

---

## PERBANDINGAN AKURASI KLASIFIKASI TINGKAT KEMISKINAN ANTARA ALGORITMA C.45 DAN NAIVE BAYES

Rita Manurung. \*,Puji Sari Ramadhan. \*\*, Moch Iswan Perangin-angin. \*\*\*

\*Program Studi Sistem Informasi, STMIK Triguna Dharma

\*\*Program Studi Sistem Informasi, STMIK Triguna Dharma

\*\*\* Program Studi Sistem Informasi, STMIK Triguna Dharma

---

### Article Info

#### Article history:

Received Jun 12<sup>th</sup>, 201x

Revised Aug 20<sup>th</sup>, 201x

Accepted Aug 26<sup>th</sup>, 201x

---

#### Keyword:

C.45

Naive Bayes

Data Mining

Kemiskinan

---

### ABSTRACT

*Kemiskinan merupakan salah satu masalah yang dialami oleh beberapa Negara berkembang, termasuk Indonesia. Banyak cara yang dilakukan untuk menanggulangi kemiskinan, diantaranya dengan program bantuan sosial untuk rakyat miskin. Bentuk bantuan sosial yang diberikan oleh pemerintah disesuaikan dengan tingkat kemiskinan yang ada di suatu wilayah sehingga pemberian bantuan sosial tersebut tidak salah sasaran. Pada penelitian kali ini kami menggunakan BDT (Basis Data Terpadu) yang dikeluarkan oleh TNP2K dalam menentukan klasifikasi tingkat kemiskinan. Adapun metode yang digunakan dalam penelitian ini adalah Naive Bayes Classifier (NBC) dan Algoritma C4.5 yang keduanya merupakan metode pada teknik klasifikasi data mining. Pengujian akan dilakukan dengan menggunakan 14 atribut. Hasil dari proses klasifikasi diperoleh bahwa metode C4.5 memiliki tingkat akurasi 3% lebih baik jika dibandingkan dengan metode Naive Bayes.*

**Kata Kunci:** C.45 , Naive Bayes, Data Mining, Kemiskinan

Copyright © 2019 STMIK Triguna Dharma.  
All rights reserved.

---

### Corresponding Author:

Nama :Rita Manurung

Program Studi : Sistem Informasi

STMIK Triguna Dharma

Email: Rita.manurung997@gmail.com

---

### 1. PENDAHULUAN

Pada masa sekarang ini aliran informasi telah meningkat dalam jumlah yang besar setiap harinya, peningkatan aliran informasi ini akan menyebabkan terjadinya penumpukan data berupa dokumen teks, baik secara online maupun offline. Dokumen teks yang menumpuk menyebabkan sulitnya mencari dokumen yang sesuai dengan kebutuhan. Ketersediaan data yang berlimpah yang dihasilkan dari penggunaan teknologi informasi di hampir semua bidang kehidupan menimbulkan kebutuhan untuk dapat memanfaatkan informasi dan pengetahuan yang terkandung di dalam limpahan data tersebut, yang kemudian melahirkan data mining. Data Mining merupakan proses untuk menemukan pengetahuan (*knowledge discovery*) yang ditambang dari sekumpulan data yang volumenya besar. Dalam data mining terdapat beberapa fungsi yaitu data mining untuk *Association Rules* (Aturan Asosiasi), data mining untuk *Classification* (Klasifikasi), data mining untuk *Clustering* (Cluster/Pengelompokan), data mining untuk *Prediction* (Prediksi), dan data mining untuk *Forecasting* (Peramalan).

Indonesia merupakan salah satu Negara berkembang di Asia khususnya Asia Tenggara. Salah satu masalah yang sering dihadapi oleh Negara berkembang adalah kemiskinan. Berdasarkan data yang dikeluarkan oleh Badan Pusat Statistik, angka kemiskinan di Indonesia pada tahun 1999 mencapai 47.97 juta jiwa. Pada tahun 2011 jumlah penduduk miskin menjadi 30.02 juta jiwa. Badan Pusat Statistik melakukan pendataan kependudukan khususnya masalah kemiskinan setiap 3 tahun sekali. Proses pendataan dilakukan dengan cara *door to door* langsung menuju rumah tangga sasaran. Bagi pemerintah Indonesia masalah

kemiskinan merupakan masalah lama yang belum dan sulit untuk diselesaikan. Pemerintah sendiri telah melakukan beberapa upaya dalam melakukan pengentasan kemiskinan diantaranya melalui program bantuan sosial diantaranya Bantuan langsung tunai (BLT), Program Keluarga Harapan (PKH) dan lain-lain. Salah satu kesulitan yang terkadang dihadapi oleh pemerintah adalah proses pembagian bantuan sosial yang tidak merata dan tepat sasaran.

Salah satu pemodelan yang bisa digunakan untuk klasifikasi tingkat kemiskinan tersebut adalah dengan menggunakan data mining. Dalam penelitian kali ini penulis mencoba membandingkan model klasifikasi yang dibentuk oleh teknik data mining antara algoritma *decision tree C4.5* dan *Naïve Bayes Clasifier (NBC)*.

## 2. METODE PENELITIAN

Berikut ini adalah uraian langkah-langkah dalam algoritma C4.5 untuk menyelesaikan kasus klasifikasi kemiskinan. *Dataset* yang telah ada pada tabel 3.1 akan digunakan untuk membentuk pohon keputusan.

Maka dilakukan tahapan-tahapan sebagai berikut:

*Tabel 1 Data Pendudukan Miskin*

No	NK	a	b	c	d	e	f	g	h	i	j	k	l	m	n	ket
1	K-1	T	T	T	T	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Ya
2	K-2	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Ya
3	K-3	T	T	T	T	T	T	T	T	T	T	T	T	T	T	Tidak
4	K-4	Y	T	Y	Y	T	Y	T	Y	Y	Y	Y	Y	Y	Y	Tidak
5	K-5	T	Y	Y	Y	T	Y	Y	Y	T	Y	T	Y	Y	T	Ya
6	K-6	Y	Y	T	Y	Y	Y	Y	Y	T	Y	T	Y	Y	Y	Ya
7	K-7	T	Y	Y	Y	Y	Y	Y	Y	Y	T	Y	Y	T	T	Ya
8	K-8	Y	T	T	Y	T	Y	Y	Y	Y	Y	T	Y	Y	Y	Ya
9	K-9	T	Y	T	T	T	Y	Y	Y	Y	Y	Y	Y	Y	Y	Ya
10	K-10	Y	T	Y	Y	T	Y	T	Y	Y	Y	Y	Y	Y	Y	Ya
11	K-11	T	Y	Y	Y	T	Y	Y	Y	T	Y	T	Y	Y	T	Ya
12	K-12	Y	T	Y	Y	T	Y	T	Y	Y	Y	Y	Y	Y	Y	Tidak
13	K-13	T	Y	Y	Y	Y	Y	Y	Y	Y	T	Y	Y	T	T	Ya
14	K-15	Y	T	Y	Y	T	Y	T	Y	Y	Y	Y	Y	Y	Y	Tidak

Keterangan

NK : Nama keluarga

- Luas lantai bangunan tempat tinggal kurang dari 8 m2 per orang
- Jenis lantai tempat tinggal terbuat dari tanah/bambu/kayu murahan
- Jenis dinding tempat tinggal dari bambu/ rumbia/ kayu berkualitas rendah/tembok tanpa diplester.
- Tidak memiliki fasilitas buang air besar/ bersama-sama dengan rumah tangga lain.
- Sumber penerangan rumah tangga tidak menggunakan listrik.
- Sumber air minum berasal dari sumur/ mata air tidak terlindung/ sungai/ air hujan.
- Bahan bakar untuk memasak sehari-hari adalah kayu bakar/ arang/ minyak tanah
- Hanya mengkonsumsi daging/ susu/ ayam dalam satu kali seminggu.
- Hanya membeli satu stel pakaian baru dalam setahun

- j. Hanya sanggup makan sebanyak satu/ dua kali dalam sehari  
k. Tidak sanggup membayar biaya pengobatan di puskesmas/ poliklinik  
l. Sumber penghasilan kepala rumah tangga adalah: petani dengan luas lahan 500m2, buruh tani, nelayan, buruh bangunan, buruh perkebunan dan atau pekerjaan lainnya dengan pendapatan dibawah Rp. 600.000,- per bulan  
m. Pendidikan tertinggi kepala rumah tangga: tidak sekolah/ tidak tamat SD/ tamat SD.  
Tidak memiliki tabungan/ barang yang mudah dijual dengan minimal Rp. 500.000,- seperti sepeda motor kredit/ non kredit, emas, ternak, kapal motor, atau barang modal lainnya.

**a. Perhitungan Total Entropy**

Berdasarkan persamaan 2 dapat dihitung nilai Entropy untuk keseluruhan data *sample/training/learning dataset* (S) sebagai berikut:

$$Entropy(S) = \sum_{i=0}^n -p_i * \log_2 p_i \dots \dots (2)$$

$$Entropy(Total) = \left(-\frac{4}{14} * \log_2\left(\frac{4}{14}\right)\right) + \left(-\frac{10}{14} * \log_2\left(\frac{10}{14}\right)\right)$$

$$Entropy(Total) = 0.863120569$$

$$Entropy(a-T) = \left(-\frac{1}{7} * \log_2\left(\frac{1}{7}\right)\right) + \left(-\frac{6}{7} * \log_2\left(\frac{6}{7}\right)\right) = 0.59167$$

$$Entropy(a-Y) = \left(-\frac{3}{7} * \log_2\left(\frac{3}{7}\right)\right) + \left(-\frac{4}{7} * \log_2\left(\frac{4}{7}\right)\right) = 0.98523$$

$$Gain(a) = 0.86312 - \left\{\frac{4}{14} * 0.59167\right\} + \left\{\frac{10}{14} * 0.98523\right\} = -0.00757$$

$$Entropy(b-T) = \left(-\frac{1}{7} * \log_2\left(\frac{1}{7}\right)\right) + \left(-\frac{6}{7} * \log_2\left(\frac{6}{7}\right)\right) = 0.98523$$

$$Entropy(b-Y) = \left(-\frac{3}{7} * \log_2\left(\frac{3}{7}\right)\right) + \left(-\frac{4}{7} * \log_2\left(\frac{4}{7}\right)\right) = 0.59167$$

$$Gain(b) = 0.86312 - \left\{\frac{4}{14} * 0.98523\right\} + \left\{\frac{10}{14} * 0.59167\right\} = 0.159004$$

$$Entropy(c-T) = \left(-\frac{1}{5} * \log_2\left(\frac{1}{5}\right)\right) + \left(-\frac{4}{5} * \log_2\left(\frac{4}{5}\right)\right) = 0.92573$$

$$Entropy(c-Y) = \left(-\frac{3}{10} * \log_2\left(\frac{3}{10}\right)\right) + \left(-\frac{7}{10} * \log_2\left(\frac{7}{10}\right)\right) = 0.88129$$

$$Gain(c) = 0.86312 - \left\{\frac{4}{14} * 0.92573\right\} + \left\{\frac{10}{14} * 0.88129\right\} = -0.03087$$

$$Entropy(d-T) = \left(-\frac{1}{3} * \log_2\left(\frac{1}{3}\right)\right) + \left(-\frac{2}{3} * \log_2\left(\frac{2}{3}\right)\right) = 0.91830$$

$$Entropy(d-Y) = \left(-\frac{3}{11} * \log_2\left(\frac{3}{11}\right)\right) + \left(-\frac{8}{11} * \log_2\left(\frac{8}{11}\right)\right) = 0.84535$$

$$Gain(d) = 0.86312 - \left\{\frac{4}{14} * 0.91830\right\} + \left\{\frac{10}{14} * 0.84535\right\} = -0.00307$$

$$Entropy(e-T) = \left(-\frac{0}{9} * \log_2\left(\frac{0}{9}\right)\right) + \left(-\frac{9}{9} * \log_2\left(\frac{9}{9}\right)\right) = 0$$

$$Entropy(e-Y) = \left(-\frac{0}{5} * \log_2\left(\frac{0}{5}\right)\right) + \left(-\frac{5}{5} * \log_2\left(\frac{5}{5}\right)\right) = 0$$

$$Gain(e) = 0.86312 - \left\{\frac{4}{14} * 0\right\} + \left\{\frac{10}{14} * 0\right\} = 0$$

$$Entropy(f-T) = \left(-\frac{1}{1} * \log_2\left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) = 0$$

$$Entropy(f-Y) = \left(-\frac{3}{13} * \log_2\left(\frac{3}{13}\right)\right) + \left(-\frac{10}{13} * \log_2\left(\frac{10}{13}\right)\right) = 0.77935$$

$$Gain(f) = 0.86312 - \left\{\frac{4}{14} * 0\right\} + \left\{\frac{10}{14} * 0.77935\right\} = 0.306443$$

$$Entropy(g-T) = \left(-\frac{4}{5} * \log_2\left(\frac{4}{5}\right)\right) + \left(-\frac{1}{5} * \log_2\left(\frac{1}{5}\right)\right) = 0.72193$$

$$\text{Entropy}(g - Y) = \left(-\frac{3}{9} * \log_2\left(\frac{3}{9}\right)\right) + \left(-\frac{6}{9} * \log_2\left(\frac{6}{9}\right)\right) = 0.91830$$

$$\text{Gain}(g) = 0.86312 - \left\{\frac{4}{14} * 0.72193\right\} + \left\{\frac{10}{14} * 0\right\} = 0.000926$$

$$\text{Entropy}(h - T) = \left(-\frac{1}{1} * \log_2\left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) = 0$$

$$\text{Entropy}(h - Y) = \left(-\frac{3}{13} * \log_2\left(\frac{3}{13}\right)\right) + \left(-\frac{10}{13} * \log_2\left(\frac{10}{13}\right)\right) = 0.77935$$

$$\text{Gain}(h) = 0.86312 - \left\{\frac{4}{14} * 0\right\} + \left\{\frac{10}{14} * 0.77935\right\} = 0.306442$$

$$\text{Entropy}(i - T) = \left(-\frac{1}{4} * \log_2\left(\frac{1}{4}\right)\right) + \left(-\frac{3}{4} * \log_2\left(\frac{3}{4}\right)\right) = 0.81128$$

$$\text{Entropy}(i - Y) = \left(-\frac{3}{10} * \log_2\left(\frac{3}{10}\right)\right) + \left(-\frac{7}{10} * \log_2\left(\frac{7}{10}\right)\right) = 0.81129$$

$$\text{Gain}(i) = 0.86312 - \left\{\frac{4}{14} * 0\right\} + \left\{\frac{10}{14} * 0.81128\right\} = 0.051833$$

$$\text{Entropy}(j - T) = \left(-\frac{1}{3} * \log_2\left(\frac{1}{3}\right)\right) + \left(-\frac{2}{3} * \log_2\left(\frac{2}{3}\right)\right) = 0.91830$$

$$\text{Entropy}(j - Y) = \left(-\frac{3}{11} * \log_2\left(\frac{3}{11}\right)\right) + \left(-\frac{8}{11} * \log_2\left(\frac{8}{11}\right)\right) = 0.76876$$

$$\text{Gain}(j) = 0.86312 - \left\{\frac{4}{14} * 0.91830\right\} + \left\{\frac{10}{14} * 0.84535\right\} = -0.051634$$

$$\text{Entropy}(k - T) = \left(-\frac{1}{5} * \log_2\left(\frac{1}{5}\right)\right) + \left(-\frac{4}{5} * \log_2\left(\frac{4}{5}\right)\right) = 0.72193$$

$$\text{Entropy}(k - Y) = \left(-\frac{3}{9} * \log_2\left(\frac{3}{9}\right)\right) + \left(-\frac{6}{9} * \log_2\left(\frac{6}{9}\right)\right) = 0.91830$$

$$\text{Gain}(k) = 0.86312 - \left\{\frac{4}{14} * 0.72193\right\} + \left\{\frac{10}{14} * 0.91830\right\} = -0.000926$$

$$\text{Entropy}(l - T) = \left(-\frac{1}{1} * \log_2\left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) = 0$$

$$\text{Entropy}(l - Y) = \left(-\frac{3}{13} * \log_2\left(\frac{3}{13}\right)\right) + \left(-\frac{10}{13} * \log_2\left(\frac{10}{13}\right)\right) = 0.77935$$

$$\text{Gain}(k) = 0.86312 - \left\{\frac{4}{14} * 0\right\} + \left\{\frac{10}{14} * 0.77935\right\} = 0.306441$$

$$\text{Entropy}(m - T) = \left(-\frac{1}{3} * \log_2\left(\frac{1}{3}\right)\right) + \left(-\frac{2}{3} * \log_2\left(\frac{2}{3}\right)\right) = 0.91830$$

$$\text{Entropy}(m - Y) = \left(-\frac{3}{11} * \log_2\left(\frac{3}{11}\right)\right) + \left(-\frac{8}{11} * \log_2\left(\frac{8}{11}\right)\right) = 0.84535$$

$$\text{Gain}(m) = 0.86312 - \left\{\frac{4}{14} * 0.91830\right\} + \left\{\frac{10}{14} * 0.84535\right\} = -0.00307$$

$$\text{Entropy}(n - T) = \left(-\frac{1}{5} * \log_2\left(\frac{1}{5}\right)\right) + \left(-\frac{4}{5} * \log_2\left(\frac{4}{5}\right)\right) = 0.72193$$

$$\text{Entropy}(n - Y) = \left(-\frac{3}{9} * \log_2\left(\frac{3}{9}\right)\right) + \left(-\frac{6}{9} * \log_2\left(\frac{6}{9}\right)\right) = 0.91830$$

$$\text{Gain}(n) = 0.86312 - \left\{\frac{4}{14} * 0.72193\right\} + \left\{\frac{10}{14} * 0.91830\right\} = 0.000926$$

Hasil perhitungan Entropy dan gain dapat dilihat pada tabel 3.2. di bawah ini.

Tabel 2. Hasil perhitungan entropy dan gain

Atribut	Nilai	Sum (Nilai)	Sum (Tidak)	Sum(Ya)	Entropy	Gain
a	T	7	1	6	0.59167	
	Y	7	3	4	0.98523	
						-0.000757
b	T	7	1	6	0.98523	
	Y	7	3	4	0.59167	
						0.159004
c	T	5	1	4	0.92573	
	Y	10	3	7	0.88129	
						0.03087
d	T	3	1	2	0.91830	
	Y	11	3	8	0.84535	
						0.00307
e	T	9	0	9	0	
	Y	5	0	5	0	
						0
f	T	1	1	0	0	
	Y	13	3	10	0.77935	
						0.306443
g	T	5	4	1	0.72193	
	Y	9	3	6	0.91830	
						0.000926
h	T	1	1	0	0	
	Y	13	3	10	0.77935	
						0.306442
i	T	4	1	3	0.81128	
	Y	10	3	7	0.81129	
						0.051833

j	T	3	1	2	0.91830	
	Y	11	3	8	0.76876	
						0.051634
k	T	5	1	4	0.72193	
	Y	9	3	6	0.91830	
						0.000926
l	T	1	1	0	0	
	Y	13	3	10	0.77935	
						0.306441
m	T	3	1	2	0.91830	
	Y	11	3	8	0.84535	
						0.00307
n	T	5	1	4	0.72193	
	Y	9	3	6	0.91830	
						0.000926

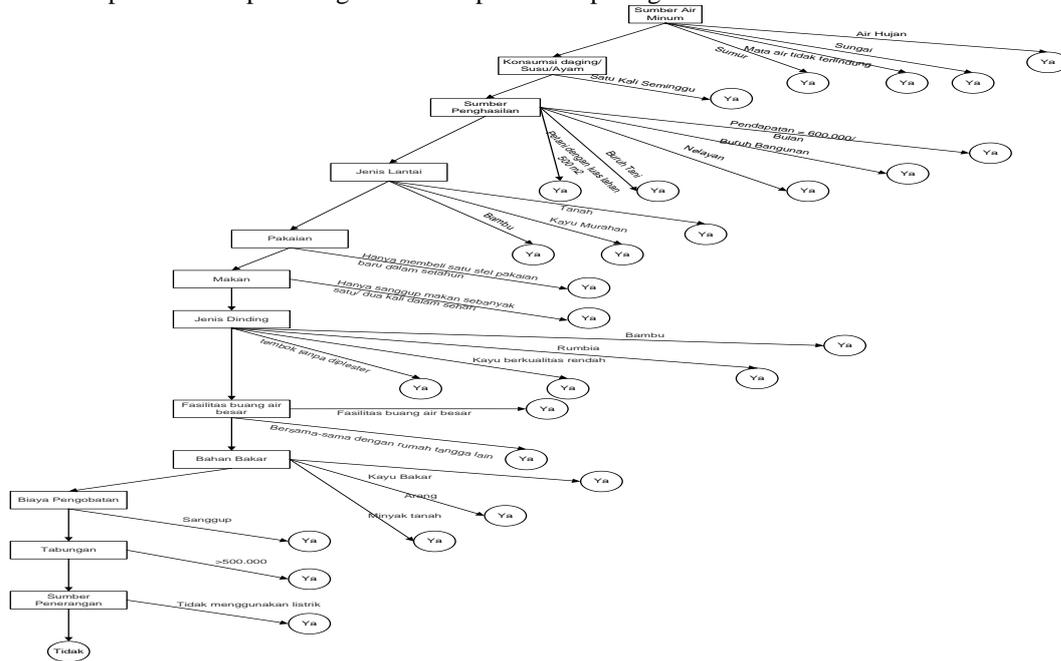
Setelah melakukan perhitungan tersebut, maka dapat dilihat bahwa nilai gain terbesar adalah Gain pada atribut "f", maka atribut "f" yang akan menjadi node akar (*root node*).

Tabel 3. Tabel Urutan Gain dari yang terbesar ke terkecil

Atribut	Gain
f	0.306443
h	0.306442
l	0.306441
b	0.159004
i	0.051833
j	0.051634
c	0.030870
d	0.003070
m	0.003070
g	0.000926
k	0.000926
n	0.000926
e	0.000000
a	-0.000757

Dari hasil perhitungan pada tabel diatas, diketahui bahwa nilai Gain terbesar yaitu pada atribut f yaitu sebesar 0, 306441. Sehingga atribut f menjadi node akar.

Pohon keputusan dari perhitungan diatas dapat dilihat pada gambar di bawah ini.



Gambar 4. Pohon Keputusan

Selanjutnya hasil dari pohon keputusan tersebut di ekstrasi menjadi rule seperti pada tabel di bawah ini :

Tabel 4. Rule berdasarkan node akhir

Id	Rule
1	IF F = "Ya" Maka Klasifikasi/kelayakan ="YA"
2	IF F = "Tidak" DAN H="Ya" MAKA Klasifikasi/kelayakan ="YA"
3	IF F = "Tidak" DAN H="Tidak" DAN L=" Ya " MAKA Klasifikasi/kelayakan ="YA"
4	IF F = "Tidak" DAN H="Tidak" DAN L= Tidak " DAN B=" Ya " MAKA Klasifikasi/kelayakan ="YA"
3	IF F = "Tidak" DAN H= "Tidak" DAN L=" Tidak " DAN B=" Tidak " DAN I=" Ya" MAKA Klasifikasi/kelayakan ="YA"
4	IF F = "Tidak" DAN H="Tidak" DAN L= Tidak " DAN B= Tidak " DAN I=" Tidak" DAN J= "Ya" MAKA Klasifikasi/kelayakan ="YA"
5	IF F = "Tidak" DAN H="Tidak" DAN L="Tidak" DAN I="Tidak" DAN C="Ya" MAKA Klasifikasi/kelayakan ="YA"
6	IF F ="Tidak" DAN H="Tidak" DAN L="Tidak" DAN I="Tidak" DAN J<>="Tidak" DAN C="Tidak" DAN D="Ya" MAKA Klasifikasi/kelayakan ="YA"
7	IF F ="Tidak" DAN H="Tidak" DAN L="Tidak" DAN B="Tidak" DAN I="Tidak" DAN J="Tidak" DAN C="Tidak" DAN D="Tidak" DAN M="Ya" MAKA Klasifikasi/kelayakan ="YA"
8	IF F ="Tidak" DAN H="Tidak" DAN B="Tidak" DAN ITidak" DAN J="Tidak" DAN C="Tidak" DAN D="Tidak" DAN M="Tidak" DAN G="Ya" MAKA Klasifikasi/kelayakan ="YA"

Tabel 3.4. Lanjutan Rule berdasarkan node akhir

9	IF F ="Tidak" DAN H="Tidak" DAN B="Tidak" DAN I="Tidak" DAN J="Tidak" DAN C="Tidak" DAN D="Tidak" DAN M="Tidak" DAN G="Tidak" DAN K =" Ya" MAKA Klasifikasi/kelayakan ="YA"
10	IF F ="Tidak" DAN H="Tidak" DAN B="Tidak" DAN I="Tidak" DAN J="Tidak" DAN C="Tidak" DAN D="Tidak" DAN M="Tidak" DAN G="Tidak" DAN K =" Tidak" DAN E=" Ya" MAKA Klasifikasi/kelayakan ="YA"
11	IF F ="Tidak" DAN H="Tidak" DAN B="Tidak" DAN I="Tidak" DAN J="Tidak" DAN C="Tidak" DAN D="Tidak" DAN M="Tidak" DAN G="Tidak" DAN K ="Tidak" DAN E=" Tidak" DAN A="Ya" MAKA Klasifikasi/kelayakan ="YA"
12	IF F ="Ya" DAN H="Ya" DAN B="Ya" DAN I="Ya" DAN J="Ya" DAN C="Ya" DAN D="Ya" DAN M="Ya" DAN G="Ya" DAN K ="Ya" DAN E=" Ya" DAN A="Ya" MAKA Klasifikasi/kelayakan ="YA"
13	IF F ="Tidak" DAN H="Tidak" DAN B="Tidak" DAN I=" Tidak " DAN J="Tidak" DAN C="Tidak" DAN D="Tidak" DAN M="Tidak" DAN G="Tidak" DAN K ="Tidak" DAN E=" Tidak" DAN A="Tidak" MAKA Klasifikasi/kelayakan ="TIDAK"

### 3.3.Perhitungan Algoritma Naive Bayes

Data yang diolah dalam perhitungan Algoritma Naive Bayes menggunakan tabel 3.1. Dasar teori yang digunakan dalam melakukan klasifikasi ini adalah teorema Bayes yang ditunjukkan oleh persamaan di bawah ini.

$$P(A|B) = (p(B|A) * p(A)) / p(B)$$

Apabila diberikan input baru, maka klasifikasi data penduduk miskin dapat ditentukan melalui langkah berikut :

#### 1. Menghitung *Class Probabilities*

$$P(\text{Keterangan} = \text{"Tidak"}) = 4/14 = 0.285714$$

$$P(\text{Keterangan} = \text{"Ya"}) = 10/14 = 0.714286$$

*Class Probabilities* ini adalah nilai peluang/probabilitas dari masing-masing kelas masalah dari tujuan yang akan diharapkan/diprediksi. Sehingga dari perhitungan berdasarkan data training yang diberikan tersebut diperoleh peluang untuk *Keterangan = Tidak* adalah 0.285714, *Keterangan = Ya* adalah 0.714286.

#### 2. Menghitung *Conditional Probabilities*

Langkah selanjutnya adalah menghitung *Conditional Probabilities*, yaitu probabilitas setiap nilai input terhadap nilai class *keterangan*. Dalam perhitungannya terdapat dua macam perhitungannya

1. Untuk data-data secara numeris dilakukan dengan perhitungan jumlah dan probabilitasnya.
2. untuk data berupa bilangan ordinal, maka perlu adanya pemodelan ulang bentuk data tersebut dengan mencari nilai *mean* dan *standart deviasi*.

Perhitungan untuk masing-masing peluang untuk kriteria/atribut *a* terhadap nilai class *keterangan*. Perhitungan detailnya adalah sebagai berikut :

##### a. Probabilitas Atribut a

$$P(a=T | \text{Keterangan}=\text{Tidak})=1/4=0.25$$

$$P(a=Y | \text{Keterangan}=\text{Tidak})=3/4=0.75$$

$$P(a=T | \text{Keterangan}=\text{Ya})=3/10=0.3$$

$$P(a=Y | \text{Keterangan}=\text{Ya})=6/10=0.6$$

Tabel 5. Hasil perhitungan Probabilitas Atribut a

Atribut a	Keterangan		Probabilitas	
	Tidak	Ya	Tidak	Ya
T	1	3	0.25	0.75

Y	3	6	0.3	0.6
---	---	---	-----	-----

b. Probabilitas Atribut b

$P(b=T \mid Keterangan=Tidak)=4/4=1$

$P(b=Y \mid Keterangan=Tidak)=0/4=0$

$P(b=T \mid Keterangan=Ya)=3/10=0.3$

$P(b=Y \mid Keterangan=Ya)=7/10=0.7$

*Tabel 6. Hasil perhitungan Probabilitas Atribut b*

Atribut b	Keterangan		Probabilitas	
	Tidak	Ya	Tidak	Ya
T	4	0	1	0
Y	3	7	0.3	0.7

a. Probabilitas Atribut c

$P(c=T \mid Keterangan=Tidak)=1/4=0.25$

$P(c=Y \mid Keterangan=Tidak)=3/4=0.75$

$P(c=T \mid Keterangan=Ya)=4/10=0.4$

$P(c=Y \mid Keterangan=Ya)=6/10=0.6$

*Tabel 7. Hasil perhitungan Probabilitas Atribut c*

Atribut c	Keterangan		Probabilitas	
	Tidak	Ya	Tidak	Ya
T	1	3	0.25	0.75
Y	4	6	0.4	0.6

b. Probabilitas Atribut d

$P(d=T \mid Keterangan=Tidak)=1/4=0.25$

$P(d=Y \mid Keterangan=Tidak)=3/4=0.75$

$P(d=T \mid Keterangan=Ya)=3/10=0.3$

$P(d=Y \mid Keterangan=Ya)=7/10=0.7$

*Tabel 8. Hasil perhitungan Probabilitas Atribut d*

Atribut d	Keterangan		Probabilitas	
	Tidak	Ya	Tidak	Ya
T	1	3	0.25	0.75
Y	3	7	0.3	0.7

c. Probabilitas Atribut e

$P(e=T \mid Keterangan=Tidak)=4/4=1$

$P(e=Y \mid Keterangan=Tidak)=0/4=0$

$P(e=T \mid Keterangan=Ya)=5/10=0.5$

$P(e=Y \mid Keterangan=Ya)=5/10=0.5$

*Tabel 9. Hasil perhitungan Probabilitas Atribut e*

Atribut e	Keterangan		Probabilitas	
	Tidak	Ya	Tidak	Ya
T	4	0	1	0
Y	0	5	0	0.5

T	4	0	1	0
Y	5	5	0.5	0.5

## d. Probabilitas Atribut f

$$P(f=T \mid \text{Keterangan=Tidak})=1/4=0.25$$

$$P(f=Y \mid \text{Keterangan=Tidak})=0/4=0$$

$$P(f=T \mid \text{Keterangan=Ya})=3/10=0.3$$

$$P(f=Y \mid \text{Keterangan=Ya})=7/10=0.7$$

Tabel 10. Hasil perhitungan Probabilitas Atribut f

Atribut f	Keterangan		Probabilitas	
	Tidak	Ya	Tidak	Ya
T	1	0	0.25	0
Y	3	7	0.3	0.7

## e. Probabilitas Atribut g

$$P(g=T \mid \text{Keterangan=Tidak})=4/4=1$$

$$P(g=Y \mid \text{Keterangan=Tidak})=0/4=0$$

$$P(g=T \mid \text{Keterangan=Ya})=0/10=0$$

$$P(g=Y \mid \text{Keterangan=Ya})=9/10=0.9$$

Tabel 11. Hasil perhitungan Probabilitas Atribut g

Atribut g	Keterangan		Probabilitas	
	Tidak	Ya	Tidak	Ya
T	4	0	1	0
Y	0	9	0	0.9

## f. Probabilitas Atribut h

$$P(h=T \mid \text{Keterangan=Tidak})=1/4=0.25$$

$$P(h=Y \mid \text{Keterangan=Tidak})=3/4=0.75$$

$$P(h=T \mid \text{Keterangan=Ya})=0/10=0$$

$$P(h=Y \mid \text{Keterangan=Ya})=10/10=1$$

Tabel 12. Hasil perhitungan Probabilitas Atribut h

Atribut h	Keterangan		Probabilitas	
	Tidak	Ya	Tidak	Ya
T	1	3	0.25	0.75
Y	0	10	0	1

## g. Probabilitas Atribut i

$$P(i=T \mid \text{Keterangan=Tidak})=1/4=0.25$$

$$P(i=Y \mid \text{Keterangan=Tidak})=3/4=0.75$$

$$P(i=T \mid \text{Keterangan=Ya})=3/10=0.3$$

$$P(i=Y \mid \text{Keterangan=Ya})=7/10=0.7$$

Tabel 13. Hasil perhitungan Probabilitas Atribut i

Atribut i	Keterangan		Probabilitas	
	Tidak	Ya	Tidak	Ya
T	1	3	0.25	0.75
Y	3	7	0.3	0.7

h. Probabilitas Atribut j  
 $P(j=T | Keterangan=Tidak)=1/4=0.25$   
 $P(j=Y | Keterangan=Tidak)=3/4=0.75$   
 $P(j=T | Keterangan=Ya)=2/10=0.3$   
 $P(j=Y | Keterangan=Ya)=8/10=0.7$

Tabel 14. Hasil perhitungan Probabilitas Atribut j

Atribut j	Keterangan		Probabilitas	
	Tidak	Ya	Tidak	Ya
T	1	3	0.25	0.75
Y	2	8	0.2	0.8

i. Probabilitas Atribut k  
 $P(k=T | Keterangan=Tidak)=1/4=0.25$   
 $P(k=Y | Keterangan=Tidak)=3/4=0.75$   
 $P(k=T | Keterangan=Ya)=3/10=0.3$   
 $P(k=Y | Keterangan=Ya)=7/10=0.7$

Tabel 15. Hasil perhitungan Probabilitas Atribut k

Atribut k	Keterangan		Probabilitas	
	Tidak	Ya	Tidak	Ya
T	1	3	0.25	0.75
Y	3	7	0.3	0.7

j. Probabilitas Atribut l  
 $P(l=T | Keterangan=Tidak)=1/4=0.25$   
 $P(l=Y | Keterangan=Tidak)=3/4=0.75$   
 $P(l=T | Keterangan=Ya)=0/10=0$   
 $P(l=Y | Keterangan=Ya)=10/10=1$

Tabel 16. Hasil perhitungan Probabilitas Atribut l

Atribut l	Keterangan		Probabilitas	
	Tidak	Ya	Tidak	Ya
T	1	3	0.25	0.75
Y	0	10	0	1

k. Probabilitas Atribut m  
 $P(m=T | Keterangan=Tidak)=1/4=0.25$   
 $P(m=Y | Keterangan=Tidak)=3/4=0.75$

$$P(m=T \mid \text{Keterangan}=Ya)=2/10=0.2$$

$$P(m=Y \mid \text{Keterangan}=Ya)=8/10=0.8$$

Tabel 17. Hasil perhitungan Probabilitas Atribut m

Atribut m	Keterangan		Probabilitas	
	Tidak	Ya	Tidak	Ya
T	1	3	0.25	0.75
Y	0.2	0.8	0.2	0.8

1. Probabilitas Atribut n

$$P(n=T \mid \text{Keterangan}=Tidak)=1/4=0.25$$

$$P(n=Y \mid \text{Keterangan}=Tidak)=3/4=0.75$$

$$P(n=T \mid \text{Keterangan}=Ya)=3/10=0.3$$

$$P(n=Y \mid \text{Keterangan}=Ya)=7/10=0.7$$

Tabel 18. Hasil perhitungan Probabilitas Atribut n

Atribut n	Keterangan		Probabilitas	
	Tidak	Ya	Tidak	Ya
T	1	3	0.25	0.75
Y	0.3	0.7	0.3	0.7

Setelah atribut dihitung maka langkah selanjutnya adalah kalikan semua nilai. Hasil sesuai dengan data X yang dicari class-nya. Berikut ini merupakan perhitungannya :

Tabel 19. Hasil sesuai dengan data X

No	Data Ke	PROBABILITAS				Tidak * CP (0.285714)	Ya * CP (0.714286)	Hasil
		Tidak	Ya	Tidak	Ya			
1	K-1	0.25*0.3	0.075	0.75*0.6	0.045	0.021429	0.032143	Ya
2	K-2	1*0.3	0.3	0*0.7	0	0.085714	0	Tidak
3	K-3	0.25*0.4=	0.1	0.75*0.6	0.045	0.028571	0.032143	Ya
4	K-4	0.25*0.3	0.075	0.75*0.7	0.0525	0.021429	0.0375	Ya
5	K-5	1*0.5	0.5	0*0.5	0	0.142857	0	Tidak
6	K-6	0.25*0.3	0.075	0*0.7	0	0.021429	0	Tidak
7	K-7	1*0	0	0*0.9	0	0	0	Tidak
8	K-8	0.25*0	0	0.75*1	0.75	0	0.535715	Ya
9	K-9	0.25*0.3	0.075	0.78*0.7	0.546	0.021429	0.39	Ya
10	K-10	0.25*0.2	0.05	0.75*0.8	0.6	0.014286	0.428572	Ya
11	K-11	0.25*0.3	0.075	0.75*0.7	0.525	0.021429	0.375	Ya

12	K-12	0.25*0	0	0.75*1	0.75	0	0.535715	Ya
13	K-13	0.25*0.2	.05	0.75*0.8	0.6	0.014286	0.428572	Ya
14	K-15	0.25*0.3	0.075	0.75*0.7	0.525	0.021429	0.375	Ya

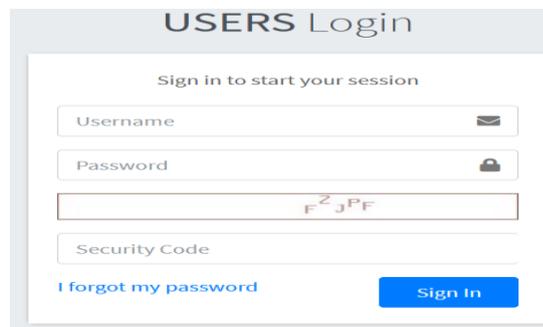
Hasil perbandingan menunjukkan bahwa kinerja algoritma Naive Bayes memiliki tingkat akurasi yang lebih tinggi dibanding algoritma C.45. Hal ini selain studi kasus yang berbeda, karakteristik dari atribut yang diolah juga berbeda. Dan hasil perbandingan yang lebih jelas akan diujicobakan di aplikasi

**3. ANALISA DAN HASIL**

Implementasi sistem adalah sebuah tahapan untuk menerapkan sistem yang telah dirancang dan dibangun. Berikut adalah tampilan sistem yang sudah dirancang dan dibangun yang menerapkan metode *Algoritma C.45 dan Naive Bayes*:

*1. Menu Login*

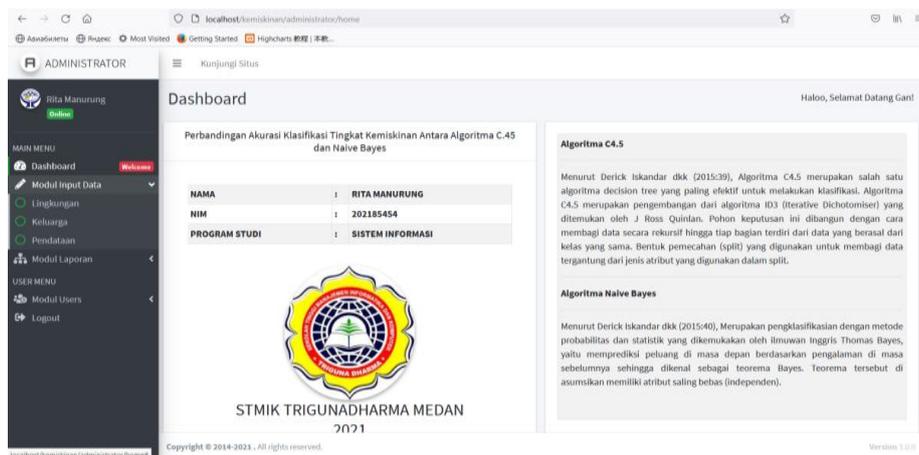
Sebelum masuk kedalam aplikasi maka *user* harus melakukan *login* terlebih dahulu. Dimana *uuser* harus menginput *username* dan *password* terlebih dahulu sesuai dengandatabase



Gambar 1 MenuLogin

*1. Menu Utama*

Halaman menu utama adalah tampilan awal ketika *user* melakukan *login* . Berikut adalah tampilan *menu utama* :



Gambar 2 Menu Utama

Pada bagian ini anda diminta untuk melakukan pengujian dengan sampling data baru dan pada bagian ini anda diminta untuk dapat menguji keakuratan sistem yang anda rancang dengan *tools-tools* yang sudah teruji dan terkalibrasi sebelumnya. Adapun hasil proses program dalam algoritma *Naive Bayes* adalah sebagai berikut.

**Perhitungan Algoritma C.45**

Total Entropy = 0.9666

Perhitungan entropy dan gain						Urutan Gain dari yang terbesar ke terkecil		
Atribut	Nilai	Sum (Nilai)	Sum (Tidak)	Sum (Ya)	Entropy	Gain	Atribut	Gain
A	T	14	4	10	0.86312	0.02038	H	0.36991
	Y	14	7	7	1.00000		K	0.35950
B	T	14	7	7	1.00000	0.04970	N	0.35950
	Y	14	4	10	0.86312		C	0.14594
C	T	8	1	7	0.54356	0.14594	I	0.12803
	Y	20	10	10	1.00000		G	0.06762
D	T	10	4	6	0.97095	-0.00018	F	0.05297
	Y	18	7	11	0.96408		B	0.04970
E	T	12	4	8	0.91830	0.00556	J	0.04514
	Y	16	7	9	0.98870		A	0.02038
F	T	4	1	3	0.81128	0.05297	E	0.00556
	Y	24	10	14	0.97987		M	0.00214
G	T	10	6	4	0.97095	0.06762	D	-0.00018
	Y	18	5	13	0.85241		L	-0.00113
H	T	2	0	2	0.00000	0.36991		
	Y	26	11	15	0.98286			
I	T	7	1	6	0.59167	0.12803		
	Y	21	10	11	0.99836			
J	T	8	2	6	0.81128	0.04514		
	Y	20	9	11	0.99277			
K	T	6	0	6	0.00000	0.35950		
	Y	22	11	11	1.00000			
L	T	5	2	3	0.97095	-0.00113		
	Y	23	9	14	0.96564			
M	T	8	3	5	0.95443	0.00214		
	Y	20	8	12	0.97095			
N	T	6	0	6	0.00000	0.35950		
	Y	22	11	11	1.00000			

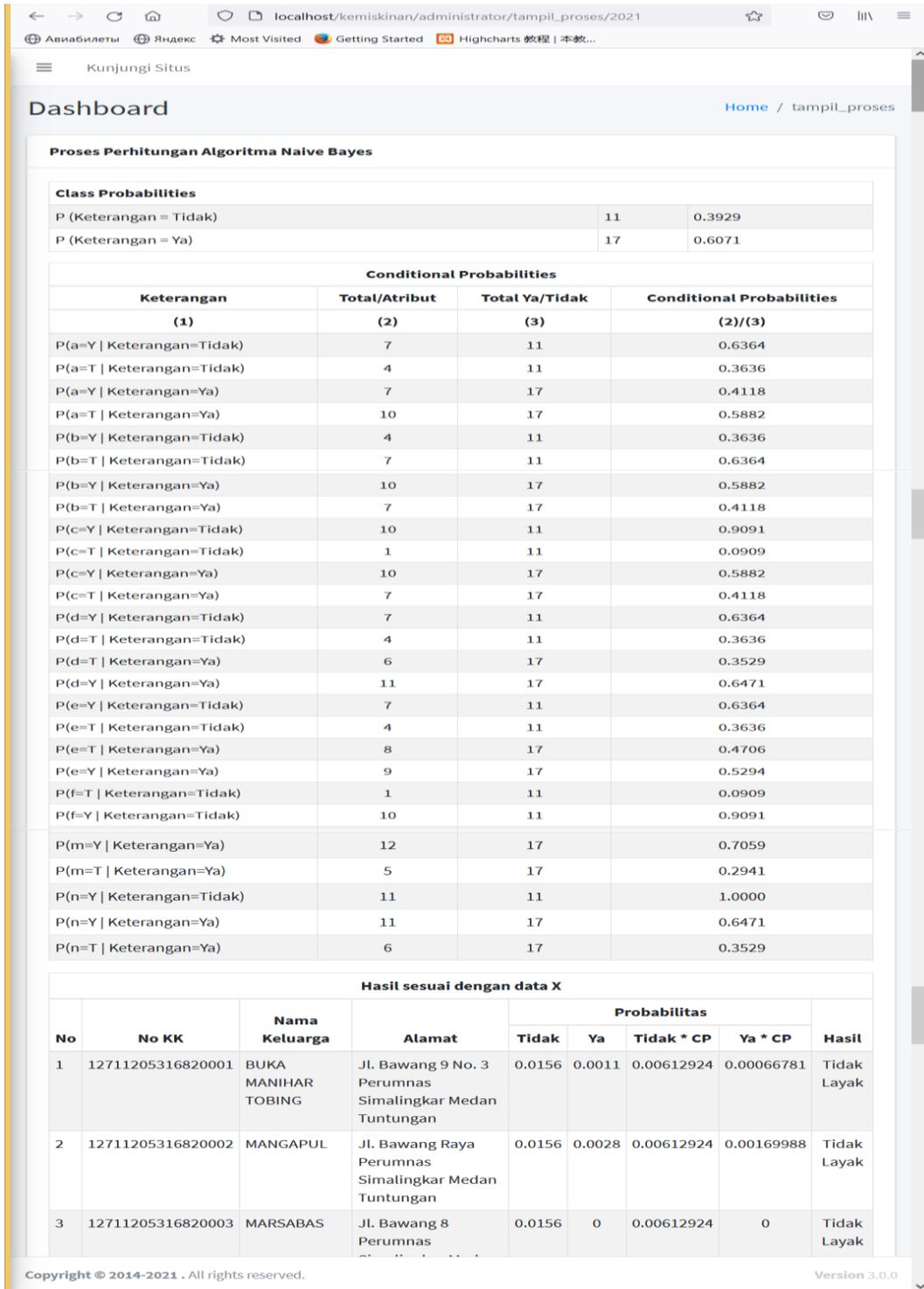
**Rule berdasarkan node akhir**

- Jika H = Y MAKA Layak
- Jika H = T dan K = Y Maka Layak
- Jika H = T dan K = T dan N = Y Maka Layak
- Jika H = T dan K = T dan N = T dan C = Y Maka Layak
- Jika H = T dan K = T dan N = T dan C = T dan I = Y Maka Layak
- Jika H = T dan K = T dan N = T dan C = T dan I = T dan G = Y Maka Layak
- Jika H = T dan K = T dan N = T dan C = T dan I = T dan G = T dan F = Y Maka Layak
- Jika H = T dan K = T dan N = T dan C = T dan I = T dan G = T dan F = T dan B = Y Maka Layak
- Jika H = T dan K = T dan N = T dan C = T dan I = T dan G = T dan F = T dan B = T dan J = Y Maka Layak
- Jika H = T dan K = T dan N = T dan C = T dan I = T dan G = T dan F = T dan B = T dan J = T dan A = Y Maka Layak
- Jika H = T dan K = T dan N = T dan C = T dan I = T dan G = T dan F = T dan B = T dan J = T dan A = T dan E = Y Maka Layak
- Jika H = T dan K = T dan N = T dan C = T dan I = T dan G = T dan F = T dan B = T dan J = T dan A = T dan E = T dan M = Y Maka Layak
- Jika H = T dan K = T dan N = T dan C = T dan I = T dan G = T dan F = T dan B = T dan J = T dan A = T dan E = T dan M = T dan D = Y Maka Layak
- Jika H = T dan K = T dan N = T dan C = T dan I = T dan G = T dan F = T dan B = T dan J = T dan A = T dan E = T dan M = T dan D = T dan L = Y Maka Tidak Layak

Urutan Gain dari yang terbesar ke terkecil			
No KK	Nama Keluarga	Alamat	Klasifikasi
12711205316820001	BUKA MANIHAR TOBING	Jl. Bawang 9 No. 3 Perumnas Simalingkar Medan Tuntungan	Layak
12711205316820002	MANGAPUL	Jl. Bawang Raya Perumnas Simalingkar Medan Tuntungan	Layak
12711205316820003	MARSABAS	Jl. Bawang 8 Perumnas Simalingkar Medan Tuntungan	Tidak Layak
12711205316820004	BELDIMAN	Jl. Tembakau Raya Perumnas Simalingkar Medan Tuntungan	Layak
12711205316820005	PITER	Jl. Sagu Perumnas Simalingkar Medan Tuntungan	Layak
12711205316820006	MARINGAN	Jl. Jahe Perumnas Simalingkar Medan Tuntungan	Layak

Copyright © 2014-2021. All rights reserved. Version 3.0.0

Gambar 14. Tampilan Halaman proses Algoritma C.45



Gambar 15. Tampilan Halaman proses Algoritma Naive Bayes

#### 4. KESIMPULAN

Berdasarkan hasil analisa dari permasalahan yang terjadi dengan kasus yang di bahas tentang pengklasifikasian penduduk/keluarga miskin dengan menerapkan metode *Decision tree C4.5* dan *Naive Bayes Clasifier (NBC)* terhadap sistem yang dirancang dan dibangun maka dapat ditarik kesimpulan sebagai berikut

1. Dari seluruh proses yang dilakukan pada sistem yang dibangun dengan 28 data *sample* didapatkan sebuah hasil dengan Algoritma C4.5 yaitu 27 data layak (96,43 %) dan 1 data tidak layak (3,57 %) sedangkan dengan menggunakan Algoritma Naive bayes yaitu 7 data layak (25 %) dan 21 data tidak layak (75 %).

2. Berdasarkan Pengujian yang telah dilakukan didapatkan sebuah kesimpulan bahwa algoritma C4.5 lebih akurat dalam Klasifikasi Data dibandingkan Naive Bayes, namun untuk Time Taken Naive Bayes lebih cepat dalam Kalasifikasi Data

3. Untuk merancang Perbandingan Akurasi Klasifikasi Tingkat Kemiskinan Antara Algoritma C.45 dan Naive Bayes yaitu dengan merancang *Use Case* diagram, *Activity* Diagram, *Class* Diagram, *Flowchart* program kemudian merancang basis data dan *interface* dimana dalam merancang *Use Case* dan *Activity* dilakukan dengan merancang setiap *Form* yang ada.

#### UCAPAN TERIMA KASIH

Terima Kasih diucapkan kepada kedua orang tuaserta keluarga yang selalu memberi motivasi, Doa dan dukungan moral maupun materi, serta pihak-pihak yang telah mendukung dalam proses pembuatan jurnal ini yang tidak dapat disebutkan satu persatu. Kiranya jurnal ini bisa memberi manfaat bagi pembaca dan dapat meningkatkan kualitas jurnal selanjutnya.

#### REFERENSI

- [1] Prof. Dr. Ir. Marimin, Msc, Ir. Hendri Tanjung, M.M., M.Ag, Haryono Prabowo, S.P., M.M, Sistem Informasi Manajemen Sumber Daya Manusia, Edisi ke I, Bogor : Grasindo, 2006
- [2] Janner Simarmata, et al, Konsep Dasar Manajemen Sistem Informasi, Edisi 1, Medan, Sumatera Utara : Yayasan Kita Menulis, 2020
- [3] Hutahean, Japerson, Konsep Sistem Informasi, Edisi 1, Yogyakarta : Budi Utama, 2015
- [4] Albert Jakobus Chandra, "Perancangan Chabot Menggunakan Dialogflow Natural Language Processing ," Jurnal Media Informatika Budidarma, Vol.4, No. 1, Januari 2020, pp.210.
- [5] Tata Sutabri, Konsep Sistem Informasi, Edisi 1, Yogyakarta : Andi, 2012
- [6] Frinces Ariwisanto Sianturi, et al, Data Mining, Teori Aplikasi Weka, Edisi 1, Kabanjahe, Sumatera Utara : CV. Rudang Mayang, 2019

#### BIBLIOGRAFI PENULIS

	<p><b>Nama Lengkap</b> : Rita Manurung  <b>NIRM</b> : 2017020447  <b>Tempat/Tgl.Lahir</b> : Lumban Huala, 23 Februari 1997  <b>Jenis Kelamin</b> : Perempuan  <b>Alamat</b> : Desa Lumban Huala, Kec Parmakasian  <b>No/Hp</b> : 0852-6171-6643  <b>Email</b> : ritamanurung997@gmail.com  <b>Program Keahlian</b> : Pemmograman Berbasis Web</p>
	<p><b>Nama Lengkap</b> : Puji Sari Ramadhan, S.Kom., M.Kom.  <b>NIDN</b> : 0126039201  <b>Tempat/Tgl.Lahir</b> :  <b>Jenis Kelamin</b> : Laki-Laki  <b>No/Hp</b> : 08116332227  <b>Email</b> : pujisariramadhan@gmail.com  <b>Pendidikan</b> : - D3 – STMIK Triguna Dharma  - S1 – STMIK Triguna Dharma  - S2 – Universitas Putra Indonesia Yptk Padang  <b>Bidang Keahlian</b> : Ilmu Komputer, Sistem Informasi</p>
	<p><b>Nama Lengkap</b> : Moch Iswan Perangin-angin, S.Kom, M.Kom  <b>NIDN</b> : 0120118902  <b>Tempat/Tgl.Lahir</b> :  <b>Jenis Kelamin</b> : Laki-laki  <b>No/Hp</b> : 0813-6252-2792  <b>Email</b> : mochammadiswan@gmail.com  <b>Pendidikan</b> : - S1 – Sekolah Tinggi Teknik Poliprofesi  - S2 – Universitas Sumatera Utara  <b>Bidang Keahlian</b> : Ilmu Kecerdasan Buatan , Manajemen Informatika</p>