

# Klasifikasi Calon Nasabah Baru Menggunakan C.45 Sebagai Dasar Pemberian Pertanggungjawaban Asuransi di PT Asuransi Central Asia Pematangsiantar

Retno Ayu Syahfitri, Agus Perdana Windarto, Harly Okprana\*

Program Studi Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Email: Retnoayu.syahfitri@gmail.com

**Abstrak**—Permasalahan yang sering timbul dalam permasalahan asuransi adalah banyaknya nasabah yang menunggak dalam membayar premi, oleh karena itu diperlukan sebuah sistem yang dapat mengklasifikasikan calon nasabah mana yang masuk ke dalam kelompok layak dan nasabah mana yang masuk ke dalam kelompok tidak layak dalam pengajuan sebagai nasabah asuransi, sehingga pihak asuransi bisa mengatasi sejak dini permasalahan tersebut. Pada penelitian ini dibangun sistem informasi pengklasifikasian penerimaan calon nasabah asuransi menggunakan *Classification Tree C4.5*. *Classification Tree C4.5* digunakan untuk menghasilkan *rule* yang diperlukan untuk proses klasifikasi kelancaran nasabah dalam pembayaran premi dan terbentuk dari hasil konversi pembangunan pohon klasifikasi (*classification tree*). Algoritma C4.5 dianggap sebagai algoritma yang sangat membantu dalam melakukan klasifikasi data karena karakteristik data yang diklasifikasikan dapat diperoleh dengan jelas, baik (*decision tree*) maupun dalam bentuk aturan atau *rule If – then* sehingga memudahkan pengguna dalam melakukan penggalian informasi terhadap data yang bersangkutan.

**Kata Kunci:** Data Mining; Klasifikasi Calon Nasabah Baru; C4.5

**Abstract**—The problem that often arises in insurance problems is the number of customers who are in arrears in paying premiums, therefore a system is needed that can classify which prospective customers are in the eligible group and which customers are in the unfit group in submitting as insurance customers. so that the insurance company can solve the problem early. In this study, an information system for the classification of acceptance of prospective insurance customers was built using the Classification Tree C4.5. Classification tree C4.5 is used to generate the rules needed for the smooth classification process of customers in paying premiums and is formed from the conversion result of the construction of a classification tree (*classification tree*). The C4.5 algorithm is considered an algorithm that is very helpful in classifying data because the characteristics of the classified data can be obtained clearly, both (*decision tree*) and in the form of rules or *If - then rules*, making it easier for users to extract information on the data in question.

**Keywords:** Data Mining; Classification of New Prospective Customers; C4.5

## 1. PENDAHULUAN

Manusia tidak dapat meramalkan apa yang akan terjadi diwaktu yang akan datang secara sempurna meskipun dengan menggunakan beberapa alat analisis. Hal itu pula yang terjadi pada perusahaan maupun individu. Risiko di masa datang dapat terjadi terhadap kehidupan seseorang misalnya saja: kematian, sakit, kecelakaan atau risiko dipecat dari pekerjaan sehingga manusia membutuhkan yang namanya asuransi. Dalam dunia bisnis, risiko yang dihadapi dapat berupa kerugian akibat kebakaran, kerusakan atau kehilangan. Oleh karena itu, setiap risiko yang dihadapi harus ditanggulangi sehingga tidak menimbulkan kerugian yang lebih besar lagi. Dapat diketahui juga asuransi sudah banyak peminatnya dikalangan masyarakat, akan tetapi pada saat ini untuk memberikan asuransi kepada nasabah masih berdasarkan proses yang tidak objektif. Proses yang tidak objektif disini maksudnya adalah proses yang data nya belum tentu keabsahannya dikarenakan salah satu alasan nya adalah *surveyor* yang bertugas untuk survei data nasabah masih banyak kesalahan di kriteria calon nasabah baru asuransi yang akan diberi pertanggungjawaban asuransi yang menyebabkan sulit menentukan kelayakan pemberian asuransi yang sering dialami oleh pihak asuransi. Permasalahan yang sering timbul dalam permasalahan asuransi adalah banyaknya nasabah yang menunggak dalam membayar premi, oleh karena itu diperlukan sebuah sistem yang dapat mengklasifikasikan calon nasabah mana yang masuk ke dalam kelompok layak dan nasabah mana yang masuk ke dalam kelompok tidak layak dalam mengasuransikan jiwa, harta benda atau agunan nya, sehingga pihak asuransi bisa mengatasi sejak dini permasalahan tersebut.

Dalam melaksanakan tugasnya PT Asuransi Central Asia (ACA) yang beralamat di Jl. Merdeka No. 246, Pahlawan Kec. Siantar Timur, Kota Pematang Siantar, Sumatera Utara ini memberikan penawaran kembali terhadap nasabah yang memiliki berbagai harta benda, asuransi jiwa atau agunan yang dapat di asuransi kan untuk mengantisipasi kerugian finansial akibat kejadian tak terduga yang menimpa properti atau jaminan kesehatan calon nasabah, dengan adanya asuransi ini nasabah akan merasa aman dan terlindungi dan tentunya dengan klasifikasi kriteria yang sesuai dengan persyaratan yang ada.

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan (*Decision Tree*). Pohon keputusan merupakan metode klasifikasi dan prediksi yang terkenal. Pohon keputusan berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain : ID3, CART, dan C4.5. Algoritma C4.5 merupakan pengembangan dari algoritma ID3, Proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan *rule* (Hidayatus Sholihah, 2018)

Adapun penelitian tentang data mining (Rani, 2015), yang membahas tentang klasifikasi nasabah menggunakan algoritma C4.5 sebagai dasar pemberian pertanggung jawaban asuransi. Klasifikasi calon nasabah asuransi baru ini bertujuan untuk memudahkan pihak asuransi dalam membuat keputusan dalam hal pemberian pertanggung jawaban asuransi. Dengan adanya klasifikasi calon nasabah asuransi baru, jika terjadi masalah dengan kasus yang sama pihak asuransi tinggal melihat aturan-aturan (*rule*) yang telah terbentuk dari pohon keputusan yang dihasilkan. Dengan metode keputusan *decision tree* menggunakan algoritma C4.5 diharapkan proses penggalian informasi lebih cepat dan optimal dengan kapasitas data yang lebih besar, sehingga kesalahan yang ditimbulkan dalam pengambilan keputusan lebih diminimalkan. Dengan begitu dapat disimpulkan bahwa metode pohon keputusan (*decision tree*) yang diproses dengan *software Rapidminer* dapat mengidentifikasi kelayakan kredit dengan baik. Pemilihan variabel (atribut kondisi dan atribut keputusan) yang akan digunakan dalam menentukan sebuah klasifikasi juga sangat mempengaruhi *rule* atau *knowledge* yang dihasilkan sehingga sistem yang dibangun dapat membantu dalam mengklasifikasikan kriteria anggota nasabah yang pengajuan perpanjangan kreditnya berpotensi dan yang tidak berpotensi dan Algoritma C4.5 dianggap sebagai algoritma yang sangat membantu dalam melakukan klasifikasi data karena karakteristik data yang diklasifikasikan dapat diperoleh dengan jelas, baik dalam bentuk struktur pohon keputusan (*decision tree*) maupun dalam bentuk aturan atau *rule If – Then* sehingga memudahkan pengguna dalam melakukan penggalian informasi terhadap data yang bersangkutan.

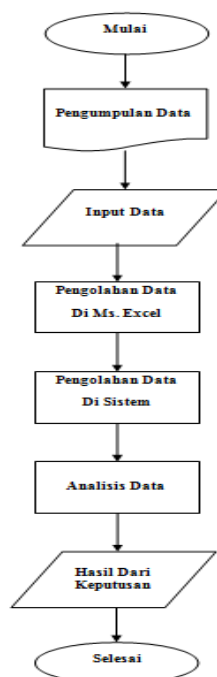
Pada penelitian ini dibangun sistem informasi pengklasifikasian penerimaan calon nasabah asuransi menggunakan *Classification Tree C4.5*. *Classification Tree C4.5* digunakan untuk menghasilkan *rule* yang diperlukan untuk proses klasifikasi nasabah dalam penentuan layak tidak layak nya nasabah dalam penerimaan manfaat asuransi dan terbentuk dari hasil konversi pembangunan pohon klasifikasi (*classification tree*), sedangkan *cost sensitive learning* digunakan untuk mengatasi masalah *missclassification cost* di setiap split sehingga berpengaruh pada kinerja pengklasifikasian, kemudian menentukan termasuk pada kelompok yang tepat..

Berdasarkan latar belakang tersebut, penulis menyimpulkan diperlukan sebuah penelitian untuk melakukan klasifikasi calon nasabah baru sebagai dasar pemberian pertanggung jawaban asuransi. Dalam hal ini, penulis menggunakan penelitian algoritma C45 dalam melakukan klasifikasi pada PT. Asuransi Central Asia seperti yang pernah dilakukan peneliti-peneliti sebelumnya.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Penelitian dilakukan berdasarkan Diagram Alur Aktifitas Kerja Penelitian yang menerangkan bagaimana suatu sistem yang berjalan dari awal sampai akhir sistem digunakan. Berikut ini adalah diagram aktifitas peneliti dan diagram program :



Gambar 1. Diagram Alur Kerja Penelitian

Gambar 1 merupakan alur aktivitas yang dilakukan penelitian pada sistem yang digunakan. gambar 1 menjelaskan penulis atau pengguna dengan mengumpulkan data terlebih dahulu setelah itu penulis menginput data ke dalam *Microsoft Excel*. Setelah itu penulis mengolah data melalui *microsoft excel* setelah itu penulis mengolah data

melalui sistem yaitu *rapidminer*. Setelah melalui proses pengolahan data di *microsoft excel* dan sistem *rapidminer* penulis melakukan analisis data dan dari hasil analisis data penulis memperoleh hasil dari keputusan yang telah diproses.

## 2.2 Metode Pengumpulan Data

Pengumpulan data yang digunakan untuk penelitian terdapat beberapa prosedur yang digunakan, yakni yang pertama adalah studi pustaka yaitu memanfaatkan perpustakaan, buku, prosiding atau jurnal sebagai referensi dalam menentukan faktor, parameter, dan label yang digunakan untuk penelitian. Kemudian yang kedua adalah data sekunder, Penulis juga menggunakan data sekunder dari register formulir calon nasabah asuransi baru di PT. Asuransi Central Asia Pematangsiantar. Dalam penelitian, data yang digunakan akan diolah dari hasil pengumpulan data nasabah yang diberikan kepada PT. Central Asia Pematangsiantar dengan sampel sebanyak 100 data dapat dilihat pada tabel dibawah ini:

**Tabel 1.** Data Calon Nasabah Asuransi

No.	Nama	Usia	Penghasilan	Jumlah Tanggungan (Orang)	Keputusan
1	Situlus Sigalingging	30	3500000	4	Layak
2	Marnaek Marojahan Sinaga	33	4000000	6	Tidak Layak
3	Pintor Sitorus	45	4450000	4	Layak
4	Erni R Marpaung	27	3000000	1	Layak
5	Pancasila Sibarani	29	4800000	6	Tidak Layak
6	Manro Nababan	34	2950000	4	Tidak Layak
...	.....	.....	.....	...	.....
100	Ratna Masni Sagala	39	4500000	0	Layak

Berdasarkan data yang dikumpulkan maka dilakukan proses seleksi dan transformasi data sehingga ada kriteria yang digunakan sebanyak 3 yakni : usia, penghasilan dan jumlah tanggungan. Adapun sub kriteria dari setiap kriteria adalah sebagai berikut:

**Tabel 2.** Kriteria Usia

Umur	24 - 40	Baik
	40 - 50	Cukup
	50 -70	Kurang

**Tabel 3.** Kriteria Penghasilan

Penghasilan	Rp. 2.500.000 – Rp. 3.000.000	Cukup
	Rp. 3.000.000 – Rp. 6.000.000	Baik
	> Rp. 6.000.000	Sangat Baik

**Tabel 4.** Kriteria Jumlah Tanggungan

Jumlah Tanggungan	0 - 1	Sangat Baik
	2 - 4	Baik
	> 6	Cukup

## 2.3 Klasifikasi

Menurut (Defiyanti 2015), Klasifikasi adalah pemrosesan untuk menemukan sebuah model atau fungsi yang menjelaskan dan mencirikan konsep atau kelas data, untuk kepentingan tertentu”. Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui (Hidayatus Sholihah 2018).

## 2.4 Algoritma C4.5

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 (Kamagi and Hansun 2014).

Pada tahap belajar dari data, algoritma C4.5 mengkonstruksi pohon keputusan dari data pelatihan, yang berupa kasus-kasus atau *record-record (tupel)* dalam *database*. Tiga prinsip kerja algoritma C4.5 pada tahap belajar dari data adalah :

- a. Pembuatan Pohon Keputusan
- b. Pemangkasan pohon keputusan dan evaluasi (opsional).

c. Pembuatan aturan-aturan dari pohon keputusan (opsional).

Rumus pada algoritma C4.5 adalah :

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

Keterangan :

S : Himpunan Kasus

A : Atribut

N : Jumlah partisi atribut A

|Si| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

Sedangkan perhitungan nilai *Entropy* dapat dilihat pada persamaan berikut ini ;

$$Entropy(A) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (2)$$

Keterangan :

S : Himpunan Kasus

A : Fitur

n : Jumlah partisi S

pi : Proporsi dari Si terhadap S

### 3. HASIL DAN PEMBAHASAN

Proses menyelesaikan masalah klasifikasi untuk penentuan kelayakan calon nasabah asuransi baru PT Asuransi Centra Asia Pematangsiantar untuk tahun 2020 ada beberapa tahapan yang harus dilakukan antara lain sebagai berikut:

- a. Langkah pertama menyiapkan *Data Training*.
- b. Menentukan kriteria dari data yang diperoleh
- c. Melakukan perhitungan nilai *entropy* dan *gain*
- d. Melakukan klasifikasi dengan menggunakan Algoritma C4.5

Dalam pengolahan data ini menggunakan metode algoritma C4.5 (perhitungan-perhitungan) atau pemodelan/rangkaian yang digunakan, atau penerapan algoritma terhadap klasifikasi calon nasabah baru. Adapun kriteria yang digunakan untuk menentukan kelayakan calon nasabah baru adalah Usia (X1), penghasilan (X2), dan jumlah tanggungan (X3). Data yang digunakan adalah data calon nasabah baru tahun 2020 dengan 100 *record* data.

**Tabel 5.** Data Olahan Calon Nasabah Baru

No.	Nama	Usia	Penghasilan	Jumlah Tanggungan (Orang)
1	Situlus Sigalingging A1	30	3500000	4
2	Marnaek Marojahan Sinaga A2	33	4000000	6
3	Pintor Sitorus A3	45	4450000	4
4	Erni R Marpaung A4	27	3000000	1
5	Pancasila Sibarani A15	29	4800000	6
6	Manro Nababan A16	34	2950000	4
...	.....	.....	.....	...
100	Ratna Masni Sagala A100	39	4500000	0

Berdasarkan tabel 5 dibagi data menjadi 2 bagian yaitu data pelatihan (*training*) yang ditunjukkan pada tabel 6 dan data pengujian (*testing*) yang ditunjukkan pada tabel 7 berikut ini:

**Tabel 6.** Data Pelatihan (*Training*)

No.	Nama	Usia	Penghasilan	Jumlah Tanggungan (Orang)
1	A1	Baik	Baik	Baik
2	A2	Baik	Baik	Cukup
3	A3	Cukup Baik	Baik	Baik
4	A4	Baik	Baik	Sangat Baik
5	A5	Baik	Baik	Cukup
6	A6	Baik	Cukup	Baik
...	.....	.....	.....	...
50	A50	Baik	Baik	Sangat Baik

**Tabel 7.** Data Pengujian (*Testing*)

No.	Nama	Usia	Penghasilan	Jumlah Tanggungan (Orang)
51	A51	Cukup Baik	Baik	Sangat Baik
52	A52	Baik	Baik	Baik
53	A53	Cukup Baik	Baik	Baik
54	A54	Baik	Cukup	Baik
55	A55	Baik	Baik	Sangat Baik
56	A56	Kurang	Baik	Baik
...	.....	....	.....	...
100	A100	Baik	Baik	Sangat Baik

Setelah menentukan data *training* dan data *testing*, langkah selanjutnya menentukan gain dengan cara perhitungan gain dan *entropy*. Untuk perhitungan algoritma C4.5 dapat diuraikan sebagai berikut:

**Langkah 1 :**

Menghitung jumlah kasus, jumlah kasus untuk kategori layak dan jumlah kasus untuk kategori tidak layak dan *Entropy* dari semua kasus berdasarkan atribut yang digunakan dengan rumus mencari *entropy* pada persamaan 2 adalah :

$$\begin{aligned}
 Entropy\ S &= \left(-\left(\frac{17}{100}\right) * \text{Log}_2\left(\frac{17}{100}\right)\right) + \left(-\left(\frac{83}{100}\right) * \text{Log}_2\left(\frac{83}{100}\right)\right) \\
 &= 0,43458686924 + 0,22311790949 \\
 &= 0,6577047787
 \end{aligned}$$

Berikut hasil lengkap perhitungan *Entropy* keseluruhan seperti yang ditunjukkan pada tabel 8 berikut :

**Tabel 8.** Hasil Perhitungan *Entropy* Total Keseluruhan

Total Kasus	Sum (Layak)	Sum (Tidak Layak)	Entropy Total
100	83	17	0,6577047787

**Langkah 2 :**

Menghitung *entropy* tiap atribut sebagai berikut:

- a. Perhitungan Entropy Usia

$$\begin{aligned}
 Entropy\ (Usia, Baik) &= \left(\left(-\frac{63}{67}\right) * \text{Log}_2\left(\frac{63}{67}\right)\right) + \left(-\frac{4}{67}\right) * \text{Log}_2\left(\frac{4}{67}\right) \\
 &= 0,08350722116 + 0,24275159 \\
 &= 0,3262588
 \end{aligned}$$

$$\begin{aligned}
 Entropy\ (Usia, Cukup) &= \left(\left(-\frac{18}{24}\right) * \text{Log}_2\left(\frac{18}{24}\right)\right) + \left(-\frac{6}{24}\right) * \text{Log}_2\left(\frac{6}{24}\right) \\
 &= 0,31127812445 + 0,5 \\
 &= 0,8112781
 \end{aligned}$$

$$\begin{aligned}
 Entropy\ (Usia, Kurang) &= \left(\left(-\frac{2}{9}\right) * \text{Log}_2\left(\frac{2}{9}\right)\right) + \left(-\frac{7}{9}\right) * \text{Log}_2\left(\frac{7}{9}\right) \\
 &= 0,48220555 + 0,28199895063 \\
 &= 0,7642045
 \end{aligned}$$

- b. Perhitungan *Entropy* Penghasilan

$$\begin{aligned}
 Entropy\ (Penghasilan, Cukup) &= \left(\left(-\frac{4}{8}\right) * \text{Log}_2\left(\frac{4}{8}\right)\right) + \left(-\frac{4}{8}\right) * \text{Log}_2\left(\frac{4}{8}\right) \\
 &= 0,5 + 0,5 \\
 &= 1
 \end{aligned}$$

$$\begin{aligned}
 Entropy\ (Penghasilan, Baik) &= \left(\left(-\frac{68}{81}\right) * \text{Log}_2\left(\frac{68}{81}\right)\right) + \left(-\frac{13}{81}\right) * \text{Log}_2\left(\frac{13}{81}\right) \\
 &= 0,21188058013 + 0,42360905804 \\
 &= 0,6354896
 \end{aligned}$$

$$\begin{aligned}
 Entropy\ (Penghasilan, Sangat Baik) &= \left(\left(-\frac{11}{11}\right) * \text{Log}_2\left(\frac{11}{11}\right)\right) + \left(-\frac{0}{11}\right) * \text{Log}_2\left(\frac{0}{11}\right) \\
 &= 0
 \end{aligned}$$

c. Perhitungan *Entropy* Jumlah Tanggungan

$$\begin{aligned} Entropy (\text{JmlhTanggungan,SB}) &= \left(-\frac{37}{39}\right) * \text{Log}_2\left(\frac{37}{39}\right) + \left(-\frac{2}{39}\right) * \text{Log}_2\left(\frac{2}{39}\right) \\ &= 0,27094678 + 0,21976421635 \\ &= 0,2918183 \end{aligned}$$

$$\begin{aligned} Entropy (\text{JmlhTanggungan,Baik}) &= \left(-\frac{43}{55}\right) * \text{Log}_2\left(\frac{43}{55}\right) + \left(-\frac{12}{55}\right) * \text{Log}_2\left(\frac{12}{55}\right) \\ &= 0,18400147 + 0,47921393733 \\ &= 0,7568336 \end{aligned}$$

$$\begin{aligned} Entropy (\text{JmlhTanggungan,Cukup}) &= \left(-\frac{3}{6}\right) * \text{Log}_2\left(\frac{3}{6}\right) + \left(-\frac{3}{6}\right) * \text{Log}_2\left(\frac{3}{6}\right) \\ &= 0,5 + 0,5 \\ &= 1 \end{aligned}$$

**Langkah 3 :**

Perhitungan Mencari Gain dari kriteria usia, penghasilan dan jumlah tanggungan dengan menggunakan persamaan 1

a. Perhitungan Gain Usia

$$\begin{aligned} Gain (\text{Total ,Usia}) &= (0,6577048) - \left(\left(\frac{67}{100}\right) * (0,3262588) + \left(\frac{24}{100}\right) * (0,8112781) + \left(\frac{9}{100}\right) * (0,7642045)\right) \\ &= 0,27308 \end{aligned}$$

b. Perhitungan Gain Penghasilan

$$\begin{aligned} Gain (\text{Total ,Penghasilan}) &= (0,6577048) - \left(\left(\frac{8}{100}\right) * (1) + \left(\frac{81}{100}\right) * (0,6354896) + \left(\frac{11}{100}\right) * (0,4836373)\right) \\ &= -0,04704 \end{aligned}$$

c. Perhitungan Gain Jumlah Tanggungan

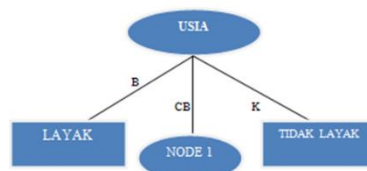
$$\begin{aligned} Gain (\text{Total ,JlhTanggungan}) &= (0,6577048) - \left(\left(\frac{39}{100}\right) * (0,2918183) + \left(\frac{55}{100}\right) * (0,7568336) + \left(\frac{6}{100}\right) * (1)\right) \\ &= 0,06764 \end{aligned}$$

Perhitungan nilai entropy dan gain untuk semua atribut dilakukan untuk mendapatkan nilai gain tertinggi yang akan dijadikan sebagai akar. Hasil perhitungannya dapat dilihat pada tabel 9 berikut ini :

**Tabel 9.** Perhitungan Node 1

Atribut		Jumlah	Layak	Tidak Layak	Entropy	Gain
Total		100	83	17	0.65770478	0.175626
Usia	Baik	67	63	4	0.32625881	
	Cukup Baik	24	18	6	0.81127812	
	Kurang	9	2	7	0.76420451	
Penghasilan						-0.04704
Penghasilan	Cukup	8	4	4	1	
	Baik	81	68	13	0.63548964	
	Sangat Baik	11	11	0	0	
Jmlh Tanggungan						0.067637
Jmlh Tanggungan	Sangat Baik	39	37	2	0.29181826	
	Baik	55	43	12	0.75683363	
	Cukup	6	3	3	1	

Dari tabel 9 diatas dapat diketahui bahwa atribut yang memiliki gain tertinggi adalah atribut usia yaitu 0,175626. Dengan demikian, atribut usia dapat menjadi node akar. Ada tiga nilai dari atribut usia yaitu Baik, Cukup Baik dan Kurang. Atribut Baik pada usia mengklasifikasikan kasus yang layak sedangkan atribut Kurang pada usia mengklasifikasikan kasus yaitu tidak layak. Dari hasil perhitungan tersebut dapat digambarkan pohon keputusan sementara tampak seperti gambar 2 yaitu:



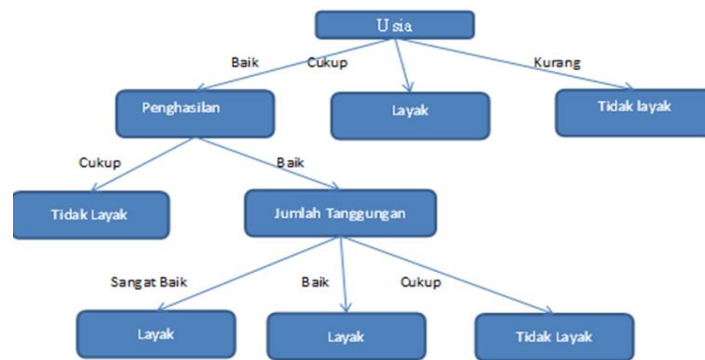
**Gambar 2.** Pohon Keputusan Hasil Perhitungan Node 1

Tabel 10. Perhitungan Node 1.1

Atribut	Jumlah	Layak	Tidak Layak	Entropy	Gain
	100	83	17	0.657704779	0.220659
Cukup	15	13	15	0.178924094	
Baik	27	26	9	0.580752033	
Sangat Baik	9	5	1	0.823323282	0.444965
Sangat Baik	13	10	13	0.291162787	
Baik	18	15	11	0.653385739	
Cukup	20	19	1	0.286396957	

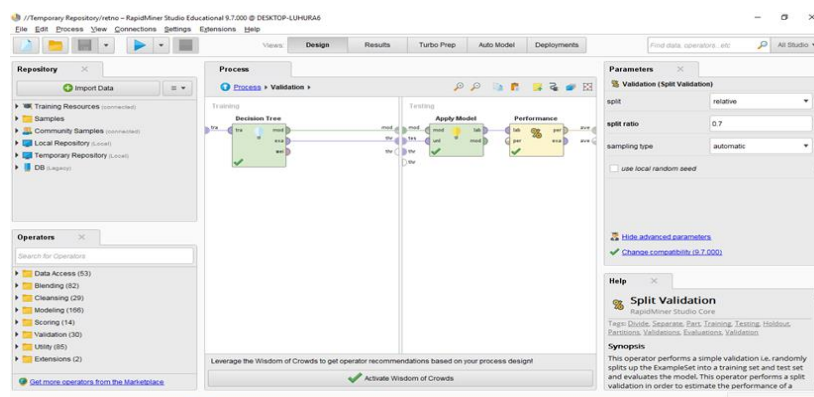
Dari tabel 10 diatas dapat diketahui bahwa atribut yang memiliki gain tertinggi adalah atribut jumlah tanggungan yaitu 0.44496. dengan demikian atribut jumlah tanggungan cabang keputusan terakhir dalam kasus ini. Ada tiga nilai dari atribut jumlah tanggungan yaitu Sangat Baik, Baik, dan Cukup. Jadi Sangat Baik pada jumlah tanggungan mengklasifikasikan kasus yaitu Layak sedangkan atribut Baik pada jumlah tanggungan mengklasifikasikan kasus yaitu Layak dan atribut Cukup mengklasifikasikan kasus yaitu Tidak layak.

Dari hasil perhitungan tersebut dapat digambarkan pohon keputusan sementara yang tampak seperti gambar 3 dibawah ini:



Gambar 3. Pohon Keputusan Perhitungan Node 1.1

Untuk melakukan pengujian pada penelitian ini dilakukan dengan menggunakan aplikasi *Rapidminer Studio 9.7*. *Software Rapidminer 9.7* menyediakan prosedur *data mining* dan *machine learning* termasuk: ETL (*extraction, transformation, loading*), *data preprocessing*, *visualisasi*, *modelling* dan evaluasi. Data calon nasabah asuransi baru yang telah diperoleh dan sudah dianalisis selanjutnya akan diolah dengan menggunakan aplikasi *Rapidminer Studio 9.7* untuk mengetahui klasifikasi kelayakan calon nasabah baru.



Gambar 4. Tampilan port decision tree, apply model dan performance

Setelah dilakukan perhitungan dan pengujian pada data dengan menggunakan Algoritma C4.5 dan telah dilakukan penghubungan port-port dari operator *decision tree*, operator *apply model* dan operator *performance* maka akan menampilkan hasil :



Gambar 5. Tampilan *Decision Tree*

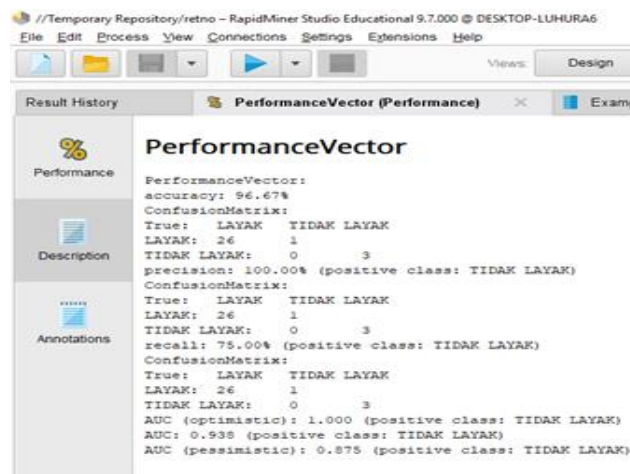
Dari pohon keputusan yang terbentuk dengan menggunakan algoritma C4.5 diperoleh *rule* berupa teks yang terbentuk dari proses penggalian data.

### Tree

```
Tanggung = Baik: LAYAK {LAYAK=40, TIDAK LAYAK=4}
Tanggung = Cukup
| Usia = Baik: LAYAK {LAYAK=4, TIDAK LAYAK=0}
| Usia = Cukup: TIDAK LAYAK {LAYAK=0, TIDAK LAYAK=6}
| Usia = Kurang: TIDAK LAYAK {LAYAK=0, TIDAK LAYAK=5}
Tanggung = Sangat Baik: LAYAK {LAYAK=41, TIDAK LAYAK=0}
```

Gambar 6. Deskripsi *Decision Tree*

Gambar 6 menunjukkan hasil deskripsi secara lengkap dari pohon keputusan yang telah terbentuk dengan menggunakan Algoritma C4.5. Dari hasil deskripsi juga menunjukkan bahwa Algoritma C4.5 baik digunakan dalam proses penggalian data untuk menarik beberapa kesimpulan yang divisualisasikan ke dalam pohon keputusan. Berikut merupakan *rule* yang diperoleh pohon keputusan yang terbentuk.



Gambar 7. Nilai *Performance Vektor Algoritma C4.5*

Dari hasil *performance vector* diatas dapat diketahui bahwa penerapan Algoritma C4.5 dalam mengklasifikasikan calon nasabah baru dengan menggunakan *software RapidMiner* memperoleh tingkat akurasi sebesar 96,67%. Dari gambar diatas dijelaskan bahwa klasifikasi layak memiliki nilai 26 dan klasifikasi tidak layak memiliki nilai 3 dengan nilai *precision* sebesar 91,38% dan nilai *recall* sebesar 75,00%.

## 4. KESIMPULAN

Dari uraian yang sudah dibahas sebelumnya dapat ditarik kesimpulan metode pohon keputusan (decision tree) yang diproses dengan software Rapidminer dapat mengidentifikasi kelayakan calon nasabah baru dengan baik. Pemilihan variabel (atribut kondisi dan atribut keputusan) yang akan digunakan dalam menentukan sebuah klasifikasi juga mempengaruhi rule atau knowledge yang dihasilkan. Sistem yang dibangun dapat membantu dalam mengklasifikasikan kriteria calon nasabah baru yang pengajuan asuransi nya layak atau tidak layak. Algoritma C4.5 dianggap sebagai algoritma yang sangat membantu dalam melakukan klasifikasi data karena karakteristik data yang diklasifikasikan dapat diperoleh dengan jelas, baik (decision tree) maupun dalam bentuk aturan atau rule If – then sehingga memudahkan pengguna dalam melakukan penggalian informasi terhadap data yang bersangkutan. Dari hasil *performance vector* diatas dapat diketahui bahwa penerapan Algoritma C4.5 dalam mengklasifikasikan calon nasabah baru dengan menggunakan *software RapidMiner* memperoleh tingkat akurasi sebesar 96,67%. Dari gambar diatas dijelaskan bahwa



klasifikasi layak memiliki nilai 26 dan klasifikasi tidak layak memiliki nilai 3 dengan nilai *precision* sebesar 91.38% dan nilai *recall* sebesar 75,00%.

## REFERENCES

- [1] Azwanti, Nurul, 'Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada Pt. Capella Dinamik Nusantara Cabang Muka Kuning', *Informatika Mulawarman: Jurnal Ilmiah Ilmu Komputer*, 13.1 (2018), 33 <<https://doi.org/10.30872/jim.v13i1.629>>
- [2] Fitrianda, Meilina Indah, 'Digital Digital Repository Repository Universitas Universitas Jember Jember Digital Digital Repository Repository Universitas Universitas Jember', 2013
- [3] Putra, Dede Wira Trise, 'Algoritma C4.5 Untuk Menentukan Tingkat Kelayakan Motor Bekas Yang Akan Dijual', *Jurnal TEKNOIF*, 4.1 (2016), 16–22
- [4] Rani, Larissa Navia, 'Klasifikasi Nasabah Menggunakan Algoritma', *Jurnal KomTekInfo Fakultas Ilmu Komputer*, 2.2 (2015), 33–38
- [5] Yuli, Mardi, 'Jurnal Edik Informatika Data Mining: Klasifikasi Menggunakan Algoritma C4 . 5 Data Mining Merupakan Bagian Dari Tahapan Proses Knowledge Discovery in Database ( KDD ) . Jurnal Edik Informatika', *Jurnal Edik Informatika*, 2.2 (2017), 213–19
- [6] Defiyanti, Sofi, 'Integrasi Metode Klasifikasi Dan Clustering Dalam Data Mining', January 2015, 2015, 39–44
- [7] Santoso, Teguh Budi and Dela Sekardiana. 2019. "Penerapan Algoritma C4.5 Untuk Penentuan Kelayakan Pemberian Kredit ( Studi Kasus : Koperia - Koperasi Warga Komplek Gandaria ) Implementation of C4 . 5 Algorithm To Determine The Feasibility of Loan." *Jurnal Algoritma, Logika Dan Komputasi* 11(1):130–37.
- [8] Firmansyah, Meidy and Rusito Aufany. 2016. "Implementasi Metode Decision Tree Dan Algoritma C4.5 Untuk Klasifikasi Data Nasabah Bank." *Infokam* XII(1):1–12.
- [9] Hidayatus Sholihah, Fiqih Satria. 2018. "Implementasi Data Mining Klasifikasi Nasabah Potensial Menggunakan Algoritma C4.5 (Studi Kasus Adira Dinamika Multi Finance)." 2:179.
- [10] Purnamasai, Dewi and Bambang Hendrawan. 2013. "Analisis Kelayakan Bisnis Usaha Roti Ceriwis Sebagai Oleh-Oleh Khas Kota Batam Dewi." *Jurnal Akuntansi, Ekonomi Dan Manajemen Bisnis* 3(1):83–87.
- [11] Maulana, Algifanri and Alfannisa Annurullah Fajrin. 2018. "Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma Fp-Growth Pada Data Transaksi Penjualan Spare Part Motor." *Klik - Kumpulan Jurnal Ilmu Komputer* 5(1):27.
- [12] Fitrianda, Meilina Indah. 2013. "Digital Digital Repository Repository Universitas Universitas Jember Jember Digital Digital Repository Repository Universitas Universitas Jember."
- [13] Sucipto, Adi. 2015. "Pada Koperasi Simpan Pinjam Dengan Menggunakan METODE Algoritma Klasifikasi C4.5." *Jurnal DISPROTEK* 6(1):75–87.
- [14] Fatmawati, Fatmawati. 2016. "Perbandingan Algoritma Klasifikasi Data Mining Model C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Diabetes." *None* 13(1):50–59.
- [15] Yuli, Mardi. 2017. "Jurnal Edik Informatika Data Mining: Klasifikasi Menggunakan Algoritma C4 . 5 Data Mining Merupakan Bagian Dari Tahapan Proses Knowledge Discovery in Database ( KDD ) . Jurnal Edik Informatika." *Jurnal Edik Informatika* 2(2):213–19.
- [16] Iswandy, Eka, Dosen Sekolah Tinggi Manajemen Informatika Komputer, and STMIK Jayanusa Padang. 2015. "Sistem Penunjang Keputusan Untuk Menentukan Dan Santunan Sosial Anak Nagari Dan Penyaluran Bagi Mahasiswa Dan Pelajar Kurang Mampu." *Jurnal TEKNOIF* 3(2).