



Perbandingan Metode *Classification and Regression Trees* (CART) dengan *Naïve Bayes Classification* (NBC) dalam Klasifikasi Status Gizi Balita di Kelurahan Pagesangan Barat

Nurul Insan^{a,*}, Mustika Hadijati^b, Irwansyah^c

^a Program Studi Matematika, FMIPA, Universitas Mataram, Jl. Majapahit No. 62, Mataram, Indonesia, 83125.
Email: nurulinsan5@yahoo.com

^b Program Studi Matematika, FMIPA, Universitas Mataram, Jl. Majapahit No. 62, Mataram, Indonesia, 83125.
Email: mustika.hadijati@unram.ac.id

^c Program Studi Matematika, FMIPA, Universitas Mataram, Jl. Majapahit No. 62, Mataram, Indonesia, 83125.
Email: irw@unram.ac.id

ABSTRACT

This study aims to compare the Classification and Regression Trees (CART) and Naïve Bayes Classification (NBC) methods in classifying the nutritional status of toddlers in West Pagesangan by looking at their accuracy and also knowing the variables that influence the classification of toddler nutritional status. The data used in this study were toddlers who come to the posyandu in May 2019, with predictor variables used namely gender, ages, weight, mother's employment status, mother's education level, number of children and parents income. The result showed that Naïve Bayes Classification (NBC) is better in classifying the nutritional status of toddlers in West Pagesangan than Classification and Regression Trees (CART). This can be seen from the accuracy values obtained with three comparisons of training data and testing data. In the comparison of 90% of training data: 10% of testing data, obtained an accuracy value of 90% for NBC and 85% for CART, in the comparison of 80% of training data: 20% of testing data, obtained an accuracy value of 82.5% for NBC and 80% for CART, while in comparison 70% training data : 30% testing data, obtained an accuracy value 72% for NBC and 70% for CART. This study also showed that significant variables the classification of nutritional status of toddlers in West Pagesangan village are age, gender, weight and parents income.

Keywords: Accuracy, Classification, Classification and Regression Trees (CART), Naïve Bayes Classification (NBC), Toddler Nutrition Status.

ABSTRAK

Penelitian ini bertujuan untuk membandingkan metode *Classification and Regression Trees* (CART) dan *Naïve Bayes Classification* (NBC) dalam mengklasifikasikan status gizi balita di Kelurahan Pagesangan Barat dengan

* Corresponding author.

Alamat e-mail: nurulinsan5@yahoo.com

melihat nilai akurasi dan juga mengetahui variabel yang berpengaruh terhadap pengklasifikasian status gizi balita. Data yang digunakan dalam penelitian ini yaitu balita yang datang ke posyandu Kelurahan Pagesangan Barat pada bulan Mei 2019, dengan variabel prediktor yang digunakan yaitu jenis kelamin, umur, berat badan, status pekerjaan ibu, tingkat pendidikan ibu, jumlah anak dan penghasilan orang tua. Dari hasil penelitian menunjukkan bahwa, *Naïve Bayes Classification* (NBC) lebih baik dalam mengklasifikasikan status gizi balita Kelurahan Pagesangan Barat dari pada *Classification and Regression Trees* (CART). Hal ini dapat diketahui dari nilai akurasi yang diperoleh dengan tiga perbandingan data *training* dan data *testing* yang digunakan. Pada perbandingan 90% data *training* : 10% data *testing*, diperoleh nilai akurasi 90% untuk NBC dan 85% untuk CART, pada perbandingan 80% data *training* : 20% data *testing*, diperoleh nilai akurasi 82.5% untuk NBC dan 80% untuk CART sedangkan pada perbandingan 70% data *training* : 30% data *testing*, diperoleh nilai akurasi 72% untuk NBC dan 70% untuk CART. Dari penelitian ini juga diketahui variabel yang berpengaruh terhadap pengklasifikasian status gizi balita Kelurahan Pagesangan Barat yaitu umur, jenis kelamin, berat badan dan penghasilan orang tua.

Kata Kunci: Akurasi, *Classification and Regression Trees* (CART), Klasifikasi, *Naïve Bayes Classification* (NBC), Status Gizi Balita.

Diserahkan: 04-05-2020; Diterima: 15-05-2020;

Doi: <https://doi.org/10.29303/emj.v3i1.68>

1. Pendahuluan

Status gizi dapat diukur dengan indeks berat badan per umur (BB/U), tinggi badan per umur (TB/U), dan berat badan per tinggi badan (BB/TB). Gizi kurang dan gizi buruk merupakan status gizi yang didasarkan pada indeks berat badan per umur (BB/U). Berdasarkan Pemantauan Status Gizi (PSG) tahun 2018 yang dilaksanakan oleh Kementerian Kesehatan, diketahui bahwa persentase gizi buruk di Nusa Tenggara Barat meningkat dari 4.3% pada tahun 2017 menjadi 5.9% pada tahun 2018, sedangkan persentase gizi kurang dari 18.3% pada tahun 2017 menjadi 20.5% pada tahun 2018 (Kemenkes, 2019).

Beberapa metode yang dapat digunakan dalam pengklasifikasian dari suatu objek penelitian diantaranya *Classification and Regression Tress* (CART), *K-Nearest Neighbor*, *Naïve Bayes Classification* (NBC), *Neural Network* dan *Support Vector Machine*. Namun pada penelitian ini peneliti membandingkan metode CART dan NBC, karena tidak ada asumsi yang harus dipenuhi oleh kedua metode tersebut. CART merupakan metode yang dapat diterapkan pada data dengan ukuran obyek dan variabel yang besar. Jika variabel terikatnya bersifat kontinu maka pohon yang dihasilkan disebut pohon regresi (*regression tree*), namun jika variabel terikatnya adalah kategorik maka pohon yang dihasilkan disebut dengan pohon klasifikasi (*classification trees*).

NBC merupakan salah satu teknik klasifikasi. NBC adalah sebuah metode penggolongan berdasarkan probabilitas sederhana dan dirancang untuk dipergunakan dengan asumsi bahwa antar satu kelas dengan kelas yang lain tidak saling tergantung (independen). Pada klasifikasi NBC, proses pembelajaran lebih ditekankan pada mengestimasi probabilitas. Keuntungan dari pendekatan ini yaitu

pengklasifikasian akan mendapatkan nilai error yang lebih kecil ketika data set berjumlah besar (Berry dan Brown, 2006).

2. Landasan Teori

2.1. Klasifikasi

Klasifikasi adalah sebuah proses untuk mencari model atau fungsi yang menjelaskan dan membedakan kelas atau konsep dari data, dengan tujuan untuk menggunakan model dan melakukan prediksi dari kelas suatu objek dimana tidak diketahui label dari kelas tersebut. Model yang ada berasal dari analisis kumpulan *training* data (objek data dimana label kelas diketahui) (Han dan Kamber, 2006).

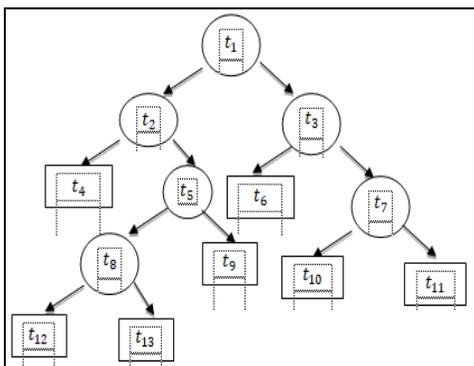
2.2. *Classification and Regression Trees* (CART)

Classification and Regression Trees (CART) adalah salah satu metode nonparametrik yang digunakan untuk analisis klasifikasi yaitu dengan teknik keputusan baik untuk variabel respon kategorik maupun kontinu. Apabila variabel respon berskala kontinu maka CART menghasilkan pohon regresi (*regression trees*), sedangkan apabila variabel respon berskala kategorik, maka CART menghasilkan pohon klasifikasi (*classification trees*). Tujuan utama CART adalah untuk mendapatkan suatu kelompok data yang akurat sebagai pencari dari suatu pengklasifikasian dan menentukan prediksi struktur data (Pratiwi dan Zain, 2014).

2.3. Pohon CART

Pohon keputusan dibentuk dengan menggunakan algoritma penyekatan rekursif secara biner (*binary recursive partitioning*). Pemilahan dilakukan dengan

memilah data menjadi dua kelompok, yaitu kelompok yang masuk simpul kiri dan yang masuk simpul kanan. Pemilahan dilakukan pada tiap simpul sampai didapatkan suatu simpul terminal atau akhir. Variabel yang memilah pada simpul utama adalah variabel terpenting dalam menduga kelas dari amatan (Rahmi, dkk., 2017).



Gambar 1.1 – Struktur Pohon Classification and Regression Trees (CART)

2.4. Pembentukan Pohon Klasifikasi

2.4.1. Proses Pemilahan Pemilah

Proses pemilahan dimulai dari simpul utama yang terdiri dari data yang akan dipilah. Setiap pemilahan hanya bergantung pada nilai yang berasal dari satu variabel independen. Untuk variabel independen kontinu X_j dengan ruang sampel berukuran n dan terdapat n nilai amatan sampel yang berbeda, maka akan terdapat $n-1$ pemilahan yang berbeda. Sedangkan untuk X_j adalah variabel kategori nominal bertaraf L , maka akan diperoleh pemilahan sebanyak $2^{L-1}-1$. Tetapi jika variabel X_j adalah kategori ordinal maka akan diperoleh $L-1$ pemilahan yang mungkin (Hartati, dkk., 2012).

Proses pemilahan pada masing-masing simpul induk didasarkan pada *goodness of split criterion* (kriteria pemilahan terbaik). Kriteria pemilahan terbaik ini dibentuk berdasarkan fungsi *impurity* (fungsi keheterogenan) yaitu untuk mengukur tingkat keheterogenan suatu kelas dari suatu simpul tertentu dalam pohon klasifikasi. Fungsi keheterogenan yang dapat digunakan adalah indeks Gini (Ratnaningrum, dkk., 2016).

Adapun rumus dari indeks Gini yang dapat digunakan adalah sebagai berikut (Tanjung dan Kartiko, 2017).

$$i(t) = 1 - \sum_j P^2(j|t) \quad (2.1)$$

dengan :

$i(t)$ = Indeks Gini

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

$P(k|t)$ = proporsi kelas k pada simpul t

Goodness of Split merupakan suatu evaluasi pemilahan oleh s pada simpul t . Pemilah terbaik adalah pemilah yang memiliki nilai *Goodness of Split* tertinggi. *Goodness of Split* didefinisikan sebagai penurunan keheterogenan yaitu sebagai berikut (Hartati, dkk., 2012).

$$\varphi(s, t) = \Delta i(s, t) = i(t) - P_L i(t_L) - P_R i(t_R) \quad (2.2)$$

dengan :

$\varphi(s, t)$ = kriteria *Goodness of Split*

$i(t)$ = indeks Gini

$P_L i(t_L)$ = proporsi pengamatan dari simpul t menuju simpul kiri

$P_R i(t_R)$ = proporsi pengamatan dari simpul t menuju simpul kanan

2.4.2. Penentuan Simpul Terminal

Simpul t dapat dijadikan simpul terminal jika tidak terdapat penurunan keheterogenan yang berarti pada pemilahan, hanya terdapat satu pengamatan ($n=1$) pada tiap simpul anak, atau adanya batasan minimum n serta adanya batasan jumlah level atau tingkat kedalaman pohon maksimal (Breiman, dkk., 1993).

2.4.3. Proses Pelabelan Kelas

Penandaan label kelas pada simpul terminal dilakukan berdasarkan aturan jumlah terbanyak, yaitu (Hartati, dkk., 2012).

$$\begin{aligned} p(j_0|t) &= \max_j \{p(j|t)\} \\ &= \max_j \frac{N_j(t)}{N(t)} \end{aligned} \quad (2.3)$$

dengan,

$p(j|t)$ = proporsi kelas j pada simpul t

$N_j(t)$ = jumlah pengamatan kelas j pada simpul t

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

2.4.4. Proses Pemangkasan Pohon Klasifikasi (Pruning)

Pohon klasifikasi maksimal yang terbentuk dimungkinkan berukuran sangat besar. Semakin banyak pemilahan yang dilakukan maka tingkat akurasi semakin tinggi, tetapi dengan ukuran yang sangat besar akan sulit dipahami sehingga menyebabkan *overfitting* (pencocokan nilai yang sangat kompleks) untuk data baru. Masalah tersebut diatasi dengan pemangkasan pada pohon klasifikasi maksimal untuk mendapatkan pohon klasifikasi yang optimal (Lewis, 2000).

Ukuran pemangkasan yang digunakan untuk memperoleh ukuran pohon yang layak yaitu :

$$R(T) = \sum_{t \in T} r(t)p(t) \quad (2.4)$$

dengan :

$R(t)$ = probabilitas terjadinya kesalahan klasifikasi yang disebabkan oleh pohon klasifikasi yang terbentuk

$p(t)$ = proporsi amatan yang masuk dalam simpul t

$r(t)$ = probabilitas terjadinya kesalahan klasifikasi di dalam sebuah simpul t tertentu yang didefinisikan sebagai berikut :

$$r(t) = 1 - \max_j P(j|t) \quad (2.5)$$

Langkah awal proses pemangkasan dilakukan terhadap T_t yaitu subpohon dari T_{max} dengan mengambil t_L yang merupakan simpul anak kiri dan t_R yang merupakan simpul anak kanan hasil dari pemilahan simpul induk t . Jika diperoleh dua simpul anak dan simpul induk yang memenuhi persamaan

$$R(t) = R(t_L) + R(t_R) \quad (2.6)$$

dengan :

$R(t_L)$ = probabilitas terjadinya kesalahan klasifikasi yang disebabkan oleh pohon klasifikasi yang terbentuk pada simpul kiri

$R(t_R)$ = probabilitas terjadinya kesalahan klasifikasi yang disebabkan oleh pohon klasifikasi yang terbentuk pada simpul kanan

maka simpul anak t_L dan t_R dipangkas. Proses ini diulang sampai tidak ada lagi pemangkasan yang mungkin (Breiman, dkk., 1993).

2.5. Naïve Bayes Classification (NBC)

Klasifikasi *Naïve Bayes* adalah klasifikasi berdasarkan teorema Bayes dan digunakan untuk

menghitung probabilitas tiap kelas dengan asumsi bahwa antar satu kelas dengan kelas yang lain tidak saling tergantung (*independen*). Pada metode ini, semua atribut akan memberikan kontribusinya dalam pengambilan keputusan, dengan bobot atribut yang sama penting dan setiap atribut saling bebas satu sama lain (Kusumadewi, 2009).

Persamaan dari teorema Bayes adalah sebagai berikut:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (2.7)$$

dengan :

X = Data dengan kelas yang belum diketahui

H = Hipotesis data merupakan suatu kelas spesifik

$P(X)$ = Probabilitas X

$P(H)$ = Probabilitas hipotesis H (*prior probability*)

$P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi X (*posterior probability*)

$P(X|H)$ = Probabilitas X berdasarkan kondisi pada hipotesis H

Untuk menjelaskan teorema *Naïve Bayes*, perlu diketahui bahwa klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, teorema Bayes di atas disesuaikan sebagai berikut :

$$P(C|F_1, \dots, F_n) = \frac{P(C)P(F_1, \dots, F_n|C)}{P(F_1, \dots, F_n)} \quad (2.8)$$

dengan variabel C merepresentasikan kelas, sementara variabel F_1, \dots, F_n merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (*posterior*) adalah peluang masuknya sampel C (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas C (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga *evidence*). Karena itu, rumus di atas dapat pula ditulis secara sederhana sebagai berikut :

$$Posterior = \frac{Prior \times likelihood}{evidence} \quad (2.9)$$

Nilai *evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut nantinya

akan dibandingkan dengan nilai-nilai posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan $P(C|F_1, \dots, F_n)$ menggunakan aturan perkalian sebagai berikut :

$$\begin{aligned}
 P(C|F_1, \dots, F_n) &= P(C) P(F_1, \dots, F_n|C) \quad (2.10) \\
 &= P(C)P(F_1|C)P(F_2, \dots, F_n|C, F_1) \\
 &= P(C)P(F_1|C)P(F_2|C, F_1) \\
 &\quad P(F_3, \dots, F_n|C, F_1, F_2) \\
 &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2) \\
 &\quad P(F_4, \dots, F_n|C, F_1, F_2, F_3) \\
 &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2) \dots \\
 &\quad P(F_n|C, F_1, F_2, F_3, \dots, F_{n-1})
 \end{aligned}$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleks faktor-faktor syarat yang mempengaruhi nilai probabilitas yang hampir mustahil untuk dianalisa satu persatu. Akibatnya perhitungan tersebut menjadi sulit untuk dilakukan. Di sinilah digunakan asumsi independensi yang sangat kuat (naif), bahwa masing-masing petunjuk (F_1, F_2, \dots, F_n) saling bebas (independen satu sama lain). Dengan asumsi tersebut maka berlaku suatu persamaan sebagai berikut :

$$\begin{aligned}
 P(F_i|F_j) &= \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} \\
 &= P(F_i) \quad (2.11)
 \end{aligned}$$

Untuk $i \neq j$, sehingga,

$$P(F_i|C, F_j) = P(F_i|C) \quad (2.12)$$

Dari persamaan di atas dapat disimpulkan bahwa asumsi independensi naif tersebut membuat syarat peluang menjadi sederhana, sehingga perhitungan menjadi mungkin untuk dilakukan. Selanjutnya, penjabaran $P(C|F_1, \dots, F_n)$ dapat disederhanakan menjadi

$$\begin{aligned}
 P(C|F_1, \dots, F_n) &= P(C)P(F_1|C)P(F_2|C) \dots P(F_n|C) \\
 &= P(C) \prod_{i=1}^n P(F_i|C) \quad (2.13)
 \end{aligned}$$

Persamaan 2.13 merupakan model dari teorema *Naïve Bayes Classification* yang selanjutnya akan digunakan dalam proses klasifikasi (Bustami, 2014).

2.6. Ketepatan Klasifikasi

Apparent Error Rate (APER) merupakan suatu ukuran yang digunakan untuk menyatakan nilai proporsi sampel yang salah diklasifikasikan. Jika n_1 merupakan jumlah anggota observasi yang diklasifikasikan $Y = 1$ dan n_2 merupakan jumlah anggota observasi yang diklasifikasikan $Y = 2$, maka bentuk matriks konfusinya yaitu seperti pada Tabel 2.1 (Johnson,2007).

Tabel 2.1 – Matriks Konfusi

Observasi	Prediksi	
	Kelas 1	Kelas 2
Kelas 1	n_{11}	n_{12}
Kelas 2	n_{21}	n_{22}

Berikut diberikan persamaan untuk menghitung nilai APPER.

$$APER = \frac{n_{12} + n_{21}}{n_1 + n_2} \quad (2.14)$$

dengan :

- n_{11} = jumlah amatan kelas 1 yang tepat diklasifikasikan
- n_{12} = jumlah amatan kelas 1 yang salah diklasifikasikan sebagai kelas 2
- n_{22} = jumlah amatan kelas 2 yang tepat diklasifikasikan sebagai kelas 2
- n_{21} = jumlah amatan kelas 2 yang salah diklasifikasikan sebagai kelas 1.

Jadi, untuk menghitung nilai ketepatan klasifikasi digunakan $1 - APER$.

3. Metode Penelitian

Data yang digunakan dalam penelitian ini adalah data primer yang diperoleh melalui survey yang dilakukan pada bulan Mei 2019 di enam posyandu yang ada di kelurahan Pagesangan Barat, yaitu posyandu Gubuk Mamben, posyandu Timbrah, posyandu Kekalik Kebon, posyandu Posinggahan, posyandu Saren dan posyandu Kekalik Baru. Alat survey yang digunakan yaitu wawancara dan observasi. Sampel dalam penelitian ini yaitu semua balita yang datang posyandu pada bulan Mei 2019 di enam posyandu yang ada di kelurahan Pagesangan Barat dengan jumlah 200 data balita. Pada penelitian ini alat yang digunakan untuk analisis data yaitu *program SPSS 16* dan *R 3.6.0*. Dalam penelitian ini variabel yang digunakan adalah variabel respon (Y) yang bersifat kategorik dan tujuh variabel prediktor (X) yang bersifat kategorik.

Berikut diberikan variabel penelitian yang digunakan.

Tabel 3.1 – Variabel Penelitian

Variabel	Nama Variabel	Keterangan	Kategori
Y	Status Gizi Balita	1: Status Gizi Lebih 2: Status Gizi Baik 3: Status Gizi Kurang 4: Status Gizi Buruk	
X_1	Jenis Kelamin	1: Laki-laki 2: Perempuan	Nominal
X_2	Umur	1: 0-5 bulan 2: 6-11 bulan 3: 12-23 bulan 4: 24-59 bulan	Ordinal
X_3	Berat Badan	1: 3-8 Kg 2: 9-14 Kg 3: 15 - 20 Kg	Ordinal
X_4	Status Pekerjaan Ibu	1: Bekerja 2: Tidak Bekerja	Nominal
X_5	Tingkat Pendidikan Ibu	1: Rendah (\leq SMP) 2: Menengah (SMA / SMK) 3: Tinggi ($>$ SMA / SMK)	Ordinal
X_6	Banyak Anak	1: ≤ 2 2: > 2	Ordinal
X_7	Penghasilan Orang Tua	1: Rendah ($<$ 1.500.000) 2: Sedang (1.500.000 – 2.400.000) 3: Tinggi (2.500.000 – 3.500.000) 4: Sangat Tinggi ($>$ 3.500.000)	Ordinal

Penelitian ini terdiri dari beberapa tahap yaitu:

1. Membagi data menjadi data *training* dan data *testing*.

2. Analisis CART menggunakan data *training*, dengan langkah-langkah sebagai berikut:
 - a. Menentukan kemungkinan pemilah
 - b. Menentukan pemilah terbaik berdasarkan *Goodness of split*
 - c. Menentukan simpul terminal
 - d. Pembentukan pohon klasifikasi maksimal
 - e. Melakukan pemangkasan pohon klasifikasi maksimal
 - f. Menentukan pohon klasifikasi optimal
 - g. Menentukan ketepatan pohon klasifikasi optimal dengan data *testing*
3. Analisis NBC menggunakan data *training*, dengan langkah-langkah sebagai berikut :
 - a. Menghitung probabilitas prior ($P(Y)$)
 - b. Menghitung probabilitas atribut terhadap masing-masing kelas ($P(X_j|Y)$).
 - c. Menginput data *testing*
 - d. Menghitung perkalian probabilitas prior dengan probabilitas atribut ada masing-masing kelas ($P(Y) \times P(X_j|Y)$).
 - e. Mencari nilai maksimum ($P(Y) \times P(X_j|Y)$).
 - f. Menentukan ketepatan klasifikasi dengan data *training*.
4. Membandingkan ketepatan klasifikasi CART dan NBC.

4. Hasil dan Pembahasan

4.1. Pembagian Data

Dari jumlah keseluruhan data balita yaitu sebanyak 200 data, ada tiga perbandingan data *training* dan data *testing* yang digunakan dalam penelitian ini, yaitu perbandingan yang pertama 180 data (90%) sebagai data *training* dan 20 data (10%) sebagai data *testing*. Perbandingan yang kedua yaitu 160 data (80%) sebagai data *training* dan 40 data (20%) sebagai data *testing* dan perbandingan yang ketiga yaitu 140 data (70%) sebagai data *training* dan 60 data (30%) sebagai data *testing*.

4.2. Analisis Classification And Regression Trees (CART)

Perbandingan data *training* dan data *testing* yang digunakan pada perhitungan ini yaitu 90% data *training* (180 data) dan 10% data *testing* (20 data), dengan cara yang sama untuk perbandingan data lainnya.

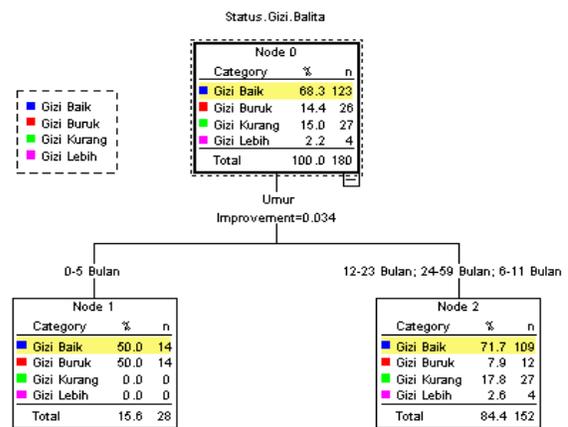
4.2.1. Proses Pemilahan Simpul

Proses pemilahan dimulai dari simpul utama yang terdiri dari data yang akan dipilah. Pemilahan dilakukan untuk memilah data menjadi dua kelompok yaitu kelompok yang masuk simpul kiri dan yang masuk simpul kanan. Pemilah terbaik adalah pemilah yang memiliki sifat heterogen paling tinggi. Ukuran sifat heterogen ini diukur menggunakan indeks gini, sedangkan untuk mengetahui pemilah terbaik dilihat dari nilai *Goodness of Split*. Nilai *Goodness of Split* pada semua kemungkinan pemilah pada masing-masing variabel dapat dilihat pada Tabel 4.1.

Tabel 4.1 – Goodness of Split Masing - Masing Variabel

No	Pemilah Simpul Kiri	Simpul Kanan	Goodness of Split
<i>Variabel Jenis Kelamin</i>			
1.	Laki-laki	Perempuan	0.032
<i>Variabel Umur</i>			
1.	0-5 bulan	6-11 bulan, 12-23 bulan, 24-59 bulan	0.034
2.	0- 5 bulan, 6-11 bulan	12-23 bulan, 24-59 bulan	0.017
3.	0- 5 bulan, 6-11 bulan, 12-23 bulan	24-59 bulan	0.009
<i>Variabel Berat Badan</i>			
1.	3-8 Kg	9-14 Kg, 15-20 Kg	0.022
2.	3-8 Kg, 9-14 Kg	15-20 Kg	0.011
<i>Variabel Status Pekerjaan Ibu</i>			
1.	Bekerja	Tidak Bekerja	0.006
<i>Variabel Tingkat Pendidikan Ibu</i>			
1.	Rendah	Menengah, Tinggi	0.004
2.	Rendah, Menengah	Tinggi	0.007
<i>Variabel Jumlah Anak</i>			
1.	≤2	>2	0.004
<i>Variabel Penghasilan Orang Tua</i>			
1.	Rendah	Sedang, Tinggi, Sangat Tinggi	0.006
2.	Rendah, Sedang	Sangat Tinggi, Tinggi	0.003
3.	Rendah, Sedang, Tinggi	Sangat Tinggi	0.002

Berdasarkan Tabel 4.1 pemilah terbaik untuk simpul 0 atau simpul akar adalah variabel umur dengan kriteria pemilahan 0-5 bulan pada simpul kiri (simpul 1) dan 6-11 bulan, 12-23 bulan, 24-59 bulan pada simpul kanan (simpul 2), sehingga dalam penelitian ini variabel umur merupakan variabel yang paling berpengaruh terhadap klasifikasi status gizi balita karena memiliki nilai *Goodness of Split* tertinggi dari variabel lainnya. Artinya variabel umur mampu menurunkan tingkat keheterogenan paling besar. Proses pemilahan dapat dilihat pada Gambar 4.1.



Gambar 4.1 – Pemilah Utama

4.2.2. Proses Pelabelan Kelas

Proses pelabelan kelas pada simpul-simpul yang terbentuk berdasarkan aturan jumlah anggota kelas terbanyak yaitu jika $P(j_0|t) = \max_j P(j|t)$ maka $j_0 = j$ dengan $j =$ gizi lebih, gizi baik, gizi kurang dan gizi buruk. Sebagai contoh yaitu simpul 2 pada Gambar 4.1

$$P(\text{Gizi Lebih}|\text{simpul 2}) = \frac{4}{152} = 0.026$$

$$P(\text{Gizi Baik}|\text{simpul 2}) = \frac{109}{152} = 0.717$$

$$P(\text{Gizi Kurang}|\text{simpul 2}) = \frac{27}{152} = 0.178$$

$$P(\text{Gizi Buruk}|\text{simpul 2}) = \frac{12}{152} = 0.079$$

Jadi, simpul 2 diberi label kelas Gizi Baik karena proporsi kelas Gizi Baik lebih besar dari pada proporsi kelas lainnya.

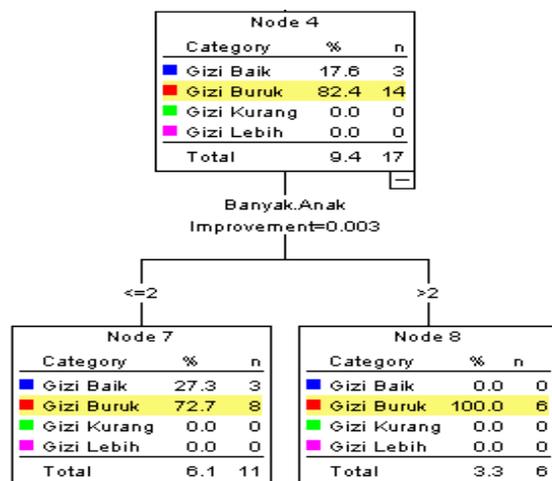
4.2.3. Pohon Klasifikasi Maksimal

Pada pohon klasifikasi maksimal yang terbentuk mempunyai kedalaman pohon 5, yaitu kedalaman

maksimal yang terbentuk berdasarkan data *training* pada penelitian ini. Sedangkan minimum pada simpul induknya adalah 2 dan minimum pada simpul anaknya adalah 1. Hasil pembentukan model berupa pohon klasifikasi dengan simpul yang terbentuk sebanyak 37, simpul terminal 19 dengan kedalaman pohon 5.

4.2.4. Proses Pemangkasan Pohon Klasifikasi (Pruning)

Setelah memperoleh pohon klasifikasi maksimal, langkah berikutnya adalah melihat apakah pohon klasifikasi maksimal perlu dilakukan proses pemangkasan (*pruning*) atau tidak. Tahap pemangkasan bertujuan untuk menghindari adanya kasus *overfitting* (pencocokan nilai yang sangat kompleks) yang disebabkan oleh pohon klasifikasi maksimal yang terbentuk dimungkinkan berukuran sangat besar. Proses pemangkasan pohon klasifikasi maksimal dimulai dengan mengambil t_L yang merupakan simpul kiri dan t_R yang merupakan simpul kanan dari T_{max} yang dihasilkan dari simpul induk t . Jika diperoleh dua simpul anak dan simpul induk yang memenuhi persamaan $R(t) = R(t_L) + R(t_R)$, maka t_L dan t_R dipangkas. Proses diulang sampai tidak ada lagi pemangkasan yang mungkin dilakukan. Sebagai contoh simpul yang dipangkas yaitu simpul 4.



Gambar 4. 2 –Pohon Klasifikasi Maksimal Yang Dipangkas

Pada simpul 4 diperoleh :

$$r(t) = 1 - \max P(j|t)$$

$$r(\text{simpul } 4) = 1 - \max P(j|\text{simpul } 4)$$

$$= 1 - 0.824 = 0.176$$

$$P(t) = \frac{N(t)}{N}$$

$$P(\text{simpul } 4) = \frac{17}{180} = 0.094$$

$$R(\text{simpul } 4) = r(\text{simpul } 4) \times P(\text{simpul } 4)$$

$$= 0.176 \times 0.094 = 0.016$$

Selanjutnya dihitung nilai $R(t_L)$ dan $R(t_R)$ pada simpul anak, yaitu simpul 7 dan simpul 8.

Pada simpul 7 diperoleh :

$$r(\text{simpul } 7) = 1 - \max P(j|\text{simpul } 7)$$

$$= 1 - 0.727 = 0.273$$

$$P(t) = \frac{N(t)}{N}$$

$$P(\text{simpul } 7) = \frac{11}{180} = 0.061$$

$$R(\text{simpul } 7) = r(\text{simpul } 7) \times P(\text{simpul } 7)$$

$$= 0.273 \times 0.061 = 0.016$$

Pada simpul 8 diperoleh :

$$r(t) = 1 - \max P(j|t)$$

$$r(\text{simpul } 8) = 1 - \max P(j|\text{simpul } 8)$$

$$= 1 - 1 = 0$$

$$P(\text{simpul } 8) = \frac{6}{180} = 0.033$$

$$R(\text{simpul } 8) = r(\text{simpul } 8) \times P(\text{simpul } 8)$$

$$= 0 \times 0.033 = 0$$

$$R(\text{simpul } 7) + R(\text{simpul } 8) = 0.016 + 0$$

$$= 0.016$$

$$= R(\text{simpul } 4)$$

Karena $R(\text{simpul } 4) = R(\text{simpul } 7) + R(\text{simpul } 8)$ terpenuhi, sehingga dilakukan pemangkasan pada simpul tersebut.

4.2.5. Pohon Klasifikasi Optimal

Pohon klasifikasi optimal yang terbentuk dari proses pemangkasan menghasilkan 13 simpul yang terdiri atas 1 simpul akar dan 7 simpul terminal dengan kedalaman pohon sebesar 5. Variabel prediktor yang

masuk ke dalam pohon optimal yaitu umur, jenis kelamin, berat badan dan penghasilan orang tua.

4.2.6. Prediksi Pohon Klasifikasi

Sebagai contoh, memprediksi kelas status gizi pada data *testing* pertama menggunakan pohon klasifikasi optimal yang terbentuk dengan kriteria jenis kelamin = perempuan, umur = 12-23 bulan, berat badan = 3-8 Kg, status pekerjaan ibu = tidak bekerja, tingkat pendidikan ibu = rendah, banyak anak ≤ 2 dan penghasilan orang tua = rendah yang diprediksi masuk ke dalam kelas **gizi kurang**.

4.2.7. Ketepatan Klasifikasi CART

Hasil prediksi yang diperoleh selanjutnya digunakan untuk menguji ketepatan klasifikasi pada penerapan metode *Classification And Regression Trees* (CART). Uji ketepatan klasifikasi dilakukan menggunakan matriks konfusi pada Tabel 4.2.

Tabel 4.2 – Matriks Konfusi Ketepatan Klasifikasi CART

Aktual	Prediksi				Total
	Gizi Lebih	Gizi Baik	Gizi Kurang	Gizi Buruk	
Gizi Lebih	0	0	0	0	0
Gizi Baik	0	13	1	0	14
Gizi Kurang	0	1	3	0	4
Gizi Buruk	0	1	0	1	2
Total	0	15	4	1	20

Ketepatan Klasifikasi = $1 - APER$

$$APER = \frac{n_{12}+n_{13}+n_{14}+n_{21}+n_{23}+n_{24}+n_{31}+n_{32}+n_{34}+n_{41}+n_{42}+n_{43}}{n_1+n_2+n_3+n_4}$$

$$= \frac{0+0+0+0+1+0+0+1+0+0+1+0}{0+15+4+1} = 0.15$$

Ketepatan Klasifikasi = $1 - APER = 0.85 = 85\%$

Jadi, dengan diperolehnya nilai APER sebesar 0.15, maka diperoleh ketepatan klasifikasinya 0.85 atau sebesar 85%.

4.3. Naïve Bayes Classification (NBC)

4.3.1. Probabilitas Status Gizi (Y)/ Probabilitas Prior (P(Y))

Dari 180 data *training* yang digunakan, diketahui bahwa terdapat 4 data balita dengan status gizi lebih (Y_0), 123 data balita dengan status gizi baik (Y_1), 27 data balita dengan status gizi kurang (Y_2) dan 26 data balita dengan status gizi buruk (Y_3). Probabilitas prior untuk kemungkinan kelas gizi lebih dapat dihitung dengan cara membagi jumlah status gizi lebih (Y_0) yang muncul pada *training* dengan total data *training*. Perhitungan probabilitas untuk setiap kelas dapat dilihat pada Tabel 4.3

Tabel 4.3 – Probabilitas Status Gizi Balita

Kategori status gizi	Jumlah balita	Probabilitas kategori status gizi
Lebih	4	0.022
Baik	123	0.683
Kurang	27	0.150
Buruk	26	0.144
Total	180	1

4.3.2. Probabilitas Atribut Terhadap Masing-masing Kelas ($P(X_j|Y)$)

Adapun nilai probabilitas setiap atribut terhadap masing-masing kelas status gizi adalah sebagai berikut.

a. Probabilitas Status Gizi dengan Jenis Kelamin

Probabilitas setiap kelas status gizi balita dengan variabel jenis kelamin dapat dilihat pada Tabel 4.4.

Tabel 4.4 – Probabilitas Status Gizi dengan Jenis Kelamin

Jenis Kelamin	Probabilitas status gizi			
	Lebih	Baik	Kurang	Buruk
Laki-laki	0.25	0.577	0.556	0.038
Perempuan	0.75	0.423	0.444	0.962
Jumlah	1	1	1	1

b. Probabilitas Status Gizi dengan Umur

Probabilitas setiap kelas status gizi balita dengan variabel umur dapat dilihat pada Tabel 4.5.

Tabel 4.5 – Probabilitas Status Gizi dengan Umur

Umur	Probabilitas status gizi			
	Lebih	Baik	Kurang	Buruk
0-5 bulan	0	0.114	0	0.539
6-11 bulan	0	0.195	0.148	0.077
12-23 bulan	0.25	0.219	0.370	0.115
24-59 bulan	0.75	0.472	0.482	0.269
Jumlah	1	1	1	1

c. Probabilitas Status Gizi dengan Berat Badan

Probabilitas setiap kelas status gizi balita dengan variabel berat badan dapat dilihat pada Tabel 4.6.

Tabel 4.6 – Probabilitas Status Gizi dengan Berat Badan

Berat badan	Probabilitas status gizi			
	Lebih	Baik	Kurang	Buruk
3-8 Kg	0	0.309	0.519	0.692
9-14 Kg	0.25	0.577	0.481	0.308
15-20 Kg	0.75	0.114	0	0
Jumlah	1	1	1	1

d. Probabilitas Status Gizi dengan Status Pekerjaan Ibu

Probabilitas setiap kelas status gizi balita dengan variabel status pekerjaan ibu dapat dilihat pada Tabel 4.7.

Tabel 4.7 – Probabilitas Status Gizi dengan Status Pekerjaan Ibu

Status pekerjaan ibu	Probabilitas status gizi			
	Lebih	Baik	Kurang	Buruk
Bekerja	0.25	0.244	0.074	0.192
Tidak bekerja	0.75	0.756	0.926	0.808
Jumlah	1	1	1	1

e. Probabilitas Status Gizi dengan Tingkat Pendidikan Ibu

Probabilitas setiap kelas status gizi balita dengan variabel tingkat pendidikan ibu dapat dilihat pada Tabel 4.8.

Tabel 4.8 – Probabilitas Status Gizi dengan Tingkat Pendidikan Ibu

Tingkat Pendidikan ibu	Probabilitas status gizi			
	Lebih	Baik	Kurang	Buruk
Rendah	0	0.480	0.667	0.5
Menengah	0.75	0.341	0.296	0.423
Tinggi	0.25	0.179	0.037	0.077
Jumlah	1	1	1	1

f. Probabilitas Status Gizi dengan Banyak Anak

Probabilitas setiap kelas status gizi balita dengan variabel banyak anak dapat dilihat pada Tabel 4.9.

Tabel 4.9 – Probabilitas Status Gizi dengan Banyak Anak

Jumlah anak	Probabilitas status gizi			
	Lebih	Baik	Kurang	Buruk
≤ 2	0.5	0.715	0.630	0.577
> 2	0.5	0.285	0.370	0.423
Jumlah	1	1	1	1

g. Probabilitas Status Gizi dengan Penghasilan Orang Tua

Probabilitas setiap kelas status gizi balita dengan variabel penghasilan orang tua dapat dilihat pada Tabel 4.10.

Tabel 4.10 – Probabilitas Status Gizi dengan Penghasilan Orang Tua

Penghasilan orang tua	Probabilitas status gizi			
	Lebih	Baik	Kurang	Buruk
Rendah	0	0.301	0.519	0.346
Sedang	0	0.333	0.259	0.346
Tinggi	0.25	0.211	0.148	0.192
Sangat tinggi	0.75	0.155	0.074	0.115
Jumlah	1	1	1	1

4.3.3. Simulasi Perhitungan Data Testing dengan Metode Naïve Bayes Classification (NBC)

Berikut ini dilakukan simulasi untuk perhitungan data *testing* pertama :

Jika diketahui balita berjenis kelamin perempuan, umur 12-23 bulan, berat badan 3-8 Kg, status pekerjaan ibunya tidak bekerja, tingkat pendidikan ibunya rendah, dan orang tuanya mempunyai anak ≤ 2 serta penghasilan orang tua rendah.

1. Probabilitas status gizi balita

Probabilitas tiap kelas status gizi balita dapat dilihat pada Tabel 4.3

$$P(Y = \text{Gizi Lebih}) = 0.022$$

$$P(Y = \text{Gizi Baik}) = 0.683$$

$$P(Y = \text{Gizi Kurang}) = 0.15$$

$$P(Y = \text{Gizi Buruk}) = 0.144$$

2. Probabilitas kasus yang sama dengan kelas yang sama

Probabilitas status gizi dengan variabel jenis kelamin perempuan dapat dilihat pada Tabel 4.4.

$$P(\text{Jenis Kelamin} = \text{Perempuan} \mid Y = \text{Gizi Lebih}) = 0.75$$

$$P(\text{Jenis Kelamin} = \text{Perempuan} \mid Y = \text{Gizi Baik}) = 0.423$$

$$P(\text{Jenis Kelamin} = \text{Perempuan} \mid Y = \text{Gizi Kurang}) = 0.444$$

$$P(\text{Jenis Kelamin} = \text{Perempuan} \mid Y = \text{Gizi Buruk}) = 0.962$$

Probabilitas status gizi dengan variabel umur 12-23 bulan dapat dilihat pada Tabel 4.5.

$$P(\text{Umur} = 12 - 23 \text{ bulan} \mid Y = \text{Gizi Lebih}) = 0.25$$

$$P(\text{Umur} = 12 - 23 \text{ bulan} \mid Y = \text{Gizi Baik}) = 0.219$$

$$P(\text{Umur} = 12 - 23 \text{ bulan} \mid Y = \text{Gizi Kurang}) = 0.370$$

$$P(\text{Umur} = 12 - 23 \text{ bulan} \mid Y = \text{Gizi Buruk}) = 0.115$$

Probabilitas status gizi dengan variabel berat badan 3-8 Kg dapat dilihat pada Tabel 4.6.

$$P(\text{Berat Badan} = 3 - 8 \text{ Kg} \mid Y = \text{Gizi Lebih}) = 0$$

$$P(\text{Berat Badan} = 3 - 8 \text{ Kg} \mid Y = \text{Gizi Baik}) = 0.309$$

$$P(\text{Berat Badan} = 3 - 8 \text{ Kg} \mid Y = \text{Gizi Kurang}) = 0.519$$

$$P(\text{Berat Badan} = 3 - 8 \text{ Kg} \mid Y = \text{Gizi Buruk}) = 0.692$$

Probabilitas status gizi dengan variabel status pekerjaan ibu tidak bekerja dapat dilihat pada Tabel 4.7.

$$P(\text{Status Pekerjaan Ibu} = \text{Tidak Bekerja} \mid Y = \text{Gizi Lebih}) = 0.75$$

$$P(\text{Status Pekerjaan Ibu} = \text{Tidak Bekerja} \mid Y = \text{Gizi Baik}) = 0.756$$

$$P(\text{Status Pekerjaan Ibu} = \text{Tidak Bekerja} \mid Y = \text{Gizi Kurang}) = 0.926$$

$$P(\text{Status Pekerjaan Ibu} = \text{Tidak Bekerja} \mid Y = \text{Gizi Buruk}) = 0.808$$

Probabilitas status gizi dengan variabel tingkat pendidikan ibu rendah dapat dilihat pada Tabel 4.8.

$$P(\text{Tingkat Pendidikan Ibu} = \text{Rendah} \mid Y = \text{Gizi Lebih}) = 0$$

$$P(\text{Tingkat Pendidikan Ibu} = \text{Rendah} \mid Y = \text{Gizi Baik}) = 0.480$$

$$P(\text{Tingkat Pendidikan Ibu} = \text{Rendah} \mid Y = \text{Gizi Kurang}) = 0.667$$

$$P(\text{Tingkat Pendidikan Ibu} = \text{Rendah} \mid Y = \text{Gizi Buruk}) = 0.5$$

Probabilitas status gizi dengan variabel banyak anak ≤ 2 dapat dilihat pada Tabel 4.9

$$P(\text{Banyak Anak} \leq 2 \mid Y = \text{Gizi Lebih}) = 0.5$$

$$P(\text{Banyak Anak} \leq 2 \mid Y = \text{Gizi Baik}) = 0.715$$

$$P(\text{Banyak Anak} \leq 2 \mid Y = \text{Gizi Kurang}) = 0.630$$

$$P(\text{Banyak Anak} \leq 2 \mid Y = \text{Gizi Buruk}) = 0.577$$

Probabilitas status gizi dengan variabel penghasilan orang tua dapat dilihat pada Tabel 4.10.

$$P(\text{Penghasilan Orang Tua} = \text{Rendah} \mid Y = \text{Gizi Lebih}) = 0$$

$$P(\text{Penghasilan Orang Tua} = \text{Rendah} \mid Y = \text{Gizi Baik}) = 0.301$$

$$P(\text{Penghasilan Orang Tua} = \text{Rendah} \mid Y = \text{Gizi Kurang}) = 0.519$$

$$P(\text{Penghasilan Orang Tua} = \text{Rendah} \mid Y = \text{Gizi Buruk}) = 0.346$$

$$= 0.444 \times 0.370 \times 0.519 \times 0.926 \times 0.667 \times 0.630 \times 0.519 \times 0.15 = 0.00258$$

3. Kalikan semua hasil variabel Gizi Lebih, Gizi Baik, Gizi Kurang dan Gizi Buruk

$$\begin{aligned} & [P(\text{Jenis Kelamin} = \text{Perempuan} \mid Y = \text{Gizi Lebih})] \\ & \times [P(\text{Umur} = 12 - 23 \text{ bulan} \mid Y = \text{Gizi Lebih})] \\ & \times [P(\text{Berat Badan} = 3 - 8 \text{ Kg} \mid Y = \text{Gizi Lebih})] \\ & \times [P(\text{Status Pekerjaan Ibu} = \text{Tidak Bekerja} \mid Y = \text{Gizi Lebih})] \\ & \times [P(\text{Tingkat Pendidikan Ibu} = \text{Rendah} \mid Y = \text{Gizi Lebih})] \\ & \times [P(\text{Jumlah Anak} \leq 2 \mid Y = \text{Gizi Lebih})] \\ & \times [P(\text{Penghasilan Orang Tua} = \text{Rendah} \mid Y = \text{Gizi Lebih})] \times [P(Y = \text{Gizi Lebih})] \\ & = 0.75 \times 0.25 \times 0 \times 0.75 \times 0 \times 0.5 \times 0 \times 0.022 = 0 \end{aligned}$$

$$\begin{aligned} & [P(\text{Jenis Kelamin} = \text{Perempuan} \mid Y = \text{Gizi Baik})] \\ & \times [P(\text{Umur} = 12 - 23 \text{ bulan} \mid Y = \text{Gizi Baik})] \\ & \times [P(\text{Berat Badan} = 3 - 8 \text{ Kg} \mid Y = \text{Gizi Baik})] \\ & \times [P(\text{Status Pekerjaan Ibu} = \text{Tidak Bekerja} \mid Y = \text{Gizi Baik})] \\ & \times [P(\text{Tingkat Pendidikan Ibu} = \text{Rendah} \mid Y = \text{Gizi Baik})] \\ & \times [P(\text{Jumlah Anak} \leq 2 \mid Y = \text{Gizi Baik})] \\ & \times [P(\text{Penghasilan Orang Tua} = \text{Rendah} \mid Y = \text{Gizi Baik})] \times [P(Y = \text{Gizi Baik})] \\ & = 0.423 \times 0.219 \times 0.309 \times 0.756 \times 0.480 \times 0.715 \times 0.301 \times 0.683 = 0.00153 \end{aligned}$$

$$\begin{aligned} & [P(\text{Jenis Kelamin} = \text{Perempuan} \mid Y = \text{Gizi Kurang})] \\ & \times [P(\text{Umur} = 12 - 23 \text{ bulan} \mid Y = \text{Gizi Kurang})] \\ & \times [P(\text{Berat Badan} = 3 - 8 \text{ Kg} \mid Y = \text{Gizi Kurang})] \\ & \times [P(\text{Status Pekerjaan Ibu} = \text{Tidak Bekerja} \mid Y = \text{Gizi Kurang})] \\ & \times [P(\text{Tingkat Pendidikan Ibu} = \text{Rendah} \mid Y = \text{Gizi Kurang})] \\ & \times [P(\text{Jumlah Anak} \leq 2 \mid Y = \text{Gizi Kurang})] \\ & \times [P(\text{Penghasilan Orang Tua} = \text{Rendah} \mid Y = \text{Gizi Kurang})] \times [P(Y = \text{Gizi Kurang})] \end{aligned}$$

$$\begin{aligned} & [P(\text{Jenis Kelamin} = \text{Perempuan} \mid Y = \text{Gizi Buruk})] \\ & \times [P(\text{Umur} = 12 - 23 \text{ bulan} \mid Y = \text{Gizi Buruk})] \\ & \times [P(\text{Berat Badan} = 3 - 8 \text{ Kg} \mid Y = \text{Gizi Buruk})] \\ & \times [P(\text{Status Pekerjaan Ibu} = \text{Tidak Bekerja} \mid Y = \text{Gizi Buruk})] \\ & \times [P(\text{Tingkat Pendidikan Ibu} = \text{Rendah} \mid Y = \text{Gizi Buruk})] \\ & \times [P(\text{Jumlah Anak} \leq 2 \mid Y = \text{Gizi Buruk})] \\ & \times [P(\text{Penghasilan Orang Tua} = \text{Rendah} \mid Y = \text{Gizi Buruk})] \times [P(Y = \text{Gizi Buruk})] \\ & = 0.962 \times 0.115 \times 0.962 \times 0.808 \times 0.5 \times 0.577 \times 0.346 \times 0.144 = 0.000889 \end{aligned}$$

Berdasarkan peluang status gizi balita gizi lebih, gizi baik, gizi kurang maupun gizi buruk dapat diketahui bahwa peluang gizi kurang lebih besar dibandingkan dengan peluang status gizi lainnya yaitu sebesar 0.00258, sehingga dapat dikatakan bahwa data *testing* pertama diprediksi masuk ke dalam kelas status gizi kurang.

4.3.4. Ketepatan Klasifikasi NBC

Berdasarkan model-model probabilitas yang didapatkan, dilakukan pengujian terhadap data *testing* yaitu 10% dari total keseluruhan data. Data pengujian tersebut sebanyak 20 data balita. Uji ketepatan klasifikasi dilakukan menggunakan matriks konfusi pada Tabel 4.11.

Tabel 4.11 – Matriks Konfusi Ketepatan Klasifikasi NBC

Aktual	Prediksi				Total
	Gizi Lebih	Gizi Baik	Gizi Kurang	Gizi Buruk	
Gizi Lebih	0	0	0	0	0
Gizi Baik	0	14	0	0	14
Gizi Kurang	0	1	3	0	4
Gizi Buruk	0	1	0	1	2
Total	0	16	3	1	20

$$APER = \frac{n_{12}+n_{13}+n_{14}+n_{21}+n_{23}+n_{24}+n_{31}+n_{32}+n_{34}+n_{41}+n_{42}+n_{43}}{n_1+n_2+n_3+n_4}$$

$$= \frac{0+0+0+0+0+0+1+0+0+1+0}{0+16+3+1} = \frac{2}{20} = 0.1$$

Ketepatan Klasifikasi = $1 - APER = 0.9 = 90\%$

Jadi, dengan memperhatikan nilai APER sebesar 0.10, maka ketepatan klasifikasi yang diperoleh adalah 0.90 atau sebesar 90%.

4.4. Perbandingan Klasifikasi Classification And Regression Trees (CART) dan Naïve Bayes Classification (NBC)

Setelah diperoleh nilai akurasi dari metode Classification And Regression Trees (CART) dan Naïve Bayes Classification (NBC), maka dilakukan perbandingan metode berdasarkan nilai akurasi yang diperoleh untuk memilih metode terbaik. Pada Tabel 4.12 berikut diberikan perbandingan akurasi dari kedua metode tersebut.

Tabel 4.12 – Perbandingan Classification And Regression Trees (CART) dan Naïve Bayes Classification (NBC)

Perbandingan Data Training dan Data Testing (%)	Ketepatan Klasifikasi	
	CART	NBC
90 : 10	85%	90%
80 : 20	80%	82.5%
70 : 30	70%	72%

Berdasarkan Tabel 4.12 didapatkan informasi bahwa ketepatan klasifikasi atau akurasi dari Classification And Regression Trees (CART) sebesar 85% dengan perbandingan data training dan data testing yaitu 90%:10%, 80% dengan perbandingan data training dan data testing 80%:20% dan 70% dengan perbandingan data training dan data testing 70%:30%. Ketepatan klasifikasi atau akurasi dari Naïve Bayes Classification (NBC) sebesar 90% dengan perbandingan data training dan data testing 90%:10%, 82.5% dengan perbandingan data training dan data testing 80%:20% dan 72.5% dengan perbandingan data training dan data testing 70%:30%. Berdasarkan akurasi yang didapatkan dari kedua metode tersebut maka dapat dikatakan bahwa Naïve Bayes Classification lebih baik dalam mengklasifikasikan status gizi balita.

Dari perbandingan data training dan data testing yang digunakan dapat diketahui pula bahwa semakin banyak data training yang digunakan maka nilai akurasinya akan semakin baik dan dari penelitian ini dapat diketahui pula bahwa salah satu kelebihan CART dibandingkan NBC yaitu meskipun nilai akurasinya CART lebih kecil tetapi metode ini dapat menentukan variabel yang berpengaruh terhadap pengklasifikasian status gizi balita di Kelurahan Pagesangan Barat, sedangkan NBC tidak dapat menentukan variabel yang berpengaruh terhadap pengklasifikasian status gizi balita.

5. Kesimpulan

Berdasarkan hasil penelitian dan pembahasan dapat disimpulkan bahwa

1. Algoritma Naïve Bayes Classification lebih baik dalam mengklasifikasikan status gizi balita di Kelurahan Pagesangan Barat. Hal ini dikarenakan nilai akurasi dari ketiga perbandingan data training dan data testing menunjukkan Naïve Bayes Classification (NBC) lebih besar dibandingkan Classification and Regression Trees (CART) yaitu pada perbandingan data training dan data testing 90%:10% menghasilkan akurasi 90% untuk Naïve Bayes Classification dan 85% untuk Classification and Regression Trees (CART). Pada perbandingan data training dan data testing 80% : 20% menghasilkan akurasi 82.5% untuk Naïve Bayes Classification dan 80% untuk Classification and Regression Trees (CART). Pada perbandingan data training dan data testing 70%:30% menghasilkan akurasi 72% untuk Naïve Bayes Classification dan 70% untuk Classification and Regression Trees (CART).
2. Variabel yang memberikan pengaruh dalam pengklasifikasian status gizi balita di Kelurahan Pagesangan Barat menggunakan Classification and Regression Trees (CART) adalah variabel umur, jenis kelamin, berat badan, dan penghasilan orang tua, sedangkan metode Naïve Bayes Classification (NBC) tidak dapat menentukan variabel yang berpengaruh dalam pengklasifikasian status gizi balita.

DAFTAR PUSTAKA

- Berry, M. W. dan Browne, M. (2006). *Lecture Notes In Data Mining*. USA: Word Scientific.
- Bustami. (2014). Penerapan Algoritma Naïve Bayes Untuk Mengklasifikasikan Para Nasabah, *Jurnal Informatika*, Vol. 8, No. 1, pp. 885-887.

- Breiman, L., dkk. (1993). *Classification and Regression Trees (CART)*. New York: Chapman And Hall.
- Han, J. dan Kamber, M. (2001). *Data Mining: Concepts and Techniques Second Edition*. San Fransisco: Morgan Kaufmann Publishers.
- Hartati, A., Zain, I., dan Ulama, B. S. S. (2012). Analisis Cart (Classification And Regression Trees) pada Faktor-Faktor yang Mempengaruhi Kepala Rumah Tangga di Jawa Timur Melakukan Urbanisasi. *Jurnal Sains dan Seni ITS*, Vol. 1, No. 1, pp. 100-105.
- Johnson, R. A. dan Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis, Sixth Edition*. USA: Pearson Education.
- Kementerian Kesehatan. (2013). *Situasi Keluarga Berencana di Indonesia*. Jakarta: Kementerian Kesehatan RI.
- Kementerian Kesehatan. (2019). *Data dan Informasi Profil Kesehatan Indoensia 2018*. Jakarta: Kementerian Kesehatan RI.
- Kusumadewi, S. (2009). Klasifikasi Status Gizi Menggunakan Naïve Bayes Classification. *Comm IT*, Vol. 3, No. 1, pp. 6-11.
- Lewis, R. J. (2000). An Introduction T Classification and Regression Trees (CART) Analysis, *Present at The 2000 Annual Meeting of Society for Academic Emergency Medicine of Sans Fransisco, California*.
- Pratiwi, F. E. dan Zain, I. (2014). Klasifikasi Pengangguran Menggunakan CART (Classification and Regression Tree) di Provinsi Sulawesi Utara. *Jurnal Sains dan Seni Pomits*. Vol. 3, No. 1, pp. 55.
- Rahmi, I., Yoza, H., dan Rahmy, H. A. (2017). Telaah Faktor-Faktor yang Mempengaruhi Status Gizi Balita di Kota Padang Berdasarkan Berat Badan Per Tinggi Badan Menggunakan Metode CART. *Jurnal Eksakta*, Vol. 18, No. 2, pp. 87.
- Ratnaningrum, D., Mukid, M. A., dan Wuryandari, T. (2016). Analisis Klasifikasi Nasabah Kredit Menggunakan Bootstrap Aggregating Classification and Regression Trees (Bagging CART). *Jurnal Gaussian*, Vol. 5, No. 1, pp. 82-83.
- Tanjung, R. H. dan Kartiko, K. (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 dan Komputasi*, Vol. 2, No. 2, pp. 78-83.