

# Klasifikasi Penerima Bantuan Rumah Tidak Layak Huni Desa Labansari Menggunakan Algoritma C4.5

Ahmad Zaelani<sup>1</sup>Universitas Buana Perjuangan Karawang  
Karawang, Indonesia  
ifl18.ahmadzaelani@mhs.ubpkarawang.ac.idAyu Ratna Juwita<sup>2</sup>Universitas Buana Perjuangan Karawang  
Karawang, Indonesia  
ayurj@ubpkarawang.ac.idTohirin Al Mudzakir<sup>3</sup>Universitas Buana Perjuangan Karawang  
Karawang, Indonesia  
tohirin@ubpkarawang.ac.id

**Abstract** — Program bantuan rumah tidak layak huni merupakan program bantuan sosial untuk meringankan keluarga yang berpenghasilan rendah dalam membangun rumah yang layak huni. Data calon penerima bantuan rumah tidak layak huni yang digunakan pada penelitian ini sebanyak 111 data. Tidak akuratnya penyaluran bantuan rumah tidak layak huni, karena tidak ada metode dalam menentukan penerima bantuan tersebut. Penyaluran bantuan yang tidak tepat sasaran akan berdampak pada pembangunan rumah yang tidak selesai. Maka untuk memperkecil kesalahan dalam mengambil keputusan, data diklasifikasi menggunakan algoritma C4.5. Perhitungan algoritma C4.5 menghasilkan nilai akurasi, presisi, *recall* dan pohon keputusan dari data yang diolah. Pengujian dilakukan menggunakan excel mendapatkan akurasi 100% presisi 100% dan *recall* 100%. Pengujian menggunakan bahasa pemrograman python mendapatkan akurasi 100%, presisi 100% dan *recall* 100%.

**Kata kunci** — algoritma C4.5, klasifikasi, bantuan sosial.

## I. PENDAHULUAN

Rumah yang layak huni menjadi keinginan setiap keluarga, namun masyarakat yang berpenghasilan rendah tidak mampu membangun rumah yang layak huni. Oleh sebab itu salah satu upaya pemerintah memperbaiki rumah masyarakat yang tidak layak huni dengan program bantuan sosial rumah tidak layak huni atau rutilahu. Program rutilahu merupakan pemerataan pembangunan disetiap Desa, sasaran dari program rutilahu ditunjukan kepada masyarakat berpenghasilan rendah atau masyarakat yang mempunyai penghasilan terbatas.

Pemerintah Desa Labansari tidak mampu bekerja sendiri maka dari itu partisipasi dari masyarakat juga penting dalam program rutilahu, karena masyarakat itu sendiri akan terlibat baik secara langsung maupun tidak langsung dalam suksesnya program rutilahu. Penyaluran bantuan rutilahu selama ini tidak menggunakan metode apapun untuk menentukan penerima bantuan rutilahu, sehingga penyaluran bantuan rutilahu yang diberikan pemerintah tidak tepat sasaran. Dalam hal ini pemerintah Labansari harus mengambil keputusan keluarga mana yang layak untuk mendapatkan bantuan rutilahu sesuai dengan kriteria dan peraturan pemerintah. Dampak dari penyaluran bantuan rutilahu yang tidak tepat sasaran berimbas pada nilai akurasi dari penyaluran bantuan rutilahu yang tidak bagus. Untuk memperkecil kesalahan dalam mengambil keputusan maka data diklasifikasi menggunakan algoritma C4.5 sebagai metode klasifikasi. Algoritma C4.5 menghasilkan nilai akurasi, presisi, *recall* dan pohon keputusan berdasarkan data yang diolah. Pengolahan data dilakukan menggunakan excel dan bahasa pemrograman python. Semoga dengan adanya pengolahan data menggunakan metode algoritma C4.5 mampu membantu pemerintah Desa Labansari dalam mengambil keputusan.

## II. TINJAUAN PUSTAKA

### A. Algoritma C4.5

Algoritma C4.5 merupakan pohon yang memiliki struktur dimana terdapat simpul untuk medeskripsikan atribut yang akan diuji dan kelas digambarkan dengan daun [1]. Algoritma C4.5 satu dari banyak metode klasifikasi yang representasi struktur pohon, setiap simpul direpresentasikan sebagai atribut dan setiap daun direpresentasikan sebagai kelas [2]. Teknik klasifikasi menghasilkan data baru pada klasifikasi dengan memanipulasi data klasifikasi sehingga menghasilkan sejumlah aturan baru. Algoritma C4.5 diciptakan oleh J. Rose Quinlan yang merupakan pengembangan dari algoritma ID3 [3].

*Entropy* merupakan ukuran informasi yang berisikan teori untuk mengetahui *impurity* dan *homogeneity* berdasarkan karakteristik dari data yang terkumpul [4]. Perhitungan pertama *entropy* yaitu menghitung semua total nilai yang ada dan dilanjutkan dengan nilai disetiap atribut [5]. Perhitungan *entropy* menggunakan rumus seperti dibawah ini.

$$Entropy(S) = \sum_{i=1}^n -p_i * \log^2 p_i \quad (1)$$

Keterangan

S : Himpunan Kasus

n : Jumlah partisi S

p<sub>i</sub>: Proporsi dari S<sub>i</sub> terhadap S

Informasi *gain* merupakan informasi berdasarkan hasil perhitungan nilai *entropy* pada setiap atribut atribut, baik secara observasi maupun disimpulkan berdasarkan set data [4]. Informasi *gain* menggunakan metode seleksi fitur dengan diperoleh nilai akurasi yang baik [6]. Perhitungan *gain* menggunakan rumus seperti dibawah ini:

$$Gain(S,A) = Entropy(S) \sum_{n=1}^i \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Keterangan:

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|S<sub>i</sub>| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

## B. Python

Python merupakan bahasa pemrograman berorientasi objek dinamis, mudah dimengerti dan bisa untuk berbagai macam keperluan [7]. Python merupakan bahasa pemrograman tingkat tinggi (*high level programming language*) yang diciptakan oleh Guido Van Rossum pada tahun 1990 di Belanda [8]. Bahasa pemrograman python berbasis *open source* sehingga bisa digunakan secara gratis yang didukung dengan *developer productivity*, *componen integration*, *software quality* dan *program portability* [8].

Bahasa pemrograman python bersifat multiguna yang berfokus keterbacaan setiap kode, python mampu menyatukan kemampuan di setiap sintaks kode karena didukung dengan pustaka yang komprehensif [9]. Alur pengkodean python cenderung lebih sedikit oleh karena itu banyak orang yang menyukai bahasa pemrograman python, penulisannya pun lebih sedikit dan lebih cepat. Fitur yang dimiliki python cukup lengkap [10].

## C. Confusion matrix

*Confusion matrix* merupakan metode dalam menghitung akurasi yang dihasilkan diperhitungan data mining [11]. Pengujian dilakukan untuk mengetahui tingkat akurasi, presisi dan *recall* perhitungan dan memastikan perhitungan yang dilakukan benar dan sesuai [12]. Akurasi untuk mengetahui jumlah klasifikasi yang benar berdasarkan perhitungan total klasifikasi, presisi untuk mengidentifikasi perkiraan jumlah kasus *positif* yang benar dan *recall* untuk mengetahui kasus *positif* dan diidentifikasi benar. Hasil dari perhitungan dianalisa dan divalidasi secara manual, nilai dari hasil perhitungan sama atau tidak dengan nilai *confusion matrix*. Nilai dari *confusion matrix* terbentuk dari *true positif* dan *true negatif* mampu menggambarkan kinerja klasifikasi dari algoritma C4.5 [13]. Berapa banyak yang mendapatkan bantuan sosial rutilahu dan *true negatif* didapatkan dari berapa banyak yang tidak mendapatkan bantuan sosial rutilahu. Adapun matriks dan rumus akurasi, presisi dan *recall* ada dibawah ini:

Tabel 2. 1 *Confusion matrix*

Prediksi	Fakta	
	<i>Positif</i>	<i>Negatif</i>
<i>Positif</i>	<i>True positif (TP)</i>	<i>False Positif (FN)</i>
<i>Negatif</i>	<i>False Negatif (FN)</i>	<i>True negatif (TN)</i>

### 1) Akurasi

Akurasi merupakan perbandingan kasus data yang diklasifikasikan benar dengan jumlah seluruh data kasus. Perhitungan nilai akurasi dengan cara menjumlahkan nilai *True positif (TP)* dengan *True negatif (TN)* dan dibagi dengan jumlah data keseluruhan seperti rumus dibawah ini.

$$Akurasi = \frac{TP + TN}{TP + FN + FP + TN} \quad (3)$$

### 2) Presisi

Presisi merupakan tingkat kedekatan kasus *positif* yang benar. Nilai presisi didapatkan dengan membagi nilai *True positif (TP)* dengan *True positif (TP)* ditambah *False positif (FP)* seperti persamaan dibawah ini.

$$Presisi = \frac{TP}{TP + FP} \quad (4)$$

### 3) Recall

*Recall* merupakan seberapa banyak data *positif* yang teridentifikasi benar. Nilai *recall* dihitung dengan membagi *True positif (TP)* dengan *True positif (TP)* ditambah dengan *True negative (TN)* seperti persamaan dibawah ini.

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

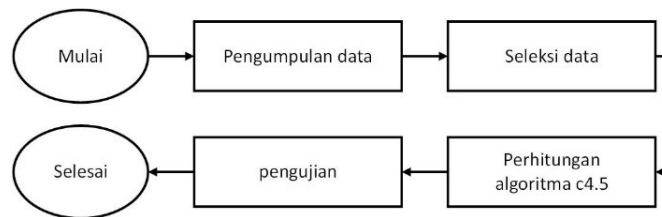
### III. METODE PENELITIAN

#### A. Bahan penelitian

Bahan penelitian tugas akhir ini adalah *dataset* calon penerima bantuan sosial rumah tidak layak huni (rutilahu) di Desa Labansari Kecamatan Cikarang Timur Kabupaten Bekasi. Data disetujui oleh Kepala Desa Labansari yang diberikan melalui Kepala seksi Kesejahteraan pada tanggal 4 Desember tahun 2021 dikantor Desa Labansari berupa *softfile* dengan *format* excel. Data bantuan sosial rutilahu didapatkan sebanyak 111 dengan parameter atribut nama, jenis kelamin, alamat, status kepemilikan, penghasilan, kondisi atap, kondisi lantai, kondisi dinding dan keputusan. Parameter atribut yang akan digunakan pada penelitian ini seperti penghasilan, kondisi atap, kondisi lantai kondisi dinding dan keputusan.

#### B. Prosedur Penelitian

Penelitian dimulai dari pengumpulan data yang didapatkan dari kantor Desa Labansari, menyeleksi data dengan memilih data yang akan diproses, menghitung *entropy* dan *gain* dengan aplikasi excel menggunakan rumus algoritma C4.5. Perhitungan kedua menggunakan bahasa pemrograman python. Setelah kedua perhitungan tersebut selesai maka dilakukan pengujian. Alur prosedur penelitian dilakukan seperti gambar dibawah ini.



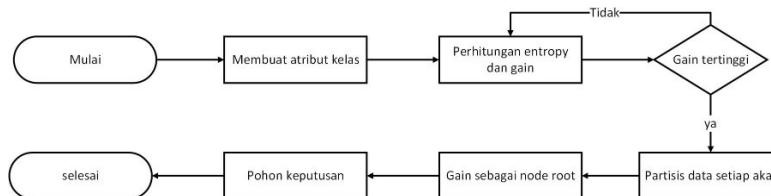
Gambar 3. 1 Alur penelitian

#### C. Seleksi Data

Seleksi data yaitu pemilihan data yang akan dipakai dalam proses mining data dengan tidak menghilangkan keaslian dari data tersebut. Atribut yang digunakan sebanyak 5 parameter yaitu penghasilan, kondisi atap, kondisi lantai, kondisi dinding dan keputusan.

#### D. Perhitungan Algoritma C4.5

Perhitungan Algoritma C4.5 menggunakan dua perhitungan yang berbeda, data yang dihitung menggunakan jumlah data yang sama dengan metode perhitungan menggunakan aplikasi excel dan bahasa pemrograman python.



Gambar 3. 2 perhitungan algoritma C4.5

#### E. Pengujian

Pengujian dilakukan untuk mengetahui seberapa akurat klasifikasi data yang dilakukan oleh algoritma C4.5. Pengujian ditinjau berdasarkan tiga aspek yaitu akurasi, presisi dan *recall* berdasarkan dari perhitungan algoritma C4.5 pada excel dan pemrograman python.

### IV. HASIL DAN PEMBAHASAN

#### A. Pengumpulan Data

Pengumpulan data rutilahu didapatkan dari kantor Desa Labansari pada tanggal 4 Desember 2021 dengan meminta izin kepada Kepala Desa untuk dijadikan bahan penelitian tugas akhir. Data yang didapatkan merupakan pengajuan baru yang akan direalisasikan tahun 2022 dengan jumlah data sebanyak 111 data dan memiliki 8 parameter atribut yaitu nama, alamat, jenis kelamin, status kepemilikan, penghasilan, kondisi atap, kondisi lantai, kondisi dinding dan keputusan.

#### B. Seleksi Data

Data yang didapatkan tidak bisa untuk dihitung oleh karena itu data harus diseleksi terlebih dahulu dengan cara melakukan pembersihan data yang tidak perlu dan penghilangan atribut yang tidak perlu. Data yang sudah diseleksi bisa dilihat pada tabel 4. 1 dengan parameter atribut penghasilan, kondisi atap, kondisi lantai, kondisi dinding dan keputusan sebagai label prediksi.

Tabel 4. 1 Data rutilahu

No	Penghasilan	Kondisiatap	Kondisi Lantai	Kondisi Dinding	Keputusan
1	Tidak Berpenghasilan	Rangka Rusak	Sebagian Terlepas	Retak	Layak
2	Rendah	Rangka Rusak	Sedikit Terlepas	Retak	Layak
3	Rendah	Rangka Rusak	Sebagian Terlepas	Retak	Layak
4	Rendah	Bocor	Sedikit Terlepas	Retak	Layak
5	Rendah	Bocor	Sedikit Terlepas	Terkelupas	Tidak Layak
6	Tidak Berpenghasilan	Rangka Rusak	Sebagian Terlepas	Retak	Layak
7	Tidak Berpenghasilan	Rangka Rusak	Sebagian Terlepas	Retak	Layak
...	.....	.....	.....	.....	.....
111	Tidak Berpenghasilan	Rangka Rusak	Sebagian Terlepas	Terkelupas	Tidak Layak

## C. Perhitungan Algoritma C4.5 menggunakan Excel

Perhitungan menggunakan aplikasi excel dilakukan untuk menghitung nilai *entropy* dan *gain* dari setiap atribut. Perhitungan tersebut akan menghasilkan pohon keputusan sebagai hasil dari perhitungan algoritma C4.5. Perhitungan *entropy* total merupakan perhitungan pertama pada algoritma C4.5. Perhitungan pada gambar 4. 1 menghasilkan nilai *gain* tertinggi pada atribut kondisi atap dengan nilai *gain* 0,280243373. Karena atribut kondisi atap memiliki nilai *gain* tertinggi maka atribut kondisi atap akan menjadi *node* pertama.

Tabel 4. 2 Perhitungan algoritma C4.5

Atribut	Nilai	Jumlah	LAYAK	TIDAK LAYAK	Entropy	Gain
<b>Total</b>	Total	111	51	60	0,995252549	
<b>PENGHASILAN</b>	Rendah	71	46	25	0,935940715	0,200708689
	Tidak Berpenghasilan	40	5	35	0,543564443	
<b>KONDISI ATAP</b>	Rangka Rusak	53	41	12	0,77170947	0,280243373
	Bocor	58	10	48	0,66319684	
<b>KONDISI LANTAI</b>	Sebagian Terlepas	54	23	31	0,984109528	0,003097066
	Sedikit Terlepas	57	28	29	0,999777967	
<b>KONDISI DINDING</b>	Retak	41	32	9	0,759275785	0,182857926
	Terkelupas	70	19	51	0,843507086	

## 1) Perhitungan Entropy

a) Nilai *entropy* total

*Entropy* (total)

$$= \left( -\frac{51}{111} * \text{imlog}_2 \left( \frac{51}{111} \right) \right) + \left( -\frac{60}{111} * \text{imlog}_2 \left( \frac{60}{111} \right) \right)$$

$$= 0,280243373$$

b) Nilai *entropy* penghasilan

*Entropy* (rendah)

$$= \left( -\frac{46}{71} * \text{imlog}_2 \left( \frac{46}{71} \right) \right) + \left( -\frac{25}{71} * \text{imlog}_2 \left( \frac{25}{71} \right) \right)$$

$$= 0,935940715$$

*Entropy* (tidak berpenghasilan)

$$= \left( -\frac{5}{40} * \text{imlog}_2 \left( \frac{5}{40} \right) \right) + \left( -\frac{35}{40} * \text{imlog}_2 \left( \frac{35}{40} \right) \right)$$

$$= 0,543564443$$

c) Nilai *entropy* kondisi atap

*Entropy* (rangka rusak)

$$= \left( -\frac{41}{53} * \text{imlog}_2 \left( \frac{41}{53} \right) \right) + \left( -\frac{12}{53} * \text{imlog}_2 \left( \frac{12}{53} \right) \right)$$

$$= 0,77170947$$

*Entropy* (bocor)

$$= \left( -\frac{10}{58} * \text{imlog}_2 \left( \frac{10}{58} \right) \right) + \left( -\frac{48}{58} * \text{imlog}_2 \left( \frac{48}{58} \right) \right)$$

$$= 0,66319684$$

d) Nilai *entropy* kondisi lantai

*Entropy* (sebagian terlepas)

$$= \left( -\frac{23}{54} * \text{imlog}_2 \left( \frac{23}{54} \right) \right) + \left( -\frac{31}{54} * \text{imlog}_2 \left( \frac{31}{54} \right) \right)$$

$$= 0,984109528$$

Entropy (sedikit terlepas)

$$= \left( -\frac{28}{59} * \log_2 \left( \frac{28}{59} \right) \right) + \left( -\frac{29}{59} * \log_2 \left( \frac{29}{59} \right) \right)$$

$$= 0,999777967$$

e) Nilai *entropy* kondisi dinding

Entropy (retak)

$$= \left( -\frac{32}{41} * \log_2 \left( \frac{32}{41} \right) \right) + \left( -\frac{9}{41} * \log_2 \left( \frac{9}{41} \right) \right)$$

$$= 0,759275785$$

Entropy (terkelupas)

$$= \left( -\frac{19}{70} * \log_2 \left( \frac{19}{70} \right) \right) + \left( -\frac{51}{70} * \log_2 \left( \frac{51}{70} \right) \right)$$

$$= 0,843507086$$

## 2) Perhitungan Gain

a) Nilai *gain* penghasilan

Gain (penghasilan)

$$= 0,995252549 - \left( \left( \frac{71}{111} \right) * 0,935940715 \right) + \left( \left( \frac{39}{111} \right) * 0,543564443 \right)$$

$$= 0,200708689$$

b) Nilai *gain* kondisi atap

Gain (kondisi atap)

$$= 0,995252549 - \left( \left( \frac{53}{111} \right) * 0,77170947 \right) + \left( \left( \frac{58}{111} \right) * 0,66319684 \right)$$

$$= 0,280243373$$

c) Nilai *gain* kondisi lantai

Gain (kondisi lantai)

$$= 0,995252549 - \left( \left( \frac{54}{111} \right) * 0,984109528 \right) + \left( \left( \frac{57}{111} \right) * 0,999777967 \right)$$

$$= 0,003097066$$

d) Nilai *gain* kondisi dinding

Gain (kondisi dinding)

$$= 0,995252549 - \left( \left( \frac{40}{111} \right) * 0,759275785 \right) + \left( \left( \frac{70}{111} \right) * 0,843507086 \right)$$

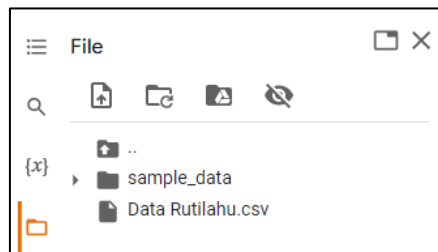
$$= 0,182857926$$

## D. Perhitungan algoritma C4.5 menggunakan python

Perhitungan sistem menggunakan bahasa pemrograman python dengan *library* yang digunakan seperti pandas, scikitlearn dan graphviz. Perhitungan klasifikasi termasuk kedalam *supervised learning* karena telah diberi label keputusan. Data yang dilatih untuk perhitungan sistem sama banyaknya dengan perhitungan manual sebanyak 111 data dengan atribut penghasilan, kondisi atap, kondisi lantai, kondisi dinding dan keputusan

### 1) Upload Dataset

Upload *dataset* ke google colab dengan *format file* csv seperti pada gambar dibawah ini.



Gambar 4. 1 Dataset yang sudah di upload

### 2) Deklarasi Library

Pendeklarasian *library* merupakan pemanggilan *library* yang digunakan untuk melakukan perhitungan.

```
#deklarasi library
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn import tree
from sklearn.metrics import classification_report
from sklearn.metrics import accuracy_score
from sklearn.metrics import precision_score
from sklearn.metrics import recall_score
import graphviz
import pandas as ps
```

Gambar 4. 2 Deklarasi *library*

## 3) Slicing Dataset

*Slicing dataset* merupakan teknik dalam memilih data dari dataset.

```
#slicing dataset
x = dataset.iloc[:, :-1].values
y = dataset.iloc[:, :-1].values
```

Gambar 4. 3 *Slicing* dataset

## 4) Klasifikasi Pohon Keputusan

Klasifikasi pohon keputusan merupakan pembuatan model pohon keputusan berdasarkan algoritma C4.5.

```
#proses klasifikasi
model = tree.DecisionTreeClassifier(random_state=0, max_depth=None,
                                   min_samples_split=2, min_samples_leaf=1,
                                   min_weight_fraction_leaf=0,
                                   max_leaf_nodes=None,
                                   min_impurity_decrease=0)
```

Gambar 4. 4 Klasifikasi pohon keputusan

## 5) Melatih Model

Melatih model data merupakan membangun pohon keputusan berdasarkan variabel model.

```
#melatih model
clf = model.fit(x, y)
```

Gambar 4. 5 Melatih model

## 6) Visualisasi Pohon Keputusan

Visualisasi digunakan untuk merepresentasikan pohon keputusan. Proses visualisasi ditunjukkan pada gambar 4. 6.

```
#visualisasi pohon keputusan
dot_data = tree.export_graphviz(model, out_file = None,
                                feature_names = dataset.columns[:-1],
                                max_depth = 4, filled = True)

[18] graph = graphviz.Source(dot_data)

[19] graph.view('pohon keputusan')

'pohon keputusan.pdf'
```

Gambar 4. 6 Visualisasi pohon keputusan

## 7) Prediksi Hasil Tes

Prediksi hasil tes yaitu untuk memprediksi nilai dari label data yang dilatih, predict hanya dapat menerima satu argumen dari data yang diuji berdasarkan data yang ada di variabel x diproses *slicing dataset*.

```
#prediksi hasil tes
y_pred = clf.predict(x)
```

Gambar 4. 7 Prediksi hasil tes

## E. Pengujian

Pengujian dataset yang telah diolah menggunakan dua metode yang berbeda, pengujian perhitungan excel diuji menggunakan *confusion matrix* dengan *tools* weka dan pengujian pemrograman python diuji menggunakan *library* scikitlearn.

1) Pengujian *confusion matrix*Tabel 4. 3 *confusion matrix*

Prediksi	Fakta	
	Layak	Tidak Layak
Layak	TF=51	FN=0
Tidak Layak	FP=0	TN=60

## a) Nilai Akurasi

$$\begin{aligned}
 \text{Akurasi} &= \frac{51 + 60}{51 + 60 + 0 + 0} * 100\% \\
 &= \frac{111}{111} * 100\% \\
 &= 100\%
 \end{aligned}$$

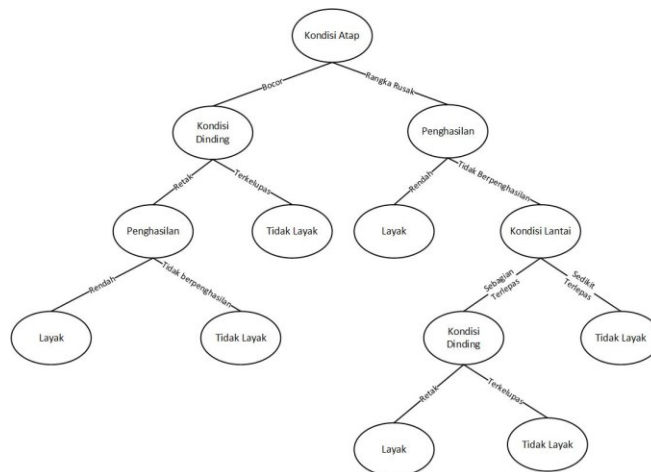
## b) Nilai Presisi

$$\begin{aligned}
 \text{Presisi} &= \frac{51}{51 + 0} * 100\% \\
 &= \frac{51}{51} * 100\% \\
 &= 100\%
 \end{aligned}$$

## c) Nilai Recall

$$\begin{aligned}
 \text{Recall} &= \frac{51}{51 + 0} * 100\% \\
 &= \frac{51}{51} * 100\% \\
 &= 100\%
 \end{aligned}$$

Berdasarkan pengujian diatas diketahui akurasi dari perhitungan menggunakan excel memiliki akurasi sebesar 100 persen, presisi sebesar 100 persen dan recall sebesar 100 persen. Dari perhitungan yang sudah dilakukan menghasilkan pohon keputusan sebagai berikut.



Gambar 4. 8 pohon keputusan perhitungan excel

## 2) Pengujian python

## a) Nilai akurasi

Perintah `accuracy_score` digunakan untuk menghitung akurasi dari prediksi yang benar dari variabel `y` dan `y_pred`.

```
#nilai akurasi
akurasi = accuracy_score(y, y_pred)
print('nilai akurasi : %d persen'%(akurasi*100))

nilai akurasi : 100 persen
```

Gambar 4. 9 Nilai akurasi

## b) Nilai presisi

Presisi merupakan rasio jumlah *positif* benar dan *positif* palsu dalam mengklasifikasi label *positif* dilabel *negatif*.

```
#nilai presisi
presisi = precision_score(y, y_pred)
print('nilai presisi : %d persen'%(presisi*100))

nilai presisi : 100 persen
```

Gambar 4. 10 Nilai presisi

## c) Nilai recall

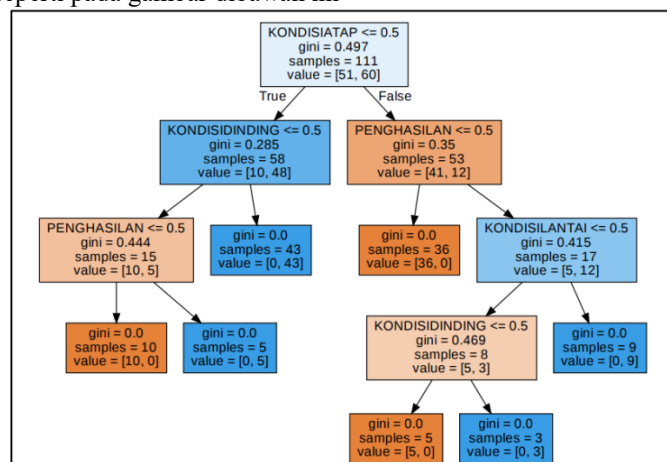
*Recall* merupakan kemampuan klasifikasi dalam menemukan seluruh sampel *positif*.

```
#nilai recall
recall = recall_score(y, y_pred)
print('nilai recall : %d persen'%(recall*100))

nilai recall : 100 persen
```

Gambar 4. 11 Nilai *recall*

Berdasarkan perhitungan yang dilakukan oleh sistem menggunakan bahasa pemrograman python menghasilkan nilai akurasi sebesar 100 persen, nilai presisi 100 persen dan nilai recall sebesar 100 persen. Dari perhitungan sistem yang telah dilakukan menghasilkan pohon keputusan seperti pada gambar dibawah ini



Gambar 4. 12 Pohon keputusan

## V. KESIMPULAN

Klasifikasi data penerima bantuan sosial rumah tidak layak huni atau rutilahu menggunakan algoritma C4.5 dengan cara melakukan perhitungan *entropy* untuk pengujian atribut dan melakukan perhitungan *gain* untuk mencari informasi nilai *gain* tertinggi. Nilai *gain* tertinggi akan menjadi sebagai *node* pertama dari pohon keputusan sampai semua data terpartisi maka pohon keputusan akan terbentuk. Perhitungan menggunakan excel dengan mengolah data sebanyak 111 data didapatkan nilai akurasi sebesar 100 persen nilai presisi 100 persen dan nilai *recall* 100 persen. Perhitungan menggunakan bahasa pemrograman python dengan mengolah data sebanyak 111 didapatkan nilai akurasi sebesar 100 persen nilai presisi 100 persen dan nilai *recall* 100 persen.

## PENGAKUAN

Naskah ilmiah ini sebagian dari penelitian tugas akhir milik Ahmad Zaelani yang berjudul Penerapan Algoritma C4.5 Untuk Klasifikasi Penerima Bantuan Rumah Tidak Layak Huni Desa Labansari. Dengan bimbingan oleh Ayu Ratna Juwita dan Tohirin Al-Mudzakir.

## DAFTAR PUSTAKA

- [1] H. Widayu, S. D. Nasution, N. Silalahi, and Mesran, "Data Mining Untuk Memprediksi Jenis Transaksi Nasabah Pada Koperasi Simpan Pinjam Dengan Algoritma C4.5," *Media Inform. Budidarma*, vol. Vol 1, No, no. 2, p. 37, 2017.
- [2] C. Anam and H. B. Santoso, "Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa," *J. Ilm. Ilmu-Ilmu Tek.*, vol. 8, no. 1, pp. 13–19, 2018, [Online]. Available: <https://ejournal.upm.ac.id/index.php/energy/article/view/111/449>
- [3] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017, doi: 10.22202/ei.2016.v2i2.1465.
- [4] M. F. Arifin and D. Fitriana, "Penerapan Algoritma Klasifikasi C4.5 Dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus : PT Atria Artha Persada," *InComTech*, vol. 8, no. 2, pp. 87–102, 2018, doi: 10.22441/incomtech.v8i1.2198.
- [5] M. Muhamad, A. P. Windarto, and S. Suhada, "Penerapan Algoritma C4.5 Pada Klasifikasi Potensi Siswa Drop Out," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 1–8, 2019, doi: 10.30865/komik.v3i1.1688.
- [6] M. Muqorobin, K. Kusrini, and E. T. Luthfi, "Optimasi Metode Naive Bayes Dengan Feature Selection Information Gain Untuk Prediksi Keterlambatan Pembayaran Spp Sekolah," *J. Ilm. SINUS*, vol. 17, no. 1, p. 1, 2019, doi: 10.30646/sinus.v17i1.378.
- [7] R. K. Ngantung and M. A. I. Pakereng, "Model Pengembangan Sistem Informasi Akademik Berbasis User Centered Design Menerapkan Framework Flask Python," *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 1052, 2021, doi: 10.30865/mib.v5i3.3054.
- [8] M. D. Gumilar, F. Sembiring, and A. Erfina, "Implementasi Progressive Web App Pada Sistem Informasi E-Learning Untuk," *J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. Vol.10, No, pp. 187–196, 2021.
- [9] D. N. Zuraidah, M. F. Apriyadi, A. R. Fatoni, M. Al Fatih, and Y. Amrozi, "Menelisik Platform Digital Dalam Teknologi Bahasa Pemrograman," *Teknois J. Ilm. Teknol. Inf. dan Sains*, vol. 11, no. 2, pp. 1–6, 2021, doi: 10.36350/jbs.v11i2.107.
- [10] M. W. Prihatmono and A. F. Watratan, "Implementasi Algoritma C4.5 Menggunakan Python Untuk Klasifikasi Kepuasan Konsumen," *Progres*, pp. 49–55, 2019, [Online]. Available: <https://jurnal.stmikprofesional.ac.id/index.php/Progress/article/view/146/22>
- [11] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform.*, vol. 5, no. 2, pp. 697–711, 2021.
- [12] A. Jeklin, "evaluasi dan prediksi penguasaan bahasa inggris maritim menggunakan metode decision tree dan confusion matrix (studi kasus di universitas maritim amni)," no. July, pp. 1–23, 2017.
- [13] B. P. Pratiwi, A. S. Handayani, and S. Sarjana, "Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi Wsn Menggunakan Confusion Matrix," *J. Inform. Upgris*, vol. 6, no. 2, pp. 66–75, 2021, doi: 10.26877/jiu.v6i2.6552.