

# Penerapan *Decision Tree* dan *Neural Network* pada Penentuan Kategori Status Gizi Balita

<sup>1</sup>Elvia Budianita, <sup>2</sup>Okfalisa, <sup>3</sup>Ahmad Ofandi

<sup>1,2,3</sup>Teknik Informatika UIN Sultan Syarif Kasim Riau, Pekanbaru  
Jl. Subrantas Km. 15, Pekanbaru, 0761-562223  
e-mail: <sup>1</sup>elvia.budianita@uin-suska.ac.id, <sup>2</sup>okfalisa@gmail.com

## Abstrak

Permasalahan gizi dapat disebabkan oleh beberapa faktor. Meliputi pengetahuan ibu terhadap gizi anak, tingkat ekonomi, pekerjaan keluarga, bahkan lingkungan sekitar yang kurang baik. Pada penelitian ini, penentuan kategori gizi balita berdasarkan faktor pendidikan ibu, kategori ekonomi keluarga dan pekerjaan ayah dilakukan dengan pendekatan metode data mining *decision tree* C4.5 dan *Neural Network* menggunakan data Hasil Penimbangan Massal Balita Puskesmas Rimbo Data Kecamatan Pangkalan Koto Baru Kabupaten Lima Puluh Kota. Pada C4.5 memberikan hasil pohon keputusan dengan root adalah berat badan dan sebagai leaf adalah umur dan pendidikan ibu. Sedangkan tinggi badan, kategori ekonomi keluarga dan pekerjaan ayah tidak sebagai leaf pada pohon keputusan. Akurasi yang dihasilkan pada metode C4.5 adalah 62.73% untuk kriteria *gain ratio* dan 84,55% untuk kriteria *information gain*. Sedangkan pada *Neural Network* diperoleh hasil akurasi mencapai 97,27%.

**Kata kunci:** C4.5, *Decision Tree*, Kategori Gizi, *Neural Network*

## Abstract

Some researches on nutrition has been done a lot. The results of nutritional problems can be caused by several factors. It includes the mother's knowledge of child nutrition, economic level, families job, dan even poor surroundings environment. In this study, the determination of the nutritional kategori of toddlers is based on maternal education factors, family economic kategori and father's job was carried out by data mining *decision tree* C4.5 and *Neural Network* approach using data results of Mass Weighing of Toddlers Puskesmas Rimbo Data District Pangkalan Koto Baru Lima puluh kota District. The C4.5 data shows the results of the *decision tree* with root is weight and as leaf is the age and education of the mother. As the height, the family's economic kategori and father's job are not as leaf on the *decision tree*. The C4.5 method obtains an accuracy of 62.73% for the *gain ratio* criteria and 84.55% for the *information gain* criteria, while in *neural network* obtained accuracy results reached 97.27%

**Keywords:** C4.5, *Decision Tree*, *Neural Network*, Nutritional kategori

## 1. Pendahuluan

Salah satu upaya untuk menjaga daya tahan tubuh adalah memenuhi kebutuhan gizi dalam tubuh. Namun, permasalahan gizi ini masih saja belum terselesaikan. Bayi usia di bawah 5 tahun (balita) masih mengalami masalah gizi sebesar 17,7% yang dieproleh dari hasil Riset Kesehatan Dasar (Riskesdas) Kementerian Kesehatan 2018. Padahal masa balita merupakan periode emas untuk pertumbuhan fisik, kecerdasan, mental serta emosional anak sehingga mengoptimalkan kualitas hidup anak sangat dibutuhkan. Hal ini dapat dilakukan dengan memenuhi gizi yang baik, kebersihan, pemberian imunisasi, pemberian vitamin A dan pelayanan kesehatan yang bermutu serta kasih sayang dan stimulasi yang memadai pada usia Balita[1].

Beberapa penelitian tentang gizi telah banyak dilakukan. Diantaranya berdasarkan penelitian Putri dkk terdapat faktor yang berhubungan dengan kategori gizi anak balita di wilayah kerja Puskesmas Nanggalo Padang, yaitu hasil analisis bivariat menunjukkan bahwa terdapat hubungan antara pendidikan ibu, pekerjaan ibu, pendapatan keluarga, jumlah anak serta pola asuh ibu dengan kategori gizi anak balita. Sedangkan berdasarkan hasil analisis multivariat faktor pekerjaan ibu merupakan faktor yang paling berhubungan dengan kategori gizi anak balita [2]. Begitu juga dengan penelitian Puspasari dan Andriani kategori gizi balita (Berat Badan/Umur) dipengaruhi oleh pengetahuan ibu tentang gizi dan asupan makan balita meliputi energi, karbohidrat dan protein [3]. Berbeda dengan penelitian yang dilakukan oleh Sholikhah dkk, faktor yang sangat mempengaruhi kategori gizi balita di pedesaan dan perkotaanadanya adalah penyakit infeksi yang menyertai balita tersebut sedangkan faktor jarak kelahiran, pola

pengasuhan gizi, pendidikan ibu dan pekerjaan ibu tidak mempengaruhi kategori gizi pada balita di pedesaan dan perkotaan [4].

Pada penelitian ini, penentuan kategori gizi balita berdasarkan faktor pendidikan ibu, kategori ekonomi keluarga dan pekerjaan ayah dilakukan dengan pendekatan metode data mining decision tree C4.5 dan Neural Network Multi layer Perceptron menggunakan data Hasil Penimbangan Massal Balita Puskesmas Rimbo Data Kecamatan Pangkalan Koto Baru Kabupaten Lima Puluh Kota. Metode C4.5 adalah salah satu metode *decision tree* yang menghasilkan node-node sebagai atribut pengujian. Pada setiap cabang node mewakili output dari pengujian dan setiap *node leaf* (terminal node) menentukan target kelas. Node paling atas dari sebuah pohon disebut *node root* [5]. Berdasarkan node yang terbentuk dapat menunjukkan variabel mana yang terpilih atau memiliki nilai gain tertinggi. Penelitian menggunakan C4.5 juga sudah banyak dilakukan salah satunya adalah penelitian yang dilakukan Abdullah et al melakukan penelitian dengan membandingkan metode *decision tree* yaitu C4.5, C5.0 dan *Weighted Decision Tree* (WDT) pada data medis. Hasil penelitian tersebut bahwa metode decision tree C4.5 memiliki akurasi lebih baik daripada C5.0 dan WDT [6].

Sedangkan metode *Neural Network* (NN) secara sederhana merupakan pemodelan data statistik non-linear. Arsitektur NN dapat digunakan untuk membentuk model hubungan yang kompleks antara neuron input dan neuron output untuk menemukan pola-pola pada data. Salah satu metode NN yang digunakan adalah multilayer perceptron. Sistem pembelajaran multilayer perceptron merupakan *supervised learning* adalah sistem pembelajaran dimana setiap pengetahuan yang akan diberikan kepada sistem, diberikan terlebih dahulu suatu acuan untuk memetakan suatu masukan menjadi suatu keluaran yang diinginkan. Multi Layer perceptron memiliki hidden layer yang terletak diantara layer input dan output. Hidden ini dapat digunakan lebih dari satu hidden layer [5]. Pada penelitian yang dilakukan Amrin, dalam menentukan keputusan pemberian kredit kepada seorang calon debitur, *neural network multilayer perceptron* memperoleh akurasi sebesar 96,1% sedangkan metode *neural network radial basic function* memperoleh akurasi 89,2% [7].

Penentuan kategori gizi balita menggunakan metode *decision tree* C4.5 dan *Neural Network* ini bertujuan untuk mengetahui variabel apa yang mempengaruhi kategori gizi balita berdasarkan *root* dan *leaf* yang terbentuk pada C4.5 dan mengetahui akurasi klasifikasi yang dihasilkan metode C4.5 dan NN.

## 2. Metode Penelitian

Metode penelitian terdiri atas beberapa tahapan diantaranya :

1. Studi literature  
Studi literatur dilakukan untuk memperoleh referensi yang kuat mengenai kesehatan gizi balita, faktor yang mempengaruhi gizi, dan metode C4.5 serta Neural Network.
2. Pengambilan Data  
Data yang digunakan pada penelitian ini merupakan data sekunder. Hasil Penimbangan Massal Balita Puskesmas Rimbo Data Kecamatan Pangkalan Koto Baru Kabupaten Lima Puluh Kota berjumlah 110 data.
3. Tahap analisa  
Tahap analisa terdiri atas analisa kebutuhan data dan analisa penyelesaian menggunakan decision tree C4.5 dan Neural Network.
4. Pengujian.  
Pada penelitian ini, pengujian dilakukan menggunakan software RapidMiner Studio Free 9.3.001.



Gambar 1. Tahapan Penelitian

### A. Analisa Data

Data yang digunakan terdiri atas beberapa variabel yang akan dijadikan masukan kemudian diproses menggunakan RapidMiner dengan *modeling decision tree C4.5* dan *Neural Network*. Variabel tersebut adalah berat badan, tinggi badan, kategori ekonomi keluarga, pekerjaan ayah, dan pendidikan ibu yang dapat ditunjukkan pada table 1. Pada metode C4.5, nilai asli setiap variabel tidak diubah dalam bentuk numerik. Sedangkan pada metode *Neural Network* setiap data pada variabel diubah dalam bentuk numerik menggunakan normalisasi min-max.

Tabel 1. Variabel masukan

No	Variabel	Satuan Nilai
1	Umur	Nilai dalam bulan
2	Berat badan	Nilai dalam kg
3	Jenis Kelamin	Nilai dalam perempuan dan laki-laki
4	Kategori ekonomi keluarga	1 = Gakin 2 = Non Gakin
5	Pekerjaan Ayah	1 = Pegawai 2 = Wiraswasta 3 = Tani/Supir/Buruh/lainnya
6	Pendidikan Ibu	1 = Perguruan Tinggi 2 = Tamat SMA 3 = Tamat SMP 4 = Tamat SD

Berdasarkan metode yang digunakan yakni bersifat klasifikasi, maka pada saat melakukan proses penyelesaian menggunakan metode C4.5 dan Neural Network memerlukan kelas atau target. Target yang digunakan adalah kategori Gizi yang dihitung berdasarkan nilai Z-score berat badan terhadap umur (BB/U). Tabel 2 menunjukkan kelas dari kategori gizi yang digunakan.

Tabel 2. Kelas kategori gizi

No	Kelas
1	Gizi lebih
2	Gizi baik
3	Gizi kurang
4	Gizi buruk

### B. Analisa Penyelesaian menggunakan Decision tree C4.5 dan Neural Network

Proses yang dilakukan pada penelitian ini menggunakan metode C4.5 dan Neural Network pada RapidMiner.

### 1) Decision Tree C4.5

Pada penyelesaian menggunakan metode C4.5, data yang digunakan tidak dinormalisasi atau diubah dalam nilai numerik karena algoritma C4.5 dapat menangani data numerik dan kategorial. Metode ini dapat melakukan klasifikasi sekaligus menunjukkan hubungan antar atribut dan menghasilkan pohon keputusan. Hal pertama yang dilakukan untuk membentuk pohon keputusan adalah menentukan variabel masukan mana yang menjadi akar dari pohon keputusan tersebut dengan cara menghitung nilai *entropy* dan *gain*. Nilai gain yang paling tinggi digunakan untuk menentukan variabel masukan mana yang menjadi node dari suatu pohon keputusan. Besarnya Entropy pada ruang sampel S didefinisikan dengan:

$$Entropy(S) = -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus}$$

dimana:

S : sampel pelatihan

$p_{\oplus}$  : jumlah yang mendukung data sampel (+) terhadap suatu kriteria

$p_{\ominus}$  : jumlah yang tidak mendukung data sampel (-) terhadap suatu kriteria.

Entropi(S) = 0, jika semua contoh pada S berada dalam kelas yang sama.

Entropi(S) = 1, jika jumlah contoh bernilai (+) dan (-) dalam S adalah sama.

$0 < Entropi(S) < 1$ , jika jumlah contoh bernilai (+) dan (-) dalam S tidak sama.

Sedangkan Gain (S,A) adalah ukuran efektifitas suatu variabel dalam mengklasifikasikan data. Gain dari suatu variabel merupakan selisih antara nilai entropy total dengan entropy dari setiap variabel masukan tersebut.

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

dimana:

A : Variabel masukan

S : Himpunan kasus (Sampel)

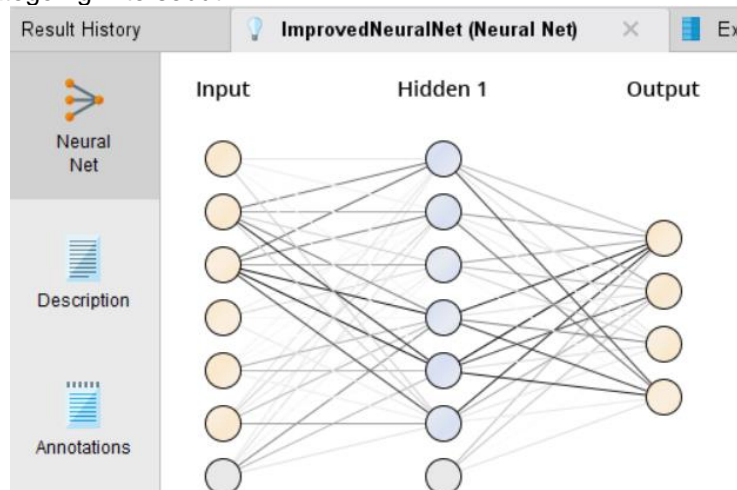
n : Jumlah partisi himpunan atribut A

$|S_i|$  : Jumlah kasus pada partisi ke-i

$|S|$  : Jumlah kasus dalam S

### 2) Neural Network

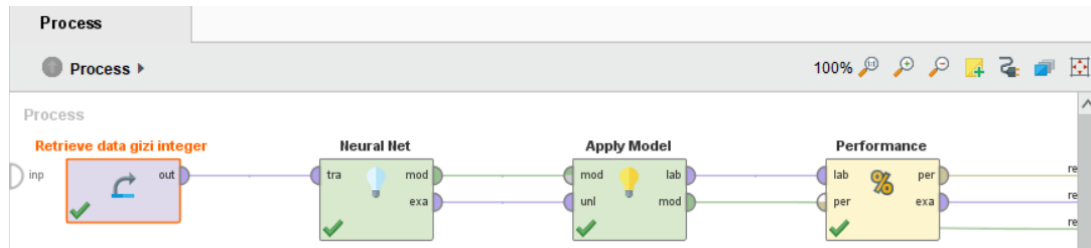
Setiap *Neural Network* terdiri dari unit pengolahan dasar yang saling berhubungan, yang disebut Neuron. *Network* melakukan proses pembelajaran dengan memodifikasi bobot hubungan antara neuron. Pada penelitian ini jenis *Neural Network* yang digunakan adalah *multi layer perceptron*. Jenis *Neural Network* ini memiliki layer yang dinamakan "*hidden*", ditengah layer input dan output. Hidden ini bersifat variabel, dapat digunakan lebih dari satu *hidden layer*. Jumlah input *layer* sama dengan jumlah variabel masukan yang digunakan yaitu 6 neuron yang terdiri atas umur, berat badan, jenis kelamin, kategori ekonomi keluarga, pekerjaan ayah, dan pendidikan ibu. Jumlah *hidden layer* adalah 6 neuron, dan output *layer* adalah 4 neuron sesuai dengan kelas kategori gizi tersebut.



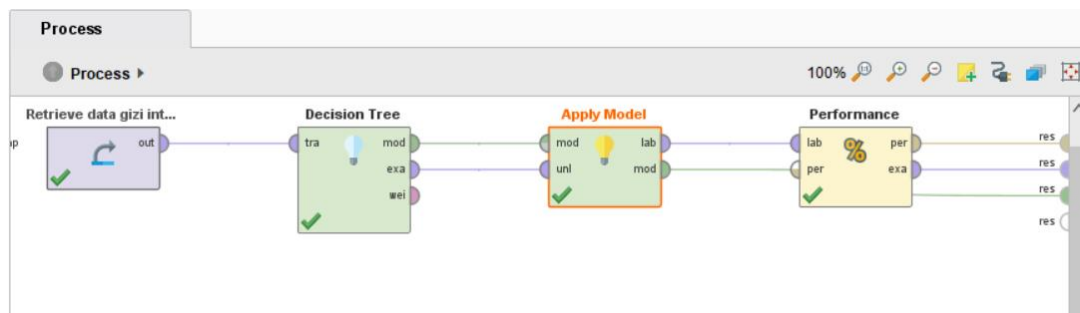
Gambar 2. Arsitektur *Neural Network* Untuk Klasifikasi Kategori Gizi

### 3. Hasil dan Pembahasan

Setelah melakukan analisa kebutuhan data dan metode, maka selanjutnya adalah melakukan pengujian terhadap metode C4.5 dan *Neural Network*. Proses pengujian menggunakan metode C4.5 dapat ditunjukkan pada gambar 3. Sedangkan Proses pengujian menggunakan metode *Neural Network* dapat ditunjukkan pada gambar 4.



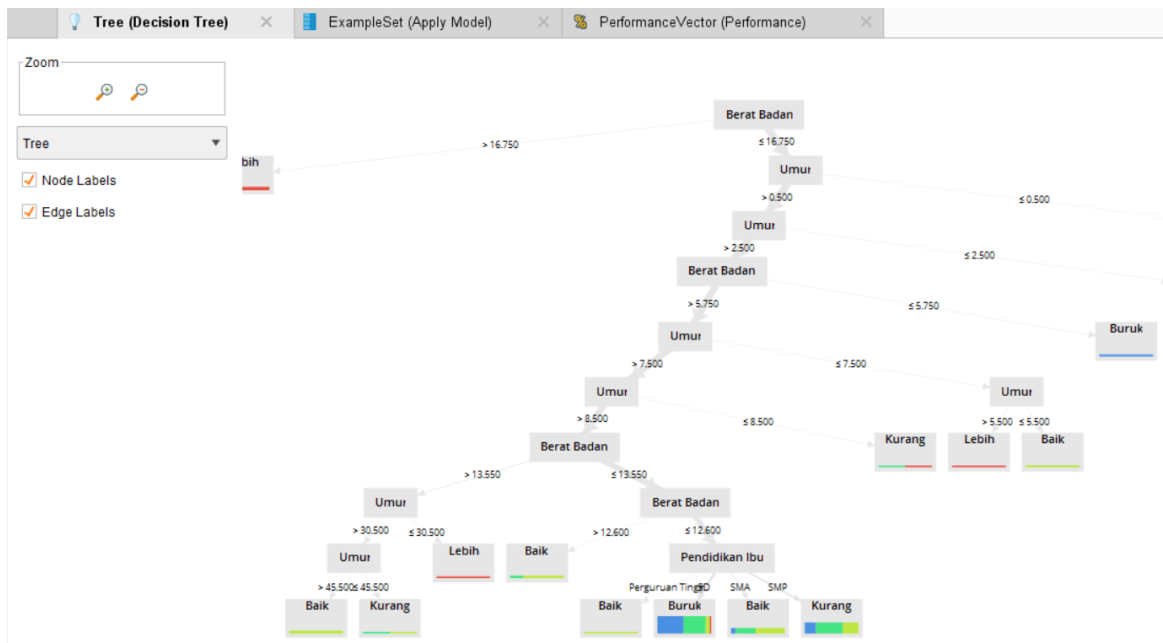
Gambar 3. Proses Pengujian menggunakan C4.5



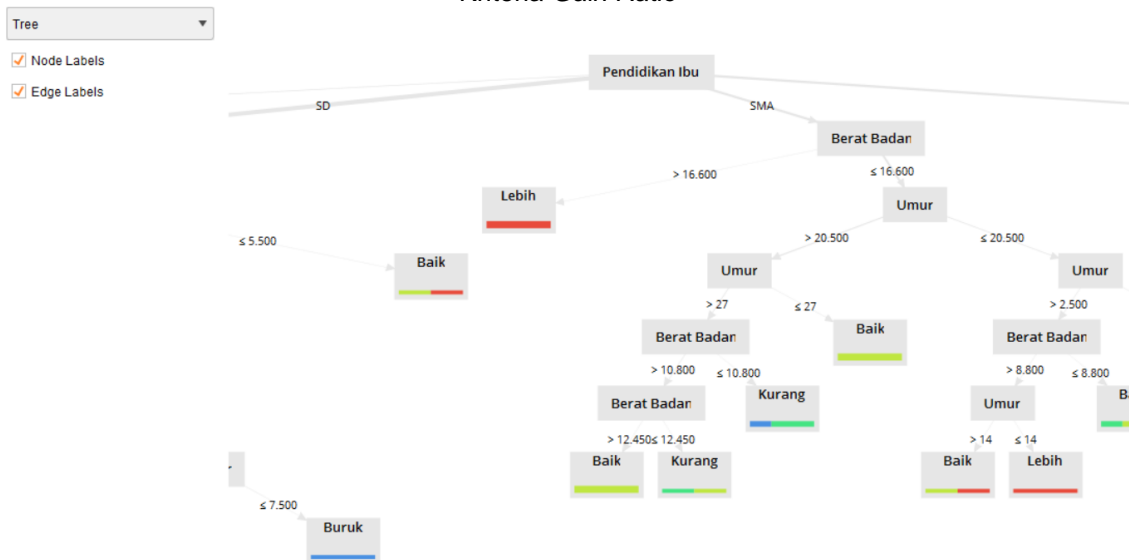
Gambar 4. Proses Pengujian menggunakan *Neural Network*

Setelah dilakukan proses pengujian, maka diperoleh hasil yang dapat dilihat pada gambar 5, 6, dan 7. Gambar 5 dan 6 menunjukkan hasil pohon keputusan penentuan kategori gizi menggunakan metode C4.5. Pada gambar 5 hasil pohon keputusan tersebut diperoleh bahwa yang menjadi *root* adalah variable berat badan, sebagai *leaf* adalah umur dan pendidikan ibu. Sedangkan jenis kelamin, kategori ekonomi keluarga dan pekerjaan ayah tidak ada. Hal ini menunjukkan bahwa tinggi badan, kategori ekonomi keluarga dan pekerjaan ayah tidak mempengaruhi hasil keputusan. Sedangkan pada gambar 6 hasil pohon keputusan tersebut diperoleh bahwa yang menjadi *root* adalah variable pendidikan ibu, sebagai *leaf* adalah umur, berat badan dan jenis kelamin. Sedangkan kategori ekonomi keluarga dan pekerjaan ayah tidak ada. Hal ini menunjukkan bahwa sama halnya dengan gambar 5 yakni kategori ekonomi keluarga dan pekerjaan ayah tidak mempengaruhi hasil keputusan.

Pada gambar 7 merupakan hasil perhitungan akurasi metode C4.5 dengan kriteria gain ratio menggunakan *confusion matrik*.



Gambar 5. Hasil Pohon Keputusan Penentuan Kategori Gizi menggunakan C4.5 dengan Kriteria *Gain Ratio*



Gambar 6. Hasil Pohon Keputusan Penentuan Kategori Gizi menggunakan C4.5 dengan Kriteria *Information Gain*

ion Tree) × ExampleSet (Apply Model) × PerformanceVector (Performance) ×

Table View Plot View

accuracy: 62.73%

	true Buruk	true Kurang	true Baik	true Lebih	class precision
pred. Buruk	22	16	3	1	52.38%
pred. Kurang	5	14	8	1	50.00%
pred. Baik	1	5	19	0	76.00%
pred. Lebih	0	0	1	14	93.33%
class recall	78.57%	40.00%	61.29%	87.50%	

Gambar 7. Hasil Akurasi Metode C4.5

Hasil pengujian parameter pada metode C4.5 dapat dilihat pada tabel 3.

Tabel 3. Pengujian parameter pada C4.5

	Kriteria	No	Minimal leaf size	Minimal size of split	Maximal depth	Confidence	Akurasi
1	Gain ratio	0,01	2	4	10	0,1	62,73%
2	Information gain	0,01	2	4	10	0,1	84,55%

Pada tabel 3 menunjukkan bahwa parameter yang digunakan pada pengujian metode C4.5 mempengaruhi hasil akurasi terutama berdasarkan kriteria *gain ratio* dan *information gain*. *Gain ratio* merupakan variasi dari *information gain*. Metode ini menghasilkan *information gain* untuk setiap atribut yang memberikan nilai atribut yang seragam. Berbeda dengan *information gain*, semua entropi dihitung. Kemudian atribut dengan entropi minimum yang dipilih untuk dilakukan perpecahan pohon (*split*) [5].

Sedangkan pada gambar 7 merupakan hasil perhitungan akurasi metode *Neural Network* menggunakan *confusion matrik*. Hasil akurasi yang diperoleh lebih baik dari pada metode C4.5 yaitu 74.55 %.

	true 4	true 3	true 2	true 1	class precision
pred. 4	21	8	0	0	72.41%
pred. 3	7	25	7	0	64.10%
pred. 2	0	2	22	2	84.62%
pred. 1	0	0	2	14	87.50%
class recall	75.00%	71.43%	70.97%	87.50%	

Gambar 8. Hasil Akurasi menggunakan *Neural Network*

Pada pengujian menggunakan *Neural Network*, dilakukan menggunakan beberapa kombinasi nilai parameter yaitu jumlah *neuron hidden layer* dan nilai *learning rate*.

Tabel 4. Pengujian Parameter *Neural Network*

No	Jumlah Neuron pada <i>Hidden Layer</i>	<i>Learning Rate</i>	Akurasi
1	6	0,01	74,55 %
2	6	0,05	84,55 %
3	8	0,01	74,55 %
4	8	0,05	86,33 %
5	10	0,01	75,45 %
6	10	0,05	97,27 %

Pada tabel 4 diperoleh akurasi terbaik pada jumlah neuron *hidden layer* bernilai 10 dan *learning rate* 0,05, dan jumlah epoch pelatihan yang digunakan pada *rapidminer* adalah 200. Hal ini dapat menunjukkan bahwa setiap variabel yang digunakan yaitu umur, berat badan, jenis kelamin, kategori ekonomi keluarga, pekerjaan ayah, dan pendidikan ibu juga mempengaruhi dalam penentuan kategori gizi balita di kota tersebut.

#### 4. Kesimpulan

Berdasarkan dari hasil pengujian yang di lakukan pada penelitian ini, dapat diambil beberapa kesimpulan sebagai berikut :

- 1) Hasil akurasi yang diperoleh menggunakan metode C4.5 pada data yang berjumlah 110 adalah 62.73% untuk kriteria *gain ratio* dan 84,55% untuk kriteria *information gain*. Sedangkan pada metode *Neural Network* diperoleh lebih baik dari pada metode C4.5 yaitu 97.27% dengan jumlah neuron pada *hidden layer* bernilai 10, *learning rate* 0,05, dan jumlah epoch pelatihan 200.
- 2) Pada C4.5 memberikan hasil pohon keputusan dengan *root* adalah berat badan ketika menggunakan kriteria *gain ratio*, dan pendidikan ibu ketika menggunakan *information gain*.

Sedangkan variable kategori ekonomi keluarga dan pekerjaan ayah tidak sebagai *leaf* pada pohon keputusan.

- 3) Pada *Neural Network* diperoleh hasil akurasi yang tinggi yang menunjukkan semua variabel yang digunakan yaitu umur, berat badan, jenis kelamin, kategori ekonomi keluarga, pekerjaan ayah, dan pendidikan ibu mempengaruhi hasil klasifikasi.

#### Daftar Pustaka

- [1] Kementerian Kesehatan Republik Indonesia. Profil Kesehatan Dinas Kesehatan Pekanbaru. Pekanbaru. 2018
- [2] Putri, R, F., Sulastri, D., dan Lestari, Y. *Faktor-Faktor yang Berhubungan dengan Kategori Gizi Anak Balita di Wilayah Kerja Puskesmas Nanggalo Padang*. Jurnal Kesehatan Andalas. 2015; 254
- [3] Puspasari, N dan Andriani, M. *Hubungan Pengetahuan Ibu tentang Gizi dan Asupan Makan Balita dengan Kategori Gizi Balita (BB/U) Usia 12-24 Bulan*. Amerta Nutr. 2017. 369-378. DOI : 10.2473/amnt.v1i4.2017.369-378
- [4] Sholikah, A., Rustiam, E, R., dan Yuniastuti, A. *Faktor - Faktor yang Berhubungan dengan Kategori Gizi Balita di Pedesaan dan Perkotaan*. Public Health Perspective Journal 2 (1) (2017) 9 – 18
- [5] C Aprilla, D., Baskoro, D, A., Ambarwati, L., dan Wicaksana, I wayan, S. *Belajar Data Mining dengan Rapid miner*. Jakarta. 2013
- [6] A. S. Abdullah, S. Selvakumar, P. Karthikeyan, and M. Venkatesh, "Comparing the Efficacy of Decision Tree and its Variants using Medical Data," Indian J. Sci. Technol., vol. 10, no. 18, pp. 1–8, 2017
- [7] Amrin. Perbandingan Metode Neural Network Radial Basis Function dan Multilayer Perceptron untuk analisa Risiko Kredit Mobil. Paradigma, Vol XX, No. 1, Maret 2018 , e-ISSN 2579 -3500