

# Implementasi Data Mining menggunakan Algoritma C-45 pada Data Masyarakat Kecamatan Garoga untuk Menentukan Pola Penerima Beras Raskin

Juan Prima Gultom<sup>1</sup>, Alex Rikki<sup>2</sup>

<sup>1</sup> STIKOM MEDAN, Jl. Jamin Ginting No.285, Padang Bulan, Medan Baru, Kota Medan, Sumatera Utara, Indonesia

<sup>2</sup> Universitas Imelda Medan, Jl. Bilal Ujung No.24,52, Pulo Brayon Darat I, Kec. Medan, Sumatera Utara, Indonesia

## ARTICLE INFORMATION

Received: Februari, 20, 2020

Revised: Maret, 2, 2020

Available online: April, 9 2020

## KEYWORDS

Data Mining, Penerima beras raskin ,  
Algoritma C4.5

## CORRESPONDENCE

Phone: +6282275847123

E-mail: alexrikisinaga@gmail.com

## A B S T R A K

Kecamatan Garoga memiliki Jumlah Desa Sebanyak 13 desa. Kesenjangan taraf kehidupan masyarakat diwilayah Kecamatan Garoga sampai saat ini belum merata, masih banyak masyarakat miskin. Program bantuan sosial salah satu komponen program jaminan sosial yang menjadi bentuk ekspresi tanggung jawab pemerintah, pemerintah daerah yang sangat peduli terhadap kondisi masyarakat yang miskin dan terlantar. Algoritma C4.5 merupakan algoritma klasifikasi pohon keputusan yang banyak digunakan karena memiliki kelebihan utama dari algoritma yang lainnya. Tujuan dari penelitian ini untuk mengetahui faktor mana saja yang paling dominan mempengaruhi data masyarakat kecamatan garoga untuk menentukan pola penerima beras raskin dengan tepat dan akurat, sehingga ini menjadi pedoman bagi pihak masyarakat ke depan untuk meningkatkan penerimaan beras raskin bagi masyarakat.

## PENDAHULUAN

Program bantuan sosial merupakan salah satu komponen program jaminan sosial yang menjadi bentuk ekspresi tanggung jawab pemerintah, pemerintah daerah yang sangat peduli terhadap kondisi masyarakat yang miskin dan terlantar diatas akar rumput (*grass root level*). Program ini merupakan implementasi undang-undang dasar 1945 pasal 34 ayat 1 yang menyatakan bahwa fakir miskin dipelihara oleh Negara [1]. Program bantuan sosial ini telah dilaksanakan dari tahun ke tahun pada Kecamatan Garoga, Tahun 2015 penerima beras raskin untuk Kecamatan Garoga berjumlah 1555 kk, dan pada Tahun 2017 penerima bantuan beras raskin sebanyak 1442 kk. Namun program tidak sesuai pelaksanaannya dikarenakan dalam pelaksanaan pemberian bantuan beras raskin ini tidak tepat sasaran, Salah sasaran ini banyak disebabkan oleh *human error*, di mana para petugas lapangan justru membagi-bagikan kupon raskin pada keluarga dekat atau teman kerabatnya. Bahkan tidak sedikit keluarga sejahtera yang "menagih jatah" beras murah tersebut [2].

Dengan menggunakan data mining, setiap kumpulan atau gudang data dapat memberikan pengetahuan penting yang menjadi informasi yang sangat berharga bagi suatu pemerintahan, seperti pada pemerintahan. Pada sebuah pemerintahan, suatu sistem informasi dapat digunakan untuk memperoleh informasi yang menunjang setiap kegiatan pada pengambilan suatu keputusan. Data mining algoritma C4.5 dapat digunakan untuk menyusun sistem yang mempunyai kemampuan melihat pola penerimaan beras raskin, untuk selanjutnya bisa menjadi strategi dalam proses pengambilan keputusan untuk penerimaan beras raskin[3], [4]. Pada penelitian ini peneliti menggunakan algoritma C4.5 dalam menentukan masyarakat yang mendapat beras raskin di Kecamatan Garoga.

Algoritma C4.5 merupakan algoritma klasifikasi pohon keputusan yang banyak digunakan karena memiliki kelebihan utama dari algoritma yang lainnya. Kelebihan algoritma C4.5 dapat menghasilkan pohon keputusan yang mudah diinterpretasikan, memiliki tingkat akurasi yang dapat diterima, efisien dalam menangani atribut bertipe diskret dan dapat menangani atribut bertipe diskret dan numerik[5], [6]. Dalam mengkonstruksi pohon, algoritma C4.5 membaca seluruh sampel data *training* dari *storage* dan memuatnya ke memori. Hal inilah yang menjadi salah satu kelemahan algoritma C4.5 dalam kategori "skalabilitas" adalah algoritma ini hanya dapat digunakan jika data training dapat disimpan secara keseluruhan dan pada waktu yang bersamaan di memori Data training yang akan digunakan oleh peneliti adalah data masyarakat Kecamatan Garoga[7].

## METODE PENELITIAN

Metode yang digunakan dalam melakukan penelitian adalah melakukan pengumpulan data dengan menetapkan lokasi lokasi / objek Penelitian di kantor Kecamatan Garoga yang berlokasi di Jl. Parluasan Garoga Kecamatan Garoga Kabupaten Tapanuli Utaras[8].

### 2.1. Data yang diperlukan

Data yang diperlukan adalah data calon penerima beras raskin pada Kecamatan Garoga dengan kriteria-kriteria yang ditetapkan oleh manajemen, antara lain sebagai berikut:

- a. Pekerjaan Suami
- b. Pekerjaan Istri

- c. Status Rumah
- d. Pendapatan

## 2.2. Metode Pengumpulan Data

Dalam penulisan Proposal skripsi ini, penulis mencari dan mengumpulkan data dan informasi yang berhubungan dalam penelitian ini dengan menggunakan beberapa tahap sebagai berikut :

Tahap I : Perencanaan

Pemilihan Judul, Perumusan Masalah dan menentukan Batasan Masalah

Tahap II : Analisa

Melakukan penelitian lapangan dengan menggunakan metode sebagai berikut ini:

1. Wawancara (Interview)

Melakukan percakapan langsung dengan pimpinan, dan pihak yang bersangkutan guna memperoleh penjelasan tentang sistem penerima beras raskin yang dipakai.

2. Pengamatan (Observation)

Dengan dilaksanakannya pengamatan langsung terhadap objek yang sedang diteliti maka penulis dapat lebih leluasa mengetahui apa sebenarnya yang menjadi kendala terhadap sistem yang sedang berjalan pada Kecamatan Garoga, sehingga diketahui pula hal-hal apa saja yang diperlukan untuk menyempurnakan sistem yang sedang berjalan tersebut.

Tahap III: Perancangan

Untuk tahap ini penulis mengumpulkan data-data dengan metode sebagai berikut :

1. Kepustakaan (*Library Research*)

Melakukan penelitian kepustakaan untuk mendapatkan landasan bersifat teori dengan mempelajari literatur dari catatan selama mengikuti perkuliahan di STIKOM Medan, contoh-contoh skripsi di perpustakaan STIKOM Medan, Jurnal-jurnal yang berkaitan serta buku-buku referensi lainnya yang berhubungan dengan penulisan penelitian ini.

2. Laboratorium (*Laboratory Research*)

a. Pengujian data yang dilakukan dengan menggunakan data-data yang dikumpulkan dan mengimplementasikan algoritma C-45

b. Metode pencarian melalui internet, yaitu mencari beberapa informasi yang dibutuhkan sebagai bahan yang sesuai untuk membantu penulisan penelitian dari internet

Tahap IV: Pengujian dan Implementasi

Pada tahap pengujian dan implementasi penulis menggunakan software Tanagra.

Tahap V: Penulisan Laporan

Pada tahap penulisan laporan penulis membuat laporan.

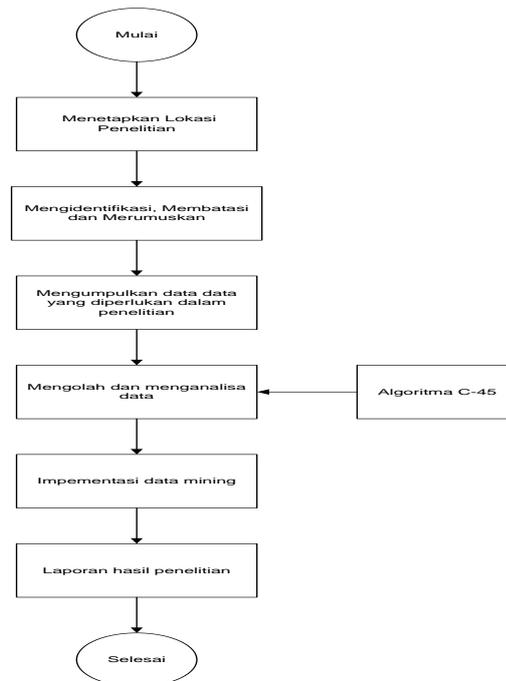
## 2.3. Metode Analisa Data

Adapun jenis penelitian yang digunakan dalam penelitian ini adalah dengan metode penelitian deskriptif kualitatif, yaitu dengan menerapkan pendekatan kualitatif yang menghasilkan data deskriptif berupa kata-kata tertulis maupun lisan dari subjek yang diamati. Teknik yang digunakan adalah observasi dan wawancara mendalam untuk menggali kebutuhan pengguna untuk mendapatkan model sistem yang mudah serta dapat digunakan dan menemukan data apa saja yang diolah atau data apa yang dihasilkan oleh sistem. Kemampuan proses analisa mencakup :

1. Kemampuan menganalisa konsep yang abstrak serta mengatur kembali kedalam pembagian logika dan menghasilkan pemecahan masalah berdasarkan pembagiannya.
2. Mampu menyerap fakta atau informasi.
3. Mampu berkomunikasi baik dalam bentuk tulisan dan lisan.
4. Mampu menerapkan algoritma C-45

## 2.4. Langkah dan Diagram Alir

Gambar berikut ini menunjukkan alir langkah penelitian yang dilakukan oleh penulis.



Gambar 1. Diagram Alir

## HASIL DAN PEMBAHASAN

Proses pengujian dilakukan untuk menentukan penerima beras raskin Kecamatan Garoga menggunakan algoritma C4.5. Algoritma C4.5 digunakan untuk dapat menentukan yang tepat untuk masing-masing input ialah pekerjaan suami, pekerjaan istri, status rumah, pendapatan, dan yang menjadi outputnya berupa pendapatan. Implementasi proses Algoritma C4.5 pada aplikasi Tanagra 1.4.50 mencakup spesifikasi hardware dan software yang dibutuhkan [9]. Berhubungan dengan aplikasi yang digunakan dalam pengujian adalah aplikasi yang menggunakan salah satu database Microsoft Excel.

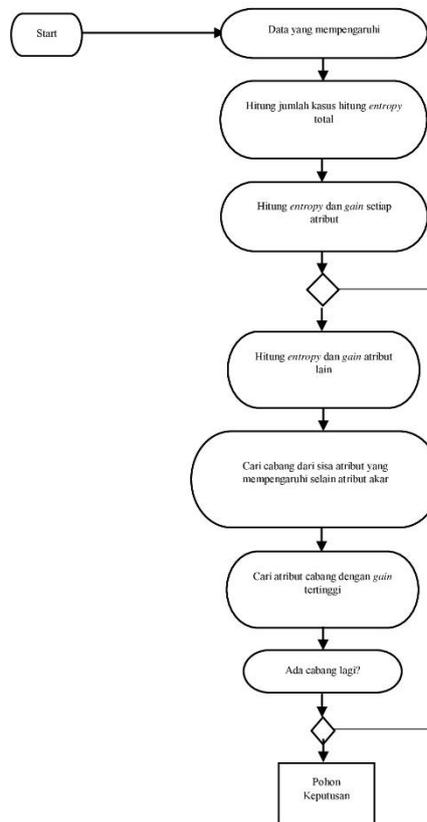
Dalam melakukan analisa pengambilan data yang dibutuhkan adalah data penerima beras raskin. data penerima beras raskin diambil dari data Kecamatan Garoga yang telah dibuat dan disebar sebelumnya. Atribut/variabel yang digunakan untuk dijadikan input adalah pekerjaan suami, pekerjaan istri, status rumah, pendapatan, dan yang menjadi outputnya berupa pendapatan. Adapun data-data Penerima Beras Raskin seperti pada tabel 1 di bawah ini:

Tabel 1. Sampel Data Penduduk

NO	Nama Keluarga	Pekerjaan Suami	Pekerjaan Istri	Status Rumah	Pendapatan	Status Rastra
1	Sastro-Amuoyo-Sipahutar-Nurhaida-Simanjuntak	Petani	Petani	Kontrakan	2000000	TIDAK
2	Anggiat-Siagian-Tiurma-Ritonga	Petani	Petani	kontrakan	2000000	TIDAK
3	Olo-Munte-Marlise-Pasaribu	Petani	Ibu-rumah-tangga	Kontrakan	2000000	YA
4	Berton-Pane-Asni-Siagian	Petani	Petani	Kontrakan	2000000	YA
5	Parnigotan-Hutabarat-Margaretta-Pakpahan	Polisi	Ibu-rumah-tangga	Rumah-Sendiri	3500000	TIDAK
6	Mangasi-Pakpahan-Riama-Pasaribu	Wiraswasta	PNS	Rumah-Sendiri	5000000	TIDAK
7	Tolopan-Hasibuan-Rusmia-Ritonga	Wiraswasta	PNS	Rumah-Sendiri	5000000	TIDAK
8	Khrisman-Hasibuan-Rincan-Ritonga	Wiraswasta	Ibu-rumah-tangga	Rumah-Sendiri	3000000	TIDAK
9	Esbon-Lubis-Masda-Hasugian	PNS	Ibu-rumah-tangga	Rumah-Sendiri	3500000	TIDAK
10	Riduan-Sipahutar-Mida-Ritonga	Petani	Petani	Rumah-Sendiri	2000000	YA
11	Ardon-Sipahutar-Rospita-Pasaribu	Wiraswasta	Petani	Rumah-Sendiri	2500000	YA
12	Bangun-Gultom-Derliana-Siregar	Wiraswasta	Ibu-rumah-tangga	Rumah-Sendiri	2500000	YA

NO	Nama Keluarga	Pekerjaan Suami	Pekerjaan Istri	Status Rumah	Pendapatan	Status Rastra
13	Irwan-Ritonga-Tumiari-Sinaga	Wiraswasta	Ibu-rumah-tangga	Kontrakan	2000000	YA
14	Siddik-Hasibuan-Mulianti	Wiraswasta	Ibu-rumah-tangga	Kontrakan	2500000	YA
15	Partuaan-Simanjuntak-Tiro-Hutajulu	Petani	Ibu-rumah-tangga	Kontrakan	2000000	YA
16	Rencong-Pane-Elvy-Mariany-Lubis	Petani	Petani	Kontrakan	2000000	YA
17	Ojahan-Siregar-Tiarma-Hasibuan	Polisi	Ibu-rumah-tangga	Rumah-Sendiri	3500000	TIDAK
18	Johannes-Lumbantobing-Mastina-Gultom	Wiraswasta	PNS	Rumah-Sendiri	5000000	TIDAK
19	Bunga-Hasibuan-Berta-Lubis	PNS	PNS	Rumah-Sendiri	6000000	TIDAK
20	Anton-Ritonga-Melda-Pandiangan	Wiraswasta	Ibu-rumah-tangga	kontrakan	2500000	YA

Metode yang digunakan penulis pada penelitian ini adalah metode Data Mining Algoritma C4.5 dimana proses utama adalah segmentasi atau penerima beras raskin yang terdapat pada Kecamatan Garoga. dan pada tahap ini penulis menseleksi data penerima beras raskin. Berikut ini merupakan diagram flowchart dari algoritma C4.5 dengan asumsi bahwa parameter input adalah pekerjaan suami, pekerjaan istri, status rumah dan pendapatan sesuai dengan penelitian yang dilakukan penulis seperti pada gambar 2 di bawah ini:



Gambar 2. Activity Diagram

### 3.1. Proses Pengklasifikasian Data Mining

Setelah data penduduk dikumpulkan, tahap selanjutnya ialah memproses data yang telah terkumpul tersebut. Dalam memproses data mining ada beberapa tahap seperti data yang dilakukan terlebih dahulu yaitu data cleaning, data selection, dan data transformation yang akan diolah menggunakan algoritma C4.5.

1. Data Cleaning  
Data cleaning untuk membersihkan data, menghapus data duplikat, menghilangkan data yang tidak penting.
2. Data selection

Dalam menentukan selection data dapat memilih data yang akan dijadikan atribut/vaiabel dari data yang ada yaitu data penerima beras raskin untuk dijadikan pohon keputusan adalah:

- a. Pekerjaan Istri
  - b. Pekerjaan Suami
  - c. Status Rumah
  - d. Pendapatan
3. Data transformasi

Data transformasi ialah data dengan mengubah format data asli menjadi data yang bisa dihitung oleh algoritma C4.5. Berikut ini adalah tabel data transformasi penerima beras raskin:

Tabel 2. Transformasi Data Penduduk

NO	Keluarga	Pekerjaan Suami	Pekerjaan Istri	Status Rumah	Pendapatan	Status Rastra
1	Sastro-Amuoyo-Sipahutar-Nurhaida-Simanjuntak	Petani	Petani	kontrakan	2000000	TIDAK
2	Anggiat-Siagian-Tiurma-Ritonga	Petani	Petani	kontrakan	2000000	TIDAK
3	Olo-Munte-Marlise-Pasaribu	Petani	Ibu-rumah-tangga	kontrakan	2000000	YA
4	Berton-Pane-Asni-Siagian	Petani	Petani	kontrakan	2000000	YA
5	Parnigotan-Hutabarat-Margaretta-Pakpahan	Polisi	Ibu-rumah-tangga	Rumah-Sendiri	3500000	TIDAK
6	Mangasi-Pakpahan-Riama-Pasaribu	Wiraswasta	PNS	Rumah-Sendiri	5000000	TIDAK
7	Tolopan-Hasibuan-Rusmia-Ritonga	Wiraswasta	PNS	Rumah-Sendiri	5000000	TIDAK
8	Khrisman-Hasibuan-Rincan-Ritonga	Wiraswasta	Ibu-rumah-tangga	Rumah-Sendiri	3000000	TIDAK
9	Esbon-Lubis-Masda-Hasugian	PNS	Ibu-rumah-tangga	Rumah-Sendiri	3500000	TIDAK
10	Riduan-Sipahutar-Mida-Ritonga	Petani	Petani	Rumah-Sendiri	2000000	YA
11	Ardon-Sipahutar-Rospita-Pasaribu	Wiraswasta	Petani	Rumah-Sendiri	2500000	YA
12	Bangun-Gultom-Derliana-Siregar	Wiraswasta	Ibu-rumah-tangga	Rumah-Sendiri	2500000	YA
13	Irwan-Ritonga-Tumiar-Sinaga	Wiraswasta	Ibu-rumah-tangga	kontrakan	2000000	YA
14	Siddik-Hasibuan-Mulianti	Wiraswasta	Ibu-rumah-tangga	kontrakan	2500000	YA
15	Partuaan-Simanjuntak-Tiro-Hutajulu	Petani	Ibu-rumah-tangga	kontrakan	2000000	YA
16	Rencong-Pane-Elvy-Mariany-Lubis	Petani	Petani	kontrakan	2000000	YA
17	Ojahan-Siregar-Tiurma-Hasibuan	Polisi	Ibu-rumah-tangga	Rumah-Sendiri	3500000	TIDAK
18	Johannes-Lumbantobing-Mastina-Gultom	Wiraswasta	PNS	Rumah-Sendiri	5000000	TIDAK
19	Bunga-Hasibuan-Berta-Lubis	PNS	PNS	Rumah-Sendiri	6000000	TIDAK
20	Anton-Ritonga-Melda-Pandiangan	Wiraswasta	Ibu-rumah-tangga	kontrakan	2500000	YA

Setelah melakukan data transformasi lalu membuat pohon keputusan dengan menghitung entropy total entropy masing-masing atribut dengan menghitung gain dan menentukan gain tertinggi. Dalam memilih atribut yang akan dijadikan sebagai, atribut akar, berdasarkan nilai gain tertinggi dari atribut yang ada. Menghitung entropy digunakan rumus sebagai berikut:

$$Entropy(S) = \sum_{i=1}^n (-P_i * \log_2 P_i)$$

Dimana:

S : Himpunan Kasus

N : Jumlah partisi S

Pi : Proporsi

Sedangkan mencari nilai gain tertinggi dengan rumus:

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

Dimana:

S : Himpunan Kasus

A : Atribut

N : Jumlah partisi atribut -A

|Si| : Jumlah kasus pada atribut -A

|S| : Jumlah seluruh kasus

Dengan menggunakan dua rumus diatas dapat dibuat tahap-tahap penyelesaian yang akan didapat entropy dan gain yang membentuk pohon keputusan.

Langkah selanjutnya adalah menentukan nilai entropy dan gainnya:

Langkah-langkah algoritma C4.5 dalam melakukan proses data mining adalah sebagai berikut:

1. Mengelompokkan setiap variabel untuk mendapatkan attribute yang akan digunakan untuk menghitung entropy dari kasus agar mendapatkan akar dan cabang dari PENERIMA BERAS RASKIN . Dalam tahap ini juga tahap klasifikasi pada data mining dilakukan, dimana data tersebut dikelompokkan dan diklasifikasikan berdasarkan variabel-variabel yang telah ditentukan. Berikut ini adalah klasifikasi data berdasarkan tabel 4.3.

Tabel 3. Klasifikasi Data

	Ket	Jlh Kasus	Iya	Tidak	Entropi	Gain
Total		20	10	10		
Pek. Suami						
	Petani	7	5	2		
	Wiraswasta	9	5	4		
	PNS	2	0	2		
	Polisi	2	0	2		
Pek. Istri						
	IRT	10	6	4		
	Petani	6	4	2		
	PNS	4	0	4		
S. Rumah						
	Kontrakan	9	7	2		
	Rumah Sendiri	11	3	8		
Pendapatan						
	2.000.000	8	6	2		
	2.500.000	4	4	0		
	3.000.000	1	0	1		
	3.500.000	3	0	3		
	5.000.000	3	0	3		
	6.000.000	1	0	1		

Berikut ini adalah nilai *entropy*:

- Entropy* total  

$$\text{Entropy total} = \text{Entropy}(S) \sum_{i=1}^n (-P_i * \log_2 P_i)$$

$$= ((-10/20 * \log_2(10/20)) + (-10/20 * \log_2(10/20)))$$

$$= 1$$
- Entropy* Pekerjaan Suami  
 Nilai atribut "Petani"  $= ((-5/7 * \log_2(5/7)) + (-2/7 * \log_2(2/7)))$   
 $= 0.8631$   
 Nilai atribut "Wiraswasta"  $= ((-5/9 * \log_2(5/9)) + (-4/9 * \log_2(4/9)))$   
 $= 0.9910$   
 Nilai atribut "PNS"  $= ((-0/2 * \log_2(0/2)) + (-2/2 * \log_2(2/2)))$   
 $= 0$   
 Nilai atribut "Polisi"  $= ((-0/2 * \log_2(0/2)) + (-2/2 * \log_2(2/2)))$   
 $= 0$
- Entropy* Pekerjaan Istri  
 Nilai atribut "IRT"  $= ((-6/10 * \log_2(6/10)) + (-4/10 * \log_2(4/10)))$   
 $= 0.9709$   
 Nilai atribut "Petani"  $= ((-4/6 * \log_2(4/6)) + (-2/6 * \log_2(2/6)))$   
 $= 0.9182$   
 Nilai atribut "PNS"  $= ((-0/4 * \log_2(0/4)) + (-4/4 * \log_2(4/4)))$   
 $= 0$
- Entropy* Status Rumah  
 Nilai atribut "Kontrakan"  $= ((-7/9 * \log_2(7/9)) + (-2/9 * \log_2(2/9)))$   
 $= 0.7642$   
 Nilai atribut "Rumah Sendiri"  $= ((-3/11 * \log_2(3/11)) + (-8/11 * \log_2(8/11)))$   
 $= 0.8453$
- Entropy* Pendapatan

Nilai atribut "2.000.000" =  $(-6/8 * \log_2(6/8) + (-2/8 * \log_2(2/8)))$   
 = 0.8112  
 Nilai atribut "2.500.000" =  $(-4/4 * \log_2(4/4) + (-0/4 * \log_2(0/4)))$   
 = 0  
 Nilai atribut "3.000.000" =  $(-0/1 * \log_2(0/1) + (-1/1 * \log_2(1/1)))$   
 = 0  
 Nilai atribut "3.500.000" =  $(-0/3 * \log_2(0/3) + (-3/3 * \log_2(3/3)))$   
 = 0  
 Nilai atribut "5.000.000" =  $(-0/3 * \log_2(0/3) + (-3/3 * \log_2(3/3)))$   
 = 0  
 Nilai atribut "6.000.000" =  $(-0/1 * \log_2(0/1) + (-1/1 * \log_2(1/1)))$   
 = 0

2. Menghitung nilai gain dari semua atribut

Berikut ini adalah nilai *gain* dari setiap kriteria.

- a. Nilai *Gain* (Pekerjaan Suami)
 
$$= 1 - ((7/20 * 0.8631) + (9/20 * 0.9910) + (2/20 * 0) + (2/20 * 0))$$

$$= 0.2519$$
- b. Nilai *Gain* (Pekerjaan Istri)
 
$$= 1 - ((10/20 * 0.9709) + (6/20 * 0.9182) + (4/20 * 0))$$

$$= 0.2391$$
- c. Nilai *Gain* (Status Rumah)
 
$$= 1 - ((9/20 * 0.7642) + (11/20 * 0.8453))$$

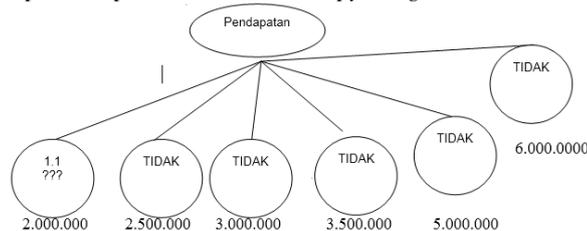
$$= 0.1911$$
- d. Nilai *Gain* (Pendapatan)
 
$$= 1 - ((8/20 * 0.8112) + (4/20 * 0) + (1/20 * 0) + (3/20 * 0) + (3/20 * 0) + (1/20 * 0))$$

$$= 0.6755$$

Tabel 4. Nilai Node 1

	Ket	Jlh Kasus	Iya	Tidak	Entropi	Gain
Total		20	10	10		
Pek. Suami						0.2519
	Petani	7	5	2	0.8631	
	Wiraswasta	9	5	4	0.9910	
	PNS	2	0	2	0	
	Polisi	2	0	2	0	
Pek. Istri						0.2391
	IRT	10	6	4	0.9709	
	Petani	6	4	2	0.9182	
	PNS	4	0	4	0	
S. Rumah						0.1911
	Kontrakan	9	7	2	0.7642	
	Rumah Sendiri	11	3	8	0.8453	
Pendapatan						0.6755
	2.000.000	8	6	2	0.8112	
	2.500.000	4	4	0	0	
	3.000.000	1	0	1	0	
	3.500.000	3	0	3	0	
	5.000.000	3	0	3	0	
	6.000.000	1	0	1	0	

Dari tabel di atas menunjukkan bahwa kriteria Pendapatan memiliki *Gain* tertinggi. Untuk *fase* selanjutnya ialah pembentukan pohon keputusan. Berikut ini adalah pohon keputusan dari nilai *entropy* dan *gain* diatas:



Gambar 3. Node 1

Selanjutnya adalah menyelesaikan untuk menghitung Node 1.1 dengan pendapatan sebagai akar sama seperti cara sebelumnya dengan menghitung nilai Entropy dari atribut yang tersisa yaitu Pekerjaan suami, Pekerjaan istri, Status Rumah setelah dihitung nilai Entropy, kemudian menghitung nilai gain untuk tiap-tiap atribut.

Tabel 5. Node 1.1

	Ket	Jlh Kasus	Iya	Tidak	Entropi	Gain
Total Pendapatan	2.000.000	8	6	2	0.8112	
Pek. Suami						
	Petani	7	5	2		
	Wiraswasta	1	1	0		
	PNS	0	0	0		
	Polisi	0	0	0		
Pek. Istri						
	IRT	3	3	0		
	Petani	5	3	2		
	PNS	0	0	0		
S. Rumah						
	Kontrakan	7	5	2		
	Rumah Sendiri	1	1	0		

## d. Entropy Pekerjaan Suami

$$\begin{aligned} \text{Nilai atribut "Petani"} &= ((-5/7 * \log_2 (5/7)) + (-2/7 * \log_2 (2/7))) \\ &= 0.8631 \end{aligned}$$

$$\begin{aligned} \text{Nilai atribut "Wiraswasta"} &= ((-1/1 * \log_2 (1/1)) + (-0/1 * \log_2 (0/1))) \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{Nilai atribut "PNS"} &= ((-0/0 * \log_2 (0/0)) + (-0/0 * \log_2 (0/0))) \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{Nilai atribut "Polisi"} &= ((-0/0 * \log_2 (0/0)) + (-0/0 * \log_2 (0/0))) \\ &= 0 \end{aligned}$$

## e. Entropy Pekerjaan Istri

$$\begin{aligned} \text{Nilai atribut "IRT"} &= ((-3/3 * \log_2 (3/3)) + (-0/3 * \log_2 (0/3))) \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{Nilai atribut "Petani"} &= ((-3/5 * \log_2 (3/5)) + (-2/5 * \log_2 (2/5))) \\ &= 0.9709 \end{aligned}$$

$$\begin{aligned} \text{Nilai atribut "PNS"} &= ((-0/0 * \log_2 (0/0)) + (-0/0 * \log_2 (0/0))) \\ &= 0 \end{aligned}$$

## f. Entropy Status Rumah

$$\begin{aligned} \text{Nilai atribut "Kontrakan"} &= ((-5/7 * \log_2 (5/7)) + (-2/7 * \log_2 (2/7))) \\ &= 0.8631 \end{aligned}$$

$$\begin{aligned} \text{Nilai atribut "Rumah Sendiri"} &= ((-1/1 * \log_2 (1/1)) + (-0/1 * \log_2 (0/1))) \\ &= 0 \end{aligned}$$

Berikut ini adalah nilai gain dari setiap kriteria.

## e. Nilai Gain (Pekerjaan Suami)

$$\begin{aligned} &= 0.8112 - ((7/8 * 0.8631) + (1/8 * 0) + (0/8 * 0) + (0/8 * 0)) \\ &= 0.0559 \end{aligned}$$

## f. Nilai Gain (Pekerjaan Istri)

$$\begin{aligned} &= 0.8112 - ((3/8 * 0) + (5/8 * 0.9709) + (0/8 * 0)) \\ &= 0.2043 \end{aligned}$$

## g. Nilai Gain (Status Rumah)

$$\begin{aligned} &= 0.8112 - ((7/8 * 0.8631) + (1/8 * 0)) \\ &= 0.0559 \end{aligned}$$

Tabel 6. Hasil Entropy dan Gain Node 1.1

	Ket	Jlh Kasus	Iya	Tidak	Entropi	Gain
Total, Pendapatan	2.000.000	8	6	2	0.8112	
Pek. Suami						0.0559
	Petani	7	5	2	0.8631	
	Wiraswasta	1	1	0	0	
	PNS	0	0	0	0	
	Polisi	0	0	0	0	
Pek. Istri						0.2043
	IRT	3	3	0	0	
	Petani	5	3	2	0.9709	
	PNS	0	0	0	0	
S. Rumah						0.0559

	Ket	Jlh Kasus	Iya	Tidak	Entropi	Gain
	Kontrakan	7	5	2	0.8631	
	Rumah Sendiri	1	1	0	0	

Tabel 7. Hasil Entropy dan Gain Node 1.3

	Ket	Jlh Kasus	Iya	Tidak	Entropi	Gain
Total, Pendapatan, pek. Istri, Status Rumah	Kontrakan	3	1	2	0.9182	
Pek. Suami						0
	Petani	3	1	2	0.9182	
	Wiraswasta	0	0	0		
	PNS	0	0	0		
	Polisi	0	0	0		

Dari tabel diatas tidak bisa dilanjutkan lagi dikarenakan sudah tidak bisa lagi dihitung nilai entropy dan nilai gainnya. Keputusan yang diambil juga sudah akurat tanpa gain yang didapat bernilai 0 ataupun 1.

### 3.2. General Rules

1. Jika pendapatan 2.000.000 pekerjaan istri petani status rumah kontrakan pekerjaan suami petani maka menerima beras raskin.
2. Jika pendapatan 2.000.000 pekerjaan istri IRT status rumah kontrakan pekerjaan suami wiraswasta maka tidak menerima beras raskin
3. .jika pendapatan 2.000.000 pekerjaan istri PNS status rumah kontrakan pekerjaan suami PNS maka tidak menerima beras raskin
4. .jika pendapatan 2.000.000 pekerjaan istri wiraswasta status rumah kontrakan pekerjaan suami polisi maka tidak menerima beras raskin

## KESIMPULAN

Berdasarkan hasil analisis dan percobaan yang dilakukan pada bab sebelumnya, maka kesimpulan yang dapat diambil adalah sebagai berikut:

1. Dengan menentukan pengklasifikasian untuk membentuk pola dengan mengolah data untuk menjadi sebuah decision tree.
2. Algoritma C4.5 dapat dengan mudah memproses data karena mudah beradaptasi pada interaksi kompleks antar variabel. Setiap cabang pohon dapat terdiri atas kombinasi dari variabel dan variabel yang sama dapat muncul lebih dari satu di berbagai cabang pohon. Hal ini dapat berguna dalam mengetahui bagaimana sebuah variabel bergantung kepada lainnya dan bagaimana spesifik variabelnya.
3. Pengujian yang dilakukan oleh data mining dengan aplikasi tanagra sangatlah mudah dalam mendapatkan informasi dari data yang import dengan format .csv dan diuji oleh tanagra

## DAFTAR PUSTAKA

- [1] L. Ramandey, "" Dampak Sosial Program Pembagian Beras Miskin (Raskin) di Kampung Amar Distrik Amar Kabupaten Timika," no. May, 2018, doi: 10.13140/RG.2.2.13516.72326.
- [2] N. Saputra, "Pendistribusian Beras Miskin (Raskin) di Kelurahan Delima Kecamatan Tampan Kota Pekanbaru Tahun 2016," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2013, doi: 10.1017/CBO9781107415324.004.
- [3] E. Ndruru and R. Limbong, "Implementasi Data Mining Dalam Pengelompokan Jurusan yang Diminati Siswa SMK Negeri 1 Lolowa'u menggunakan Metode Clustering | Ndruru | MEANS (Media Informasi Analisa dan Sistem)," *MEANS (Media Informasi Analisa dan Sistem)*, 2018. [Online]. Available: [http://ejournal.ust.ac.id/index.php/Jurnal\\_Means/article/view/273/pdfdsdx11](http://ejournal.ust.ac.id/index.php/Jurnal_Means/article/view/273/pdfdsdx11). [Accessed: 11-Jan-2020].
- [4] J. Han, *Data mining: Data mining concepts and techniques*. 2014.
- [5] E. Buulolo, N. Silalahi, Fadlina, and R. Rahim, "C4.5 Algorithm To Predict the Impact of the Earthquake," *Int. J. Eng. Res. Technol.*, vol. 6, no. 2, pp. 10–15, 2017.
- [6] H. Widayu, S. Darma, N. Silalahi, and Mesran, "Data Mining Untuk Memprediksi Jenis Transaksi Nasabah Pada Koperasi Simpan Pinjam Dengan Algoritma C4.5," *Issn 2548-8368*, vol. Vol 1, No. no. June, p. 7, 2017.
- [7] D. W. Sitohang and A. Rikki, "Implementasi Algoritma K- Means Clustering untuk Mengelompokkan Data Gizi Balita pada Kecamatan Garoga Tapanuli Utara," vol. 02, pp. 80–92, 2019.
- [8] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Bandung: PT Alfabet, 2016.
- [9] S. Haryati, A. SudarsonHaryati, S., Sudarsono, A., & Suryana, E. (2015). Implementasi Data Mining untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5. *Jurnal Media Infotama*, 11(2), 130–138.o, and E. Suryana, "Implementasi Data Mining untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5," *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.