

# Penerapan Metode Data Mining C4.5 dalam Penentuan Kelayakan Rehabilitas Rumah Warga

Aulia Sugarda\*, Saifullah, Jalaluddin, Agus Perdana Windarto, Wendi Robiansyah

Program Studi Sistem Informasi, STIKOM Tunas Bangsa, Pematang Siantar, Indonesia

Email: <sup>1</sup>\*auliasugarda2323@gmail.com, <sup>2</sup>saifullah@amiktunasbangsa.ac.id, <sup>3</sup>jalaluddin@amiktunasbangsa.ac.id,

<sup>4</sup>agus.perdana@amiktunasbangsa.ac.id, <sup>5</sup>wendirobiansyah@amiktunasbangsa.ac.id

Email Penulis Korespondensi: auliasugarda2323@gmail.com

**Abstrak**– Tujuan penelitian untuk mengetahui rumah warga yang layak direhabilitas di Kelurahan Pematang Dolok Kahean. Sumber data yang digunakan penelitian dengan menggunakan dataset yang sudah ada di Kelurahan Pematang Dolok Kahean. Solusi yang diberikan adalah untuk mengklasifikasikan tingkat kelayakan rumah warga dengan menggunakan metode data mining C4.5 serta menggunakan bantuan *software Rapidminer*. Metode ini dipilih karena merupakan salah satu metode pada pohon keputusan yang banyak dimanfaatkan untuk melakukan prediksi terhadap suatu kasus. Hasil penelitian menyebutkan nilai akurasi sistem sebesar 83,33% dengan menggunakan *split validation* dimana metode ini menghasilkan beberapa rule yang dapat digunakan dalam penentuan kelayakan rehabilitas rumah warga sehingga subsidi pemerintah dapat disalurkan secara tepat.

**Kata Kunci:** Metode C4.5; Klasifikasi; Data Mining; Kelayakan Rumah Warga; Pematang Dolok Kahean; *Rapidminer*

**Abstract**– The purpose of the study was to find out which houses deserve to be rehabilitated in Pematang Dolok Kahean Village. The source of the data used in this research is using datasets that already exist in Pematang Dolok Kahean Village. The solution given is to classify the feasibility level of residents' houses using the C4.5 data mining method and using the *Rapidminer software* assistance. This method was chosen because it is one of the most widely used decision tree methods to predict a case. The results of the study stated that the system's accuracy value was 83.33% using split validation where this method produced several rules that could be used in determining the feasibility of the rehabilitation of residents' houses so that government subsidies could be channeled appropriately.

**Keywords:** C4.5; Classification; Data Mining; Feasibility of Residents' Houses; Pematang Dolok Kahean; *Rapidminer*

## 1. PENDAHULUAN

Kemiskinan merupakan masalah sosial terpenting untuk mengetahui kesejahteraan suatu rumahtangga. Karena setiap orang berhak hidup sejahtera mendapatkan lingkungan hidup yang baik serta berhak memperoleh pelayanan kesehatan [1]. Pada Undang-Undang Nomor 01 Tahun 2011 tentang perumahan dan kawasan pemukiman menyebutkan bahwa rumah adalah bangunan gedung yang berfungsi untuk tempat tinggal yang layak huni. Saat ini jumlah masyarakat yang berpenghasilan rendah masih cukup besar, yang mengakibatkan banyak tempat tinggal warga yang termasuk kategori Rumah Tidak Layak Huni (RTLH). Kemiskinan didefinisikan dengan berbagai konsep salah satunya yaitu bank dunia mendefinisikan kemiskinan merupakan suatu ketidakmampuan untuk mencapai kehidupan layak dengan pendapatan \$1,9 perhari [2]. Dengan demikian pihak Kantor Kelurahan Pematang Dolok Kahean Rumah Tidak Layak Huni membuat beberapa kebijakan yang telah dibuat oleh kantor tersebut. Hal ini memiliki ketentuan dalam menerapkan kelayakan rehabilitas sosial rumah tidak layak huni, Penilaian dilakukan dengan melihat rumah warga yang Tidak Layak Huni. Untuk mendapatkan solusi dari permasalahan diatas penulis menggunakan teknik data mining dengan Metode C4.5 [3]–[6] agar mendapatkan solusi yang lebih efisien dalam melakukan penilaian terhadap Rumah Warga Tidak layak Huni. Proses yang semula memiliki banyak kendala terhadap masyarakat menjadi lebih cepat dalam penyelesaian data warga yang mendapatkan Rehabilitas sosial Rumah Warga tidak Layak Huni. Pada metode ini juga dapat menghindari munculnya permasalahan-permasalahan menggunakan kriteria yang jumlahnya lebih sedikit pada node internal tanpa mengurangi kualitas keputusan yang dihasilkan [7]–[10].

Untuk menentukan Kelayakan Rehabilitas Rumah Warga Tidak layak Huni dapat menggunakan hasil klasifikasi data diri yang sesuai. Klasifikasi merupakan proses pencarian model yang dapat membedakan kelas data dengan tujuan agar model tersebut dapat digunakan untuk memprediksi kelas dari suatu obyek yang belum diketahui kelasnya [11]. Salah satu metode yang digunakan untuk membuat pohon keputusan (*decision tree*) adalah Metode C4.5 yang merupakan metode sangat populer dimana digunakan oleh banyak peneliti [12], [13]. Kelebihan Metode C4.5 dapat menghasilkan pohon keputusan yang mudah diinterpretasikan, memiliki tingkat akurasi yang diterima, efisien dalam menangani atributbertipe diskret dan numeric. Penulis telah membaca beberapa penelitian terdahulu sebagai referensi diantaranya yaitu penelitian dilakukan oleh [14] yang berjudul “Penerapan Metode C4.5 untuk menentukan Kelayakan Penerima Bantuan Rehabilitas Sosial Rumah Tidak layak Huni (Studi kasus Di Pemerintahan Kabupaten Solok Selatan)”.

Berdasarkan paparan tersebut diharapkan hasil dari penelitian ini dapat menentukan Kelayakan Rehabilitas Rumah Warga Tidak layak dengan 5 variabel yang dipakai oleh Kantor Kelurahan Pematang Dolok Kahean adalah Status, Tanggungan, Pekerjaan, Bentuk Rumah, dan Lantai. Output yang akan dihasilkan merupakan hasil dari Menentukan Kelayakan rehabilitas Rumah Warga Tidak Layak Huni dengan menggunakan metode Metode C4.5.

## 2. METODOLOGI PENELITIAN

Metode yang digunakan pada penelitian adalah metode C4.5. Dalam pemodelan metode C4.5 dilakukan pada sampel dataset, kemudian menghitung *entropy* ( $S$ ) dari keseluruhan atribut, setelah *entropy* ( $S$ ) ditemukan kemudian menghitung *Gain* dari seluruh atribut, atribut yang memiliki *Gain* yang akan dipergunakan sebagai akar atau node. Selanjutnya buat cabang untuk tiap nilai, bagi kasus dalam cabang, ulangi perhitungan *Gain* sampai semua data masuk kedalam kelas yang sama. Atribut yang telah dipilih tidak lagi dibutuhkan dalam perhitungan, proses pembentukan pohon keputusan terhenti jika sudah tidak ada lagi atribut yang dipartisi dan semua tuple dalam node  $N$  telah memiliki kelas yang sama [15], [16]. Dalam kasus dataset pada penelitian ini terdiri dari 2 kelas keterangan yaitu layak dan tidak layak. Secara umum metode C4.5 untuk membuat pohon keputusan (*decision tree*) [17]–[19] adalah sebagai berikut :

- a) Pilih atribut sebagai akar (*rootnode*).

Atribut akar dapat dilakukan dengan menghitung jumlah kasus dan jumlah target atribut. Nilai *entropy* dihitung untuk menghasilkannilai atribut. Berikut ini adalah Rumus dasar perhitungan *entropy* :

$$Entropy(s) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (1)$$

Dimana :

$S$  : Himpunan Kasus

$n$  : Jumlah Partisi

$p_i$  : Proporsi dari  $S_i$  terhadap  $S$

Setelah menghitung *entropy* setiap kasus selanjutnya menghitung nilai *gain* untuk pemisah objek dengan rumus berikut ini :

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

Dimana :

$S$  : Himpunan kasus

$A$  : Atribut

$n$  : Jumlah partisi atribut  $A$

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

$|S|$  : Jumlah kasus dalam  $S$  Sebelum

Setelah semua atribut dihitung menggunakan persamaan diatas, maka atribut yang lain dijadikan sebagai *node* (akar).

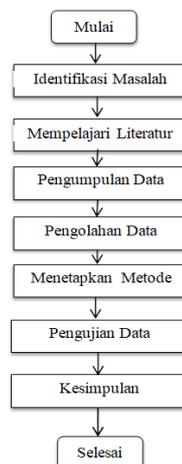
- b) Membuat cabang untuk tiap-tiap nilai.

Setelah diperoleh atribut yang mempunyai *gain* tertinggi maka atribut tersebut digunakan sebagai *node*. *Node* ini memiliki *instance* sehingga *instance* dijadikan sebagai cabang dari *node*.

- c) Membagi kasus dalam cabang.

Nilai *instance* diklasifikasikan berdasarkan makna dari nilai *instance* tersebut agar menjadi lebih sederhana

Dalam menyelesaikan penelitian, sumber dataset yang digunakan adalah observasi secara langsung pada kantor kelurahan pematang dolok kahean dalam menentukan kelayakan rumah warga tidak layak huni. Penelitian ini dilaksanakan pada bulan Maret 2022. Pada rancangan penelitian dibuat perincian atau tahap keseluruhan proses penelitian, mulai dari pengumpulan data hingga memperoleh kesimpulan dan hasil akhir yang didapat, sehingga menghasilkan sebuah keputusan yang dapat menjadi masukan untuk pegawai kantor kelurahan pematang dolok kahean. Pada tahap selanjutnya dataset ditransformasi kedalam format Microsoft excel 2010, yang kemudian akan diproses menggunakan aplikasi *Rapidminer* studio, pengujiannya menggunakan *cross validation* yang ada di dalam *Rapidminer* studio, maka akan terlihat akurasi pohon keputusan dan rulenya. Adapun rancangan alur dari penelitian yang dibuat penulis dapat dilihat pada gambar 1 dibawah ini:



Gambar 1. Alur Rancangan Penelitian

Berdasarkan Gambar 1, keterangan pada rancangan dari penelitian yang dilakukan adalah sebagai berikut :

- a) Identifikasi Masalah.  
Identifikasi masalah merupakan analisa masalah yang sedang terjadi. Dalam hal ini identifikasi masalah terkait dengan menentukan kelayakan rehabilitas sosial rumah warga tidak layak huni di kelurahan pematang dolok kahean.
- b) Mempelajari Literatur.  
Pada penyusunan penelitian banyak mempelajari buku-buku, jurnal untuk dapat mempermudah penulisan penelitian ini.
- c) Pengumpulan Data.  
Pengumpulan dataset dilakukan dengan tujuan memilih data mana yang sesuai dengan permasalahan dan dapat digunakan sebagai atribut dalam proses Algoritma C4.5.
- d) Pengolahan Data.  
Proses pengolahan data dilakukan untuk mendapatkan hasil ataupun informasi yang berguna untuk banyak orang sebagai acuan dalam pengambilan keputusan. Dimana pada proses ini data di input ke dalam Microsoft Excel, kemudian diolah menggunakan Algoritma C4.5.
- e) Menetapkan Metode.  
Metode yang digunakan dalam penelitian ini yaitu Metode data mining C4.5
- f) Pengujian Data.  
Setelah proses pengolah dataset, tahap selanjutnya melakukan pengujian data yang dilakukan dengan menggunakan bantuan *software Rapidminer* sehingga diperoleh hasil perbandingan antara pengolahan data manual dengan aplikasi *software*.
- g) Kesimpulan  
Dari hasil dan pembahasan yang dilakukan maka dapat diperoleh kesimpulan yang dapat menjadi referensi untuk membuat sebuah keputusan dan menjadi referensi untuk penelitian selanjutnya.

### 3. HASIL DAN PEMBAHASAN

Berikut tahapan pengolahan data dengan metode C4.5 untuk memperoleh pohon keputusan dalam menentukan kelayakan rehabilitas rumah warga tidak layak huni sesuai dengan data yang diperoleh. Perhitungan metode C4.5 dimulai dari dengan memilih atribut akar terlebih dahulu dengan mencari jumlah kasus keseluruhan, jumlah kasus keputusan layak dan tidak layak. Menghitung entropy dari semua kasus yang dibagi berdasarkan status, tanggungan, pekerjaan, bentuk rumah, dan lantai. Setelah itu dilakukan perhitungan gain untuk masing-masing atribut. Hasil perhitungan ditunjukkan oleh tabel 1 berikut.

**Tabel 1.** Perhitungan *Node 1*

| Node 1         | Jml kasus | Layak | Tidak Layak | Entropy    | Gain       |
|----------------|-----------|-------|-------------|------------|------------|
| Total          | 40        | 26    | 14          | 0.93406806 |            |
| Status         |           |       |             |            | 0.21583756 |
| Menikah        | 34        | 21    | 13          | 0.95968689 |            |
| Janda          | 6         | 5     | 1           | 0.65002242 |            |
| Tanggungan     |           |       |             |            | 0.00897962 |
| 2 Anak         | 11        | 7     | 4           | 0.9456603  |            |
| 3 Anak         | 16        | 11    | 5           | 0.89603823 |            |
| 4 Anak         | 9         | 6     | 3           | 0.91829583 |            |
| 5 Anak         | 4         | 2     | 2           | 1          |            |
| Pekerjaan      |           |       |             |            | 0.44724394 |
| Serabutan      | 12        | 12    | 0           |            |            |
| Bhl            | 18        | 13    | 4           | 0.82127826 |            |
| Swasta         | 10        | 1     | 9           | 0.46899559 |            |
| Bentuk Rumah   |           |       |             |            | 0.46703403 |
| Semi Permanen  | 20        | 7     | 13          | 0.93406806 |            |
| Tidak Permanen | 20        | 19    | 0           |            |            |
| Lantai         |           |       |             |            | 0.41395489 |
| Semen          | 22        | 8     | 14          | 0.9456603  |            |
| Tanah          | 18        | 18    | 0           |            |            |

Kolom entropy pada baris total dihitung dengan persamaan (1) sebagai berikut:

**Menghitung entropy total :**

$$Entropy [\text{Total}] = \left( -\frac{26}{40} \times \log_2 \left( \frac{26}{40} \right) \right) + \left( -\frac{14}{40} \times \log_2 \left( \frac{14}{40} \right) \right)$$

$$Entropy [\text{Total}] = 0,93406806$$

**Menghitung entropy dan gain Status :**

$$\begin{aligned} \text{Entropy [Status-Menikah]} &= \left(-\frac{21}{34} x \log_2 \left(\frac{21}{34}\right)\right) + \left(-\frac{13}{34} x \log_2 \left(\frac{13}{34}\right)\right) \\ &= 0,95968689 \end{aligned}$$

$$\begin{aligned} \text{Entropy [Status-Janda]} &= \left(-\frac{5}{6} x \log_2 \left(\frac{5}{6}\right)\right) + \left(-\frac{1}{6} x \log_2 \left(\frac{1}{6}\right)\right) \\ &= 0,65002242 \end{aligned}$$

$$\begin{aligned} \text{Gain [Total, Status]} &= 0,93406806 - \left(\frac{34}{40} x 0,9899928\right) + \left(\frac{6}{40} x 0,6500224\right) \\ &= 0,215583756 \end{aligned}$$

**Menghitung entropy dan gain Tanggungan :**

$$\begin{aligned} \text{Entropy [Tanggungan- 2 Anak]} &= \left(-\frac{7}{11} x \log_2 \left(\frac{7}{11}\right)\right) + \left(-\frac{4}{11} x \log_2 \left(\frac{4}{11}\right)\right) \\ &= 0,9456603 \end{aligned}$$

$$\begin{aligned} \text{Entropy [Tanggungan- 3 Anak]} &= \left(-\frac{11}{16} x \log_2 \left(\frac{11}{16}\right)\right) + \left(-\frac{5}{16} x \log_2 \left(\frac{5}{16}\right)\right) \\ &= 0,89603823 \end{aligned}$$

$$\begin{aligned} \text{Entropy [Tanggungan- 4 Anak]} &= \left(-\frac{6}{9} x \log_2 \left(\frac{6}{9}\right)\right) + \left(-\frac{3}{9} x \log_2 \left(\frac{3}{9}\right)\right) \\ &= 0,91829583 \end{aligned}$$

$$\begin{aligned} \text{Entropy [Tanggungan-5 Anak]} &= \left(-\frac{2}{4} x \log_2 \left(\frac{2}{4}\right)\right) + \left(-\frac{2}{4} x \log_2 \left(\frac{2}{4}\right)\right) \\ &= 1 \end{aligned}$$

$$\begin{aligned} \text{Gain [Total, Tanggungan]} &= 0,97099506 - \left(\frac{11}{40} x 0,9940302\right) + \left(\frac{16}{40} x 0,954434\right) + \left(\frac{9}{40} x 0,9182958\right) + \left(\frac{4}{40} x 0,91\right) = \\ &0,00897962 \end{aligned}$$

**Menghitung entropy dan gain Pekerjaan :**

$$\text{Entropy [Pekerjaan-Serabutan]} = 0$$

$$\begin{aligned} \text{Entropy [Pekerjaan-bhl]} &= \left(-\frac{13}{18} x \log_2 \left(\frac{13}{18}\right)\right) + \left(-\frac{4}{18} x \log_2 \left(\frac{4}{18}\right)\right) \\ &= 0,82127826 \end{aligned}$$

$$\begin{aligned} \text{Entropy [Pekerjaan-Swasta]} &= \left(-\frac{1}{10} x \log_2 \left(\frac{1}{10}\right)\right) + \left(-\frac{9}{10} x \log_2 \left(\frac{9}{10}\right)\right) \\ &= 0,46899559 \end{aligned}$$

$$\begin{aligned} \text{Gain [Total, Pekerjaan]} &= 0,9340686 - \left(\frac{12}{40} x 0\right) + \left(\frac{18}{40} x 0,82127826\right) + \left(\frac{10}{40} x 0,46899559\right) = \\ &0,4472243994 \end{aligned}$$

**Menghitung entropy dan gain Bentuk Rumah:**

$$\begin{aligned} \text{Entrop[BentukRumah-SemiPermanen]} &= \left(-\frac{7}{20} x \log_2 \left(\frac{7}{20}\right)\right) + \left(-\frac{13}{20} x \log_2 \left(\frac{13}{20}\right)\right) \\ &= 0,46703403 \end{aligned}$$

$$\text{Entrop[BentukRumah-TidakPermanen]} = 0$$

$$\begin{aligned} \text{Gain [Total,Bentuk Rumah]} &= 0,93406806 - \left(\frac{18}{40} x 0,65002242\right) + \left(\frac{22}{40} x 0\right) \\ &= 0,67703403 \end{aligned}$$

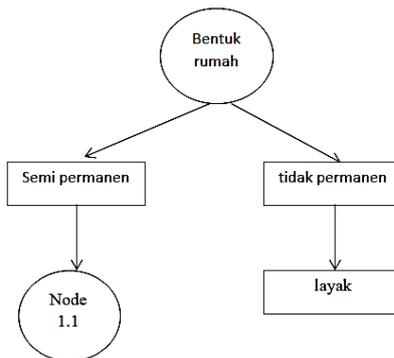
**Menghitung entropy dan gain Lantai :**

$$\begin{aligned} \text{Entropy [Lantai-Semen]} &= \left(-\frac{8}{22} x \log_2 \left(\frac{8}{22}\right)\right) + \left(-\frac{14}{22} x \log_2 \left(\frac{14}{22}\right)\right) \\ &= 0,93406806 \end{aligned}$$

$$\text{Entropy [Lantai-Tanah]} = 0$$

$$\begin{aligned} \text{Gain [Total, Lantai]} &= 0,93406806 - \left(\frac{23}{40} x 0,88654089\right) + \left(\frac{17}{40} x 0\right) = 0,46703403 \end{aligned}$$

Dari perhitungan pada tabel 1 diperoleh atribut *node* (akar) adalah bentuk rumah memiliki gain tertinggi 0,46703403, dimana terdiri dari 2 sub atribut yaitu semi permanen dan tidak permanen. Berdasarkan nilai entropy dari kedua sub atribut diatas, sub atribut semi permanen yang memperoleh keputusan. Maka perlu dilakukan perhitungan lebih lanjut untuk menentukan node akar selanjutnya, dapat digambarkan pohon keputusan dari Tabel 1 sebagai berikut:



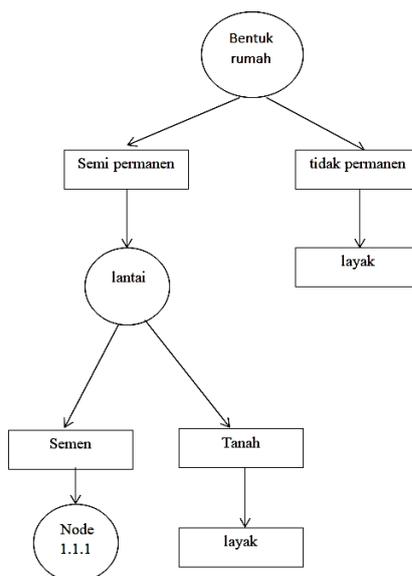
Gambar 2. Pohon Keputusan 1

Untuk perhitungan selanjutnya Algoritma C4.5 dapat ditunjukkan pada Tabel berikut ini :

Tabel 2. Perhitungan Node 1.1

| Node 1.1                   | Jml kasus | Layak | Tidak Layak | Entropy    | Gain       |
|----------------------------|-----------|-------|-------------|------------|------------|
| bentuk rumah-semi permanen | 20        | 7     | 13          | 0.93406806 |            |
| Status                     |           |       |             |            | 0.57083493 |
| Menikah                    | 18        | 6     | 12          | 0.91829583 |            |
| Janda                      | 2         | 1     | 1           | 1          |            |
| Tanggungan                 |           |       |             |            | 0.48302196 |
| 2 anak                     | 6         | 3     | 3           | 1          |            |
| 3 anak                     | 7         | 2     | 5           | 0.86312057 |            |
| 4 anak                     | 4         | 1     | 3           | 0.81127812 |            |
| 5 anak                     | 3         | 1     | 2           | 0.91829583 |            |
| Pekerjaan                  |           |       |             |            | 0.71107594 |
| serabutan                  | 2         | 2     | 0           |            |            |
| bhl                        | 9         | 5     | 4           | 0.99107606 |            |
| swasta                     | 8         | 0     | 8           |            |            |
| Lantai                     |           |       |             |            | 0.72162699 |
| semen                      | 15        | 2     | 13          | 0.56650951 |            |
| tanah                      | 5         | 5     | 0           |            |            |

Dari perhitungan Tabel 2, atribut yang menjadi *node* cabang dari atribut bentuk rumah-semi permanen adalah lantai. Dimana atribut dari lantai yang terdiri dari semen dan tanah. Tanah sudah memperoleh keputusan layak sedangkan untuk sub atribut semen belum memperoleh keputusan maka akan dilakukan perhitungan kembali. Maka dapat digambarkan pohon keputusan dari tabel 2 sebagai berikut:



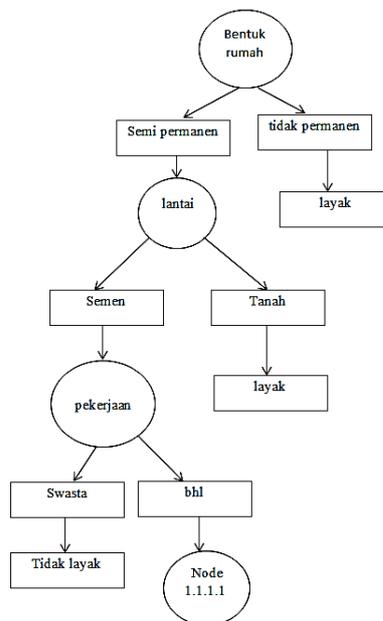
Gambar 3. Pohon keputusan 2

Dalam mencari hasil perhitungan selanjutnya pada *node* akar bentuk rumah-semi permanen dan *node* cabang lantai-semen dapat ditunjukkan pada tabel 4 berikut ini.

**Tabel 3.** Perhitungan *node* 1.1.1

| Node 1.1.1                   | Jml kasus | Layak | Tidak Layak | Entropy    | Gain       |
|------------------------------|-----------|-------|-------------|------------|------------|
| bentuk rumah - semi permanen |           |       |             |            |            |
| lantai - semen               | 15        | 2     | 13          | 0.56650951 |            |
| Status                       |           |       |             |            | 0.8569139  |
| Menikah                      | 13        | 1     | 12          | 0.39124356 |            |
| Janda                        | 2         | 1     | 1           | 1          |            |
| Tanggungjan                  |           |       |             |            | 0.78406806 |
| 2 anak                       | 3         | 0     | 3           |            |            |
| 3 anak                       | 5         | 0     | 1           |            |            |
| 4 anak                       | 4         | 1     | 3           | 0.81127812 |            |
| 5 anak                       | 3         | 1     | 2           | 0.91829583 |            |
| Pekerjaan                    |           |       |             |            | 0.79632368 |
| serabutan                    | 0         | 0     | 0           |            |            |
| bhl                          | 6         | 2     | 4           | 0.91829583 |            |
| swasta                       | 8         | 0     | 8           |            |            |

Dari hasil perhitungan pada tabel 3, atribut yang menjadi *node* cabang dari bentuk rumah-semi permanen dan lantai-semen adalah pekerjaan. Dimana atribut dari pekerjaan terdiri dari serabutan, bhl, dan swasta. Dimana swasta sudah memperoleh keputusan tidak layak. Sedangkan untuk sub atribut bhl belum memperoleh keputusan maka akan dilakukan perhitungan kembali, maka dapat digambarkan pohon keputusan dari tabel 3 sebagai berikut :



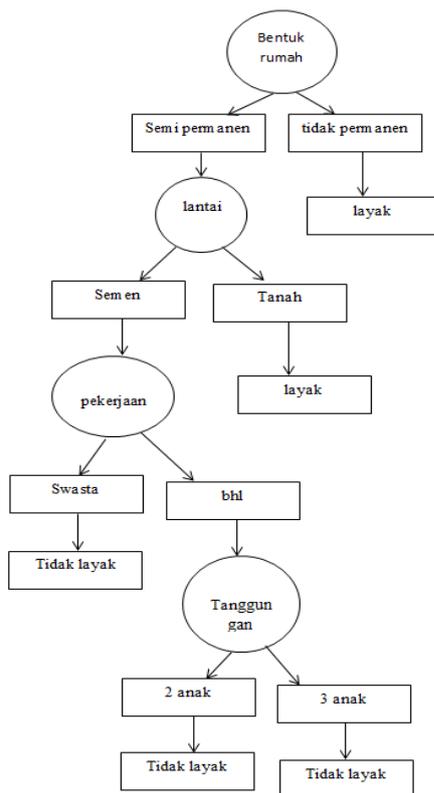
**Gambar 4.** Pohon keputusan 3

Selanjutnya melakukan perhitungan pada *node* cabang dari atribut pekerjaan-bhl dapat ditunjukkan pada Tabel 4.

**Tabel 4.** Perhitungan *Node* 1.1.1.1

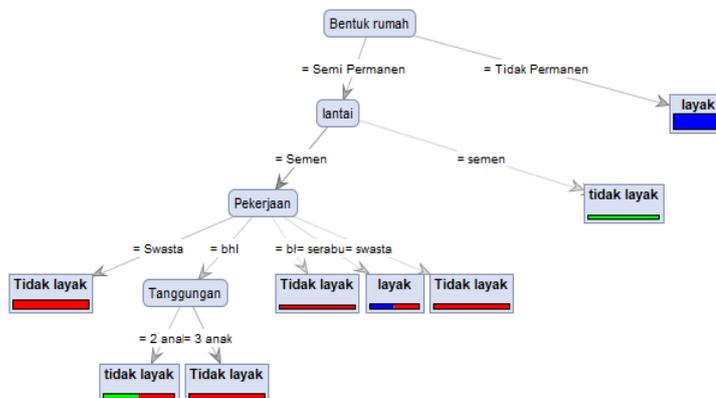
| Node 1.1.1.1                 | Jml kasus | Layak | Tidak Layak | Entropy    | Gain       |
|------------------------------|-----------|-------|-------------|------------|------------|
| bentuk rumah - semi permanen |           |       |             |            |            |
| lantai - semen               |           |       |             |            |            |
| pekerjaan - bhl              | 6         | 2     | 4           | 0.91829583 |            |
| Status                       |           |       |             |            | 0.89444399 |
| Menikah                      | 3         | 0     | 3           |            |            |
| Janda                        | 1         | 2     | 0           |            |            |
| Tanggungjan                  |           |       |             |            | 0.93406806 |
| 2 anak                       | 2         | 0     | 2           |            |            |
| 3 anak                       | 1         | 0     | 1           |            |            |
| 4 anak                       | 2         | 1     | 1           |            |            |
| 5 anak                       | 1         | 1     | 0           |            |            |

Dari hasil perhitungan pada tabel 4, atribut yang menjadi *node* cabang dari pekerjaan-bhl adalah Tanggungan. Dimana atribut tanggungan yang terdiri dari 2 anak, 3 anak, 4 anak, dan 5 anak. Dimana 2 anak memperoleh keputusan tidak layak, dan 3 anak memperoleh keputusan tidak layak. Maka proses perhitungan telah selesai dilakukan kemudian pohon keputusan perhitungan C4.5 dapat dilihat pada gambar 5.



Gambar 5. Pohon keputusan 4

Berikut adalah hasil pengolahan data dengan model pohon keputusan sesuai dengan *software Rapidminer*, dapat dilihat pada Gambar 6. sebagai berikut.



Gambar 6. Decision Tree Pada Rapidminer

Gambar 6 merupakan pohon keputusan yang dihasilkan pada *Rapidminer* dengan aturan atau *rule* yang dapat dilihat pada text view pada Gambar 7 berikut.

```

Tree
Bentuk rumah = Semi Permanen
| lantai = Semen
| | Pekerjaan = Swasta: Tidak layak {layak=0, tidak layak=0, Tidak layak=9}
| | Pekerjaan = bhl
| | | Tanggungan = 2 anak: tidak layak {layak=0, tidak layak=1, Tidak layak=1}
| | | Tanggungan = 3 anak: Tidak layak {layak=0, tidak layak=0, Tidak layak=3}
| | Pekerjaan = bhl : Tidak layak {layak=0, tidak layak=0, Tidak layak=1}
| | Pekerjaan = serabutan: layak {layak=1, tidak layak=0, Tidak layak=1}
| | Pekerjaan = swasta: Tidak layak {layak=0, tidak layak=0, Tidak layak=2}
| lantai = Tanah: layak {layak=1, tidak layak=0, Tidak layak=0}
| lantai = semen: tidak layak {layak=0, tidak layak=1, Tidak layak=0}
Bentuk rumah = Tidak Permanen: layak {layak=19, tidak layak=0, Tidak layak=0}
    
```

Gambar 7. Rule Decision Tree Pada Rapidminer

Gambar 7 menunjukkan hasil deskripsi secara lengkap dari pohon keputusan (*decision tree*) yang telah terbentuk dengan menggunakan algoritma C4.5 dari hasil deskripsi juga menunjukkan bahwa penggunaan data mining metode C4.5 baik digunakan dalam proses menggali data (*data mining process*) untuk menarik beberapa kesimpulan yang divisualisasikan dengan pohon keputusan (*decision tree*). Berikut adalah rules yang dihasilkan dari pohon keputusan tersebut :

- Jika bentuk rumah = Semi permanen dan Lantai = Semen dan pekerjaan = Swasta, maka hasilnya Tidak layak { layak = 0, tidak layak = 9}
- Jika lantai = semen dan pekerjaan = bhl dan tanggungan = 2 anak, maka hasilnya Tidak layak { layak = 0, tidak layak = 1}
- Jika lantai = semen dan pekerjaan = bhl dan tanggungan = 3 anak, maka hasilnya Tidak layak { layak = 0, tidak layak = 3}
- Jika lantai = semen dan pekerjaan = bhl, maka hasilnya Tidak layak { layak = 0, tidak layak = 1}
- Jika lantai = semen dan pekerjaan = serabutan, maka hasilnya layak { layak = 1, tidak layak = 0, tidak layak = 1}
- Jika lantai = semen dan pekerjaan = swasta, maka hasilnya tidak layak { layak = 0, tidak layak = 2}
- Jika bentuk rumah = semi permanen dan lantai = tanah, maka hasilnya layak { layak = 1, tidak layak = 0}
- Jika bentuk rumah = semi permanen lantai = semen, maka hasilnya tidak layak { layak = 0, tidak layak = 1}
- Jika bentuk rumah = tidak permanen, maka hasilnya layak { layak = 1, tidak layak = 0}

| accuracy: 83.33%  |            |                  |                  |                 |
|-------------------|------------|------------------|------------------|-----------------|
|                   | true layak | true tidak layak | true Tidak layak | class precision |
| pred. layak       | 8          | 1                | 1                | 80.00%          |
| pred. tidak layak | 0          | 0                | 0                | 0.00%           |
| pred. Tidak layak | 0          | 0                | 2                | 100.00%         |
| class recall      | 100.00%    | 0.00%            | 66.67%           |                 |

Gambar 8. Nilai Akurasi Algoritma C4.5

| PerformanceVector  |       |             |             |
|--------------------|-------|-------------|-------------|
| PerformanceVector: |       |             |             |
| accuracy: 83.33%   |       |             |             |
| ConfusionMatrix:   |       |             |             |
| True:              | layak | tidak layak | Tidak layak |
| layak:             | 8     | 1           | 1           |
| tidak layak:       | 0     | 0           | 0           |
| Tidak layak:       | 0     | 0           | 2           |
| kappa: 0.586       |       |             |             |
| ConfusionMatrix:   |       |             |             |
| True:              | layak | tidak layak | Tidak layak |
| layak:             | 8     | 1           | 1           |
| tidak layak:       | 0     | 0           | 0           |
| Tidak layak:       | 0     | 0           | 2           |

Gambar 9. Nilai Performance Algoritma C4.5

Berdasarkan pengolahan data menggunakan *software Rapidminer* didapat nilai akurasi sistem sebesar 83,33%. Dimana model yang dibentuk telah diuji tingkat akurasinya dengan menggunakan *split validation* pada aplikasi *Rapidminer 5.3*.

#### 4. KESIMPULAN

Berdasarkan pembahasan sebelumnya dapat disimpulkan bahwa tata mining dengan algoritma C4.5 dapat diterapkan untuk menentukan Kelayakan Rehabilitas Rumah Warga Tidak Layak Huni di Pematang Dolok Kahean, yang memperoleh hasil perhitungan nilai entropy dan gain pada setiap atribut. Hasil penerapan Algoritma C4.5 telah didapatkan Pohon keputusan dan rule yang menunjukkan bahwa atribut bentuk rumah sebagai atribut yang paling berpengaruh dalam keputusan Kelayakan rehabilitas Rumah Warga, kemudian diikuti dengan atribut lantai, pekerjaan dan tanggungan. Hasil penerapan Algoritma C4.5 dapat diuji dengan *Software Rapidminer* dan diperoleh yang sama dengan perhitungan manual dengan pemodelan Algoritma C4.5 menghasilkan Nilai Akurasi sebesar 83,33%..

#### REFERENCES

- [1] W. Aprianti, K. A. Hafizd, and M. R. Rizani, "Implementasi Association Rules dengan Algoritma Apriori pada Dataset Kemiskinan," *Limits: Journal of Mathematics and Its Applications*, vol. 14, no. 2, p. 57, 2017, doi: 10.12962/limits.v14i2.2933.

- [2] E. Septiya Anggrairi, A. Subagyo, D. Oktavina Radianto, and P. Perkapalan Negeri Surabaya, "Analisis Pengaruh Fasilitas Pendidikan Terhadap Tingkat Pengangguran dan Kemiskinan di Wilayah Indonesia Tahun 2018," *Riset dan Konseptual*, vol. 3, no. 2, pp. 109–115, 2019, [Online]. Available: [www.journal.unublitar.ac.id/jp](http://www.journal.unublitar.ac.id/jp).
- [3] M. Praveena and N. Bhavana, "Prediction of chronic kidney disease using C4.5 algorithm," *International Journal of Recent Technology and Engineering*, vol. 7, no. 6, pp. 721–723, 2019.
- [4] A. A. Elacio, L. L. Lacatan, A. A. Vinluan, and F. G. Balazon, "Machine learning integration of Herzberg's theory using c4.5 algorithm," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 1.1 Special Issue, pp. 57–63, 2020, doi: 10.30534/ijatcse/2020/1191.12020.
- [5] F. Syukmana *et al.*, "Predicting Relegation Clubs in Italian Serie A with Method based C4.5 Decision Tree Algorithm," *Journal of Physics: Conference Series*, vol. 1471, no. 1, 2020, doi: 10.1088/1742-6596/1471/1/012016.
- [6] N. Azwanti, "Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada Pt. Capella Dinamik Nusantara Cabang Muka Kuning," *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, vol. 13, no. 1, p. 33, 2018, doi: 10.30872/jim.v13i1.629.
- [7] K. F. Irnanda and A. P. Windarto, "Penerapan Klasifikasi C4.5 Dalam Meningkatkan Kecakapan Berbahasa Inggris dalam Masyarakat," pp. 304–308, 2020.
- [8] A. Waluyo, H. Jatnika, M. R. S. Permatasari, T. Tuslaela, I. Purnamasari, and A. P. Windarto, "Data Mining Optimization uses C4.5 Classification and Particle Swarm Optimization (PSO) in the location selection of Student Boardinghouses," *IOP Conference Series: Materials Science and Engineering*, vol. 874, no. 1, pp. 1–9, 2020, doi: 10.1088/1757-899X/874/1/012024.
- [9] W. Katrina, H. J. Damanik, F. Parhusip, D. Hartama, A. P. Windarto, and A. Wanto, "C4.5 Classification Rules Model for Determining Students Level of Understanding of the Subject," *Journal of Physics: Conference Series*, vol. 1255, no. 1, 2019, doi: 10.1088/1742-6596/1255/1/012005.
- [10] Y. S. Luvia, A. P. Windarto, S. Solikhun, and D. Hartama, "Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Keberhasilan Mahasiswa Di Amik Tunas Bangsa," *Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika)*, vol. 1, no. 1, p. 75, 2017, doi: 10.30645/jurasik.v1i1.12.
- [11] A. P. Windarto and T. Herawan, "Decision Support System on Determination of Contraception Tools as an Effort to Suppress the Number of Growth Ratios in Indonesia," *Lecture Notes in Electrical Engineering*, vol. 730, pp. 771–778, 2022, doi: 10.1007/978-981-33-4597-3\_69.
- [12] A. C. Rumahorbo and K. A. Sekarwati, "Penerapan Data Mining dengan Menggunakan C4.5 pada Klasifikasi Fasilitas Kesehatan Provinsi di Indonesia," *Jurnal Ilmiah KOMPUTASI*, vol. 19, no. 1, pp. 27–38, 2020.
- [13] K. Chokkanathan and S. Koteeswaran, "Privacy protection and perfect classification nature of C4.5 algorithm," *International Journal of Engineering and Technology(UAE)*, vol. 7, no. 2, pp. 235–238, 2018, doi: 10.14419/ijet.v7i2.24.12055.
- [14] K. Zaman, "Penerapan Data Mining Menggunakan Algoritma C4.5 Untuk Menentukan Kelayakan Penerima Bantuan Rehabilitas Sosial Rumah Tidak Layak Huni ( Studi Kasus Di Pemerintahan Kabupaten Solok Selatan )," *UPI YPTK Jurnal KomTekInfo*, vol. 3, no. 2, pp. 12–24, 2016.
- [15] R. Novita, S. Zakir, A. Nur Khomarudin, E. Maiyana, and H. Hasyim, "Use of the C4.5 Algorithm in Determining Scholarship Recipients," *Journal of Physics: Conference Series*, vol. 1779, no. 1, 2021, doi: 10.1088/1742-6596/1779/1/012009.
- [16] E. Indra, K. Ho, Arlinanda, R. Hakim, D. Sitanggang, and O. Sihombing, "Application of C4.5 Algorithm for Cattle Disease Classification," *Journal of Physics: Conference Series*, vol. 1230, no. 1, 2019, doi: 10.1088/1742-6596/1230/1/012070.
- [17] N. Azwanti, "Algoritma C4.5 Untuk Memprediksi Mahasiswa Yang Mengulang Mata Kuliah (Studi Kasus Di Amik Labuhan Batu)," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 9, no. 1, pp. 11–22, 2018, doi: 10.24176/simet.v9i1.1627.
- [18] R. R. C. Silva, W. M. Caminhas, P. C. De Lima E Silva, and F. G. Guimaraes, "A C4.5 Fuzzy Decision Tree Method for Multivariate Time Series Forecasting," *IEEE International Conference on Fuzzy Systems*, vol. 2021-July, 2021, doi: 10.1109/FUZZ45933.2021.9494439.
- [19] L. N. Rani, "Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit," *INOVTEK Polbeng - Seri Informatika*, vol. 1, no. 2, p. 126, 2016, doi: 10.35314/isi.v1i2.131.