}^2}$$

Keterangan:

P : Peluang

$X_i$  : Atribut ke-i

$x_i$  : Nilai atribut ke-i

Y : Kelas yang dicari

$y_j$  : Sub kelas yang dicari

$\sigma_{i,j}$  : Standar deviasi untuk  $X_i = x_i$  untuk  $Y = y_j$

$\sigma_{i,j}^2$  : Varian untuk  $X_i = x_i$  untuk  $Y = y_j$

Langkah-langkah perhitungan dengan *Naïve Bayes Classifier* menggunakan data kontinu adalah sebagai berikut:

- Menghitung probabilitas (*prior*) tiap kelas yang ada
- Menghitung rata-rata (*mean*) tiap fitur
- Menghitung nilai standar deviasi dari fitur
- Setelah itu menghitung nilai *likelihood* dengan menggunakan rumus:

$$Likelihood = P(X_i | Y_j)$$

- Kemudian menghitung *posterior probability* dengan menggunakan rumus:

$$P(X|Y) = P(Y) \cdot P(X|Y)$$

- Melakukan normalisasi nilai probabilitas dengan menggunakan rumus berikut :

$$NP(Y_1|X_i) = \frac{P(Y_1|X_i)}{P(Y_1|X_i) + P(Y_2|X_i)}$$

$$NP(Y_2|X_i) = \frac{P(Y_2|X_i)}{P(Y_1|X_i) + P(Y_2|X_i)}$$

Dimana:

$NP(Y_1|X_i)$  : Normalisasi nilai probabilitas kategori (*class*) pertama

$NP(Y_2|X_i)$  : Normalisasi nilai probabilitas kategori (*class*) kedua

## b. Classification and Regression Tree

*Classification and Regression Tree* (CART) merupakan salah satu metode algoritma dari salah satu teknik eksplorasi data *decision tree* (Ishaq, 2017). CART merupakan metode statistika nonparametrik yang dikembangkan untuk dapat memilih peubah yang paling mempengaruhi peubah respon kategorik maupun kontinu. *Classification and Regression Tree* (CART) merupakan salah satu algoritma dari salah satu teknik eksplorasi data yaitu teknik pohon keputusan (Andarama, dkk 2020). Kelebihan dari penggunaan algoritma CART adalah sebagai berikut:

- 1) Tidak memerlukan asumsi dan uji hipotesis karena termasuk bentuk statistika nonparametrik.
- 2) Dapat digunakan untuk menangani bentuk *dataset* yang kompleks
- 3) Tidak dipengaruhi data *outlier*
- 4) Dapat digunakan untuk berbagai tipe data baik itu kontinu maupun kategorik
- 5) Tidak perlu dilakukan transformasi karena jenis data asli tidak akan berpengaruh terhadap bentuk pohon keputusan
- 6) Mudah melakukan eksplorasi, interpretasi, serta pengambilan keputusan karena struktur data dapat dilihat secara visual.

Pohon keputusan dibentuk dengan menggunakan algoritma penyekatan rekursif secara biner (*binary recursive partitioning*). Pengertian *binary* berarti setiap simpul awal akan terbagi menjadi dua kelompok anak simpul.

Tahapan analisis menggunakan algoritma *Classification and Regression Tree* (CART) adalah sebagai berikut:

### 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:

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

$i(t)$  : 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

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

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 Breiman, dkk (1993) 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:

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

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  $t$

d. 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 keberhasilan sistem dalam menemukan informasi atau juga dapat menunjukkan derajat keandalan model dalam memprediksi data berlabel positif dengan benar.
- 2) *Specificity* digunakan untuk mengukur tingkat keberhasilan sistem yang menunjukkan derajat keandalan model dalam mendeteksi data berlabel negatif dengan benar.
- 3) *Accuracy* digunakan untuk mengukur tingkat ketepatan klasifikasi data.

Tabel 1. Ketepatan klasifikasi

<i>Confussion matrix</i>	Data Aktual	
	1	2
1	$TP/n_{11}$	$FP/n_{12}$
2	$FN/n_{21}$	$TN/n_{22}$

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

$$Sensitivity/recall = \frac{n_{11}}{n_{11} + n_{21}}$$

$$Specivity = \frac{n_{22}}{n_{12} + n_{22}}$$

$$Akurasi(1 - APER) = \frac{n_{11} + n_{22}}{n_{11} + n_{21} + n_{21} + n_{22}}$$

Keterangan:

APER : *Apparent Error Rate*

$n_{11}$  : jumlah observasi dari kelas 1 yang tepat diprediksi sebagai kelas 1

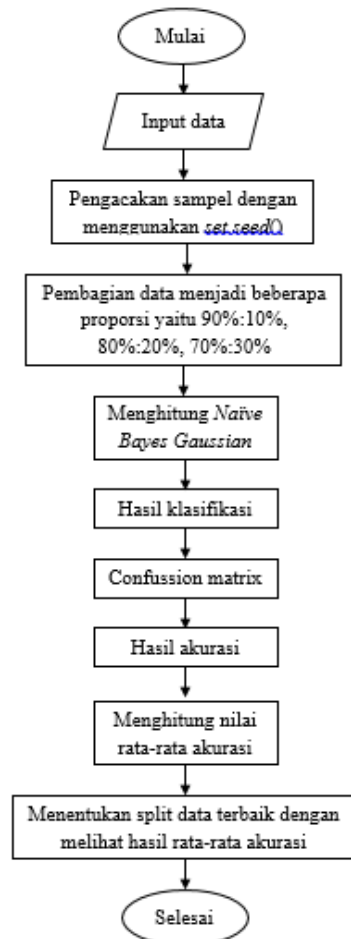
$n_{12}$  : jumlah observasi dari kelas 1 yang tepat diprediksi sebagai kelas 2

$n_{21}$  : jumlah observasi dari kelas 2 yang tepat diprediksi sebagai kelas 1

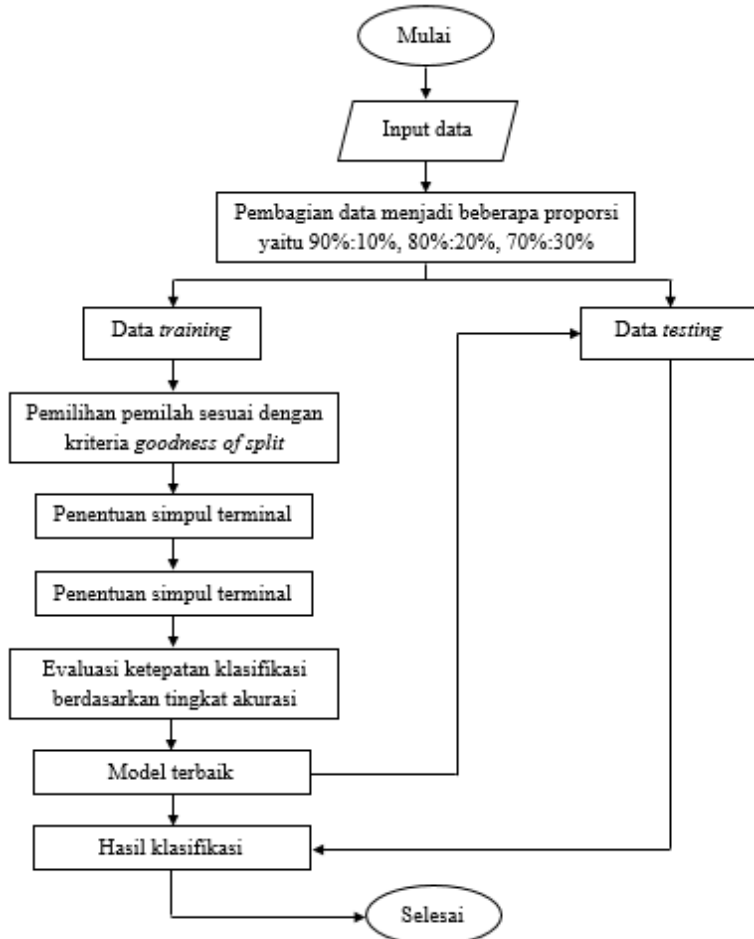
$n_{22}$  : jumlah observasi dari kelas 2 yang tepat diprediksi sebagai kelas 2

**Diagram Alir**

*Flowchart* analisis *Naïve Bayes Classifier* disajikan pada **Gambar 1**, dan *Flowchart* analisis *Classification and Regression Tree* disajikan pada **Gambar 2**.



Gambar 1. Flowchart NBC



Gambar 2. Flowchart CART

### 3. Hasil dan Pembahasan

Statistik deskriptif yang digunakan untuk mengetahui gambaran karakteristik status kemiskinan kabupaten/kota di Indonesia tahun 2022 yang disajikan dalam **Tabel 1**.

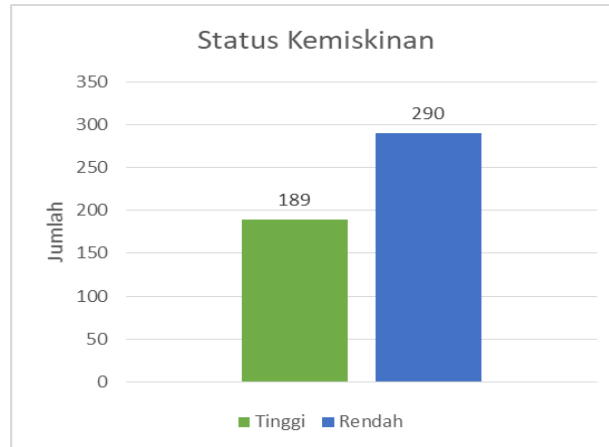
Tabel 1. Statistik Deskriptif

	Y (Persen)	X <sub>1</sub> (Ribu Rupiah)	X <sub>2</sub> (Tahun)	X <sub>3</sub> (Persen)	X <sub>4</sub> (Persen)
N	479	479	479	479	479
Nilai Minimum	2,28	4.190	1,58	0,09	34,10
Nilai Maksimum	42,03	24.221	13,03	14,73	87,69
Rata-rata	11,48	10.730,84	8,52	4,69	70,69
Median	9,49	10.588	8,35	4,35	70,36
Standar Deviasi	7,32	2.783,01	1,65	2,44	6,53

Keterangan :

Persentase penduduk miskin (Y), Pengeluaran per kapita (X<sub>1</sub>), Rata-rata lama sekolah (X<sub>2</sub>), Tingkat pengangguran terbuka (X<sub>3</sub>), Indeks pembangunan manusia (X<sub>4</sub>)

Status kemiskinan dikategorikan menjadi dua kategori yaitu kategori tinggi ( $x > \text{rata-rata}$ ) dan kategori rendah ( $x \leq \text{rata-rata}$ ). Jika suatu kabupaten/kota termasuk dalam kategori tinggi berarti kabupaten/kota tersebut termasuk miskin dan jika suatu kabupaten/kota termasuk dalam kategori rendah berarti kabupaten/kota tersebut termasuk tidak miskin. Diperoleh jumlah kabupaten/kota yang berstatus miskin dengan persentase penduduk miskin tinggi ( $x \geq \text{rata-rata}$ ) sebanyak 189 kabupaten/kota atau sebesar 39% dan 290 kabupaten/kota atau sebesar 61% yang berstatus tidak miskin dengan persentase penduduk miskin rendah ( $x > \text{rata-rata}$ ) yang disajikan pada **Gambar 3**. Histogram status kemiskinan kabupaten/kota di Indonesia tahun 2022.



**Gambar 3.** Histogram status kemiskinan kabupaten/kota di Indonesia tahun 2022.

Analisis *Naïve Bayes Classifier* dan *Classification and Regression Tree* untuk melihat status kemiskinan kabupaten/kota di Indonesia berdasarkan persentase penduduk miskin menggunakan *software* Rstudio dilakukan dengan menggunakan tiga proporsi data yaitu 90%:10%, 80%:20%, dan 70%:30% dengan pengacakan sebanyak 30 kali dengan menggunakan fungsi *set.seed* sehingga diperoleh hasil akurasi yang berbeda. Proporsi pembagian data ini bersifat subjektif tergantung peneliti yang mana proporsi data *training* lebih besar dibandingkan dengan data *testing*.

Hasil klasifikasi menggunakan metode *Naive Bayes Classifier* dengan pengacakan sebanyak 30 kali disajikan dalam **Tabel 2**.

**Tabel 2.** Rerata hasil akurasi status kemiskinan menggunakan metode *Naïve Bayes Classifier*

Proporsi	Akurasi
90% : 10%	72,57%
80% : 20%	73,00%
70% : 30%	73,13%

Berdasarkan Tabel 4.3 diperoleh nilai rata-rata akurasi status kemiskinan berdasarkan persentase penduduk miskin dengan proporsi data 90%:10% sebesar 72,57%, untuk proporsi data 80%:20% sebesar 73,00%, dan proporsi data 70%:30% sebesar 73,13%. Oleh karena itu hasil akurasi terbaik adalah hasil akurasi yang paling tinggi yaitu dengan proporsi data 70%:30% sebesar 73,13%.

**Tabel 3.** Rerata hasil akurasi status kemiskinan menggunakan metode *Classification and Regression Tree*

Proporsi	Akurasi
90% : 10%	77,81%
80% : 20%	78,31%
70% : 30%	79,67%

Berdasarkan Tabel 4.8 diperoleh nilai rata-rata akurasi status kemiskinan berdasarkan persentase penduduk miskin dengan proporsi data 90%:10% sebesar 77,81%, untuk proporsi data 80%:20% sebesar 78,31%, dan proporsi data 70%:30% sebesar 79,67%. Oleh karena itu hasil akurasi terbaik adalah hasil akurasi yang paling tinggi yaitu dengan proporsi data 70%:30% sebesar 79,67%.

Berdasarkan perbandingan hasil analisis menggunakan metode *Naïve Bayes Classifier* dan *Classification and Regression Tree* dengan pengacakan sebanyak 30 kali diketahui bahwa secara keseluruhan metode *Classification and Regression Tree* menghasilkan nilai akurasi lebih tinggi dibandingkan dengan metode *Naïve Bayes Classifier*. Hasil akurasi dari analisis menggunakan metode *Naïve Bayes Classifier* dan *Classification and Regression Tree* pada proporsi data 90%:10%, 80%:20% dan 70%:30% diketahui bahwa semakin banyak data *training* semakin kecil nilai akurasi yang dihasilkan. Hal tersebut terjadi karena jumlah data pada kelas tidak seimbang sehingga model tidak dapat mengenali kelas minoritas dengan baik atau sering disebut dengan *imbalanced classification*.

#### 4. Kesimpulan

Jumlah data yang digunakan dalam penelitian ini sebanyak 479. Diperoleh jumlah kabupaten/kota yang berstatus miskin dengan persentase penduduk miskin tinggi ( $x \geq$  rata-rata) sebanyak 189 kabupaten/kota dan 290 kabupaten/kota yang berstatus tidak miskin dengan persentase penduduk miskin rendah ( $x <$  rata-rata). Dari 30 kali pengulangan diperoleh nilai akurasi dari metode *Naïve Bayes Classifier* (NBC) pada proporsi data 90% : 10% sebesar 72,57%, 80% : 20% sebesar 73,00%, dan 70% : 30% sebesar 73,13%. Sedangkan nilai akurasi yang dihasilkan dari metode *Classification and Regression Tree* (CART) pada proporsi data 90% : 10% sebesar 77,81%, 80% : 20% sebesar 78,31%, dan 70% : 30% sebesar 79,67%. Berdasarkan perbandingan hasil analisis menggunakan metode *Naïve Bayes Classifier* (NBC) dan *Classification and Regression Tree* (CART) dengan pengacakan sebanyak 30 kali diketahui bahwa secara keseluruhan metode *Classification and Regression Tree* (CART) menghasilkan nilai akurasi lebih tinggi dibandingkan dengan metode *Naïve Bayes Classifier* (NBC).

Saran untuk penelitian selanjutnya yaitu untuk meningkatkan performa klasifikasi dengan menambahkan metode untuk mengatasi *imbalanced classification*, Untuk mendapatkan nilai akurasi yang lebih tinggi dengan menambahkan variabel, menggantikan variabel yang memiliki pengaruh lebih kecil dengan variabel lain atau dapat menggunakan metode klasifikasi lainnya dan dapat menggunakan lebih banyak proporsi data. Selain itu peneliti menyarankan untuk dapat membandingkan dengan metode klasifikasi lainnya.

#### Daftar Pustaka

- Andarama, A., Nurfalinda, & Ritha, N. (2020). Implementasi Algoritma *Classification and Regression Trees* (CART) dalam Klasifikasi Penerima Bantuan Sosial Beras Sejahtera. *Student Online Journal*, 1.
- Fatika, I. (2022). *Klasifikasi Tingkat Pengangguran Terbuka di Indonesia dengan Algoritma Classification and Regression Tree (CART) dan C4.5*. Yogyakarta: IST AKPRIND.
- Insan, N., Hadijati, M., & Irwansyah. (2020, Juni). Perbandingan Metode *Classification and Regression Tree* (CART) dengan *Naive Bayes Classification* (NBC) dalam Klasifikasi Status Gizi Balita di Kelurahan Pagesangan Barat. *Eigen Mathematics*, 3.
- Ishaq, Y. (2017). Prediksi *Google Search Engine Result Page* (SERP) Menggunakan *Classification and Regression Tree* (CART). *e-Proceeding of Engineering*, 2997-3002.



- Khakim, E. N. (2022, Oktober). Perbandingan Algoritma Klasifikasi Data Kesejahteraan Sosial Kabupaten Bantul. *Jurnal Ilmiah Sistem Informasi, Teknologi Informasi dan Sistem Komputer*, 17.
- Utami, E. S., Setyawan, Y., & Noeryanti. (2022, November). Klasifikasi Kabupaten/kota di Indonesia Berdasarkan Tingkat Keparahan dan Kedalaman Kemiskinan Menggunakan *Naive Bayes Classifier* dan *K-Nearest Neighbor*. *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST)*.
- Setio, P. B., Saputro, D. r., & Winarno, B. (2020). Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika*, 3.