

## Data Mining Algoritma C4.5 Pada Klasifikasi Kredit Koperasi Simpan Pinjam

Julkari Sinta Parapat, Anita Sindar Sinaga

Teknik Informatika, STMIK Pelita Nusantara  
Jalan Iskandar Muda No. 1 Medan, Indonesia 20154  
e-mail: julkakamilan@gmail.com, haito\_ita@yahoo.com

### Abstract

*Savings and loan cooperatives are developing from the customer's loan interest. Decision makers must be objective in determining which customers are given loans. Every data of prospective borrower customers is studied to see the pattern of customers in savings and loans at the cooperative. Data Mining is used to extract and identify information. Technique of Data Mining Decision Tree, the answer to a system is to help and look for decisions taking into account the factors that occur. This study aims to assess the customer from the data record in deciding the loan taking customer. This study aims to assess the customer from the data record in deciding the loan taking customer. Starting from data analysis, consists of 16 sample data: 7 customer data smoothly and 9 customer data defaults. In Node 1, the attribute with the highest Gain is obtained, the income is 0.6853. In Node 1.1, the attribute with the highest Gain is obtained, age, equal to 0.6215. The highest Gain is the time period: 0.9179. There are only two values from the time period attribute, 31-60, 15-30 days. attribute value 31-60 classifies the case into one, the decision is "jammed" as well as 15-30 days classifying the case into one, the decision is "smooth" so that there is no need to do further calculations.*

**Keywords:** Customer data; Customer pattern; Decision Tree; Data Mining; C4.5 algorithm

### Abstrak

Koperasi simpan pinjam berkembang dari bunga pinjaman nasabah. Pengambil keputusan harus objektif dalam menentukan nasabah yang diberi kredit meminjam. Setiap data nasabah calon peminjam dipelajari untuk melihat pola nasabah dalam simpan pinjam pada koperasi. Digunakan *Data Mining* untuk mengekstraksi dan mengidentifikasi informasi. Teknik *Data Mining Decision Tree*, jawaban akan sebuah sistem untuk membantu dan mencari keputusan dengan memperhitungkan faktor-faktor yang terjadi. Penelitian ini bertujuan menilai nasabah dari record data dalam memutuskan nasabah pengambil kredit. Analisa data, terdiri dari 16 data sample, 7 data nasabah lancar dan 9 data nasabah macet pembayaran. Setiap percabangan menyatakan kondisi yang harus dipenuhi dan tiap ujung pohon menyatakan kelas data. Pada *Node 1* diperoleh atribut dengan *Gain* tertinggi: Penghasilan, sebesar 0,6853. Pada *Node 1.1* diperoleh atribut dengan *Gain* tertinggi: umur, sebesar 0,6215. *Gain* tertinggi adalah jangka waktu: 0,9179. Ada tinggal dua nilai dari atribut jangka waktu, 31-60, 15-30 hari. nilai atribut 31-60 mengklasifikasikan kasus menjadi satu yaitu keputusannya "macet" demikian juga dengan 15-30 hari mengklasifikasi kasus menjadi satu yaitu keputusannya "lancar" sehingga tidak perlu melakukan perhitungan lebih lanjut.

**Kata Kunci:** Data nasabah; Pola nasabah; *Decision Tree*; *Data Mining*; Algoritma C4.5

### 1. Pendahuluan

Menurut Undang-Undang Republik Indonesia Nomor 25 Tahun 1992 Pasal 1 ayat (1) tentang Perkoperasian adalah Koperasi adalah badan usaha yang beranggotakan orang-orang atau badan hukum koperasi dengan melandaskan kegiatannya berdasarkan prinsip-prinsipkoperasi sekaligus sebagai gerakan ekonomi rakyat yang berdasarkan atas asas kekeluargaan [1]. Koperasi simpan pinjam adalah merupakan koperasi yang meningkatkan kesejahteraan anggotanya dengan kegiatan kredit berbunga rendah. Data calon peminjam diserahkan kepada Ketua koperasi untuk dianalisa, kemudian ketua koperasi memberikan hasil keputusannya kepada calon peminjam layak atau tidaknya calon peminjam diberi pinjaman [2]. Koperasi simpan pinjam berkembang dari bunga pinjaman nasabah. Kriteria peminjam yaitu

nasabah baru, nasabah tidak bermasalah dan nasabah bermasalah. Tipe nasabah dapat diamati dari transaksi keluar masuk uang pada data nasabah. Pemohon kredit dikategorikan lancar, kurang lancar dan macet. Dibutuhkan suatu metode dalam menentukan layak tidaknya nasabah diberikan pinjaman berbentuk kredit. *Data Mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data Mining* adalah proses pencarian pola-pola yang tersembunyi (*hidden patern*) berupa pengetahuan (*knowledge*) yang tidak diketahui sebelumnya dari suatu sekumpulan data yang mana data tersebut dapat berada di dalam *database*, data *warehouse*, atau media penyimpanan informasi yang lain [3]. Dengan nilai data yang bervariasi, kejadian diuraikan oleh koleksi atribut dan mempunyai salah satu dari satu set kelas yang eksklusif. Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Setelah sebuah pohon keputusan dibangun maka dapat digunakan untuk mengklasifikasikan *record* yang belum ada kelasnya. Dimulai dari *Node root*, menggunakan tes terhadap atribut dari *record* yang belum ada kelasnya tersebut lalu mengikuti cabang yang sesuai dengan hasil dari tes tersebut, yang akan membawa kepada *internal Node* (*Node* yang memiliki satu cabang masuk dan dua atau lebih cabang yang keluar), dengan cara harus melakukan tes lagi terhadap atribut atau *Node* daun. *Record* yang kelasnya tidak diketahui kemudian diberikan kelas yang sesuai dengan kelas yang ada pada *Node* daun. Pada pohon keputusan setiap simpul daun menandai label kelas. Proses dalam pohon keputusan yaitu mengubah bentuk data (tabel) menjadi model pohon (*tree*) kemudian mengubah model pohon menjadi aturan (*rule*) [4].

Dalam penelitian klasifikasi nasabah menggunakan algoritma C4.5 sebagai dasar pemberian kredit dijelaskan proses pengevaluasian permintaan kredit, dilakukan melalui penilaian kondisi perusahaan calon debitur yang diperkirakan dapat mempengaruhi kemampuan nasabah dalam memenuhi kewajiban. Penelitian yang menerapkan metode *Decision Tree* algoritma C4.5 dalam mengklasifikasi data penjualan bisnis gerai makanan cepat saji, bertujuan untuk klasifikasi penjualan menu makanan yang paling digemari [5].

## 2. Metode Penelitian

Dalam penelitian ini dilakukan pendekatan R&D (*Research and Development*). Tahap awal dimulai dengan menganalisa data nasabah dilanjutkan implementasi *Data Mining* pada data *sample* [6].

### 2.1. Analisa Data

Data nasabah:

#### 1. Data nasabah

Data nasabah adalah data nasabah kredit mandiri yang digunakan untuk acuan dalam pengaplikasian. Data nasabah kredit dengan atribut *id\_nasabah*, nama, umur, status, penghasilan, angsuran, jangka waktu, dan keterangan.

#### 2. Data nasabah pembayaran lancar.

Dari 16 data nasabah diambil 7 data nasabah lancar.

#### 3. Data Nasabah Pembayaran Macet

Dari 16 data nasabah diambil 9 data nasabah macet.

Proses pengolahan data kredit nasabah :

1. Data nasabah kredit dengan atribut *id\_nasabah*, nama, umur, status, penghasilan, *peminjaman\_usaha*, angsuran, jangka waktu, dan label di berikan kepada administrasi lalu di berikan kepada kepala koperasi sehingga berupa laporan data kredit nasabah, pembayaran lancar dan pembayaran macet.

2. Nasabah memberikan *document* kredit kepada administrasi, dan administrasi memberikan kepada kepala koperasi.

3. Kepala koperasi memberikan laporan data kredit nasabah pembayaran lancar dan data kredit nasabah pembayaran macet kepada administrasi, dan administrasi memberikan kepada karyawan koperasi.

Kriteria keterlambatan kredit pada koperasi yaitu pengawasan khusus, kurang lancar dan macet.

### 2.2. Data Mining

*Data Mining* sebagai proses untuk menemukan korelasi atau pola dari ratusan atau ribuan *field* dari sebuah relasional basis data yang besar. *Data Mining* adalah serangkaian

proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat pada basis data. *Data Mining* terutama digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut *Knowledge Discovery Databases* (KDD) [6]. Hal penting yang terkait di dalam *Data Mining* adalah:

1. *Data Mining* merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.
3. Tujuan *Data Mining* adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat. Kemampuan *Data Mining* dalam mencari informasi bisnis yang berharga dari basis data yang sangat besar, dapat dianalogikan dengan penambangan logam mulia dari lahan sumbernya, teknologi ini dipakai untuk [7] :
  - a) Prediksi dan sifat-sifat bisnis. *Data Mining* secara otomatis melakukan proses pencarian informasi untuk memprediksi basis data dalam jumlah besar.
  - b) Penemuan pola-pola yang tidak diketahui sebelumnya. *Data Mining* “menyapu” basis data, kemudian mengidentifikasi pola-pola yang sebelumnya tersembunyi dalam satu sapuan.

Tahapan proses *Data Mining* [8] :

1. *Data Cleaning*, tahap pembersihan data yang tidak konsisten.
2. *Data Integration*, langkah menggabungkan data dari beberapa sumber.
3. *Data Selection*, data yang tidak dikembalikan lagi ke *database* setelah proses *data cleaning*.
4. *Data Transformation*, data berubah atau bersatu menjadi bentuk yang tepat untuk menambang dengan ringkasan performa atau operasi *regrasi*.
5. *Data Mining*, merupakan proses yang digunakan untuk mengolah suatu data dengan menggunakan metode.
6. *Evaluation and presentation*, pengidentifikasian pola berdasarkan tindakan yang digunakan.
7. *Knowledge*, hasil yang dicapai berupa pengetahuan atau sebuah informasi.

### 2.3. *Decision Tree*

*Decision Tree* merupakan metode klasifikasi yang paling sering digunakan. Dalam pengerjaannya tidak memerlukan waktu yang lama dan hasilnya pun mudah untuk dipahami dan banyak penelitian dalam kasus ini sering menggunakan *Decision Tree* untuk mendapatkan hasil yang maksimal [9]. Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

1. Pilih atribut sebagai akar
2. Buat cabang untuk masing-masing nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Tiga jenis *Node* pada *Decision Tree* :

- a. *Root Node*, merupakan *Node* paling atas, pada *Node* ini tidak ada masukan dan bisa tidak mempunyai luaran lebih dari satu.
- b. *Internal Node*, merupakan *Node* percabangan, pada *Node* ini hanya terdapat satu masukan dan mempunyai luaran minimal 2.
- c. *Leaf Node* atau *terminal Node*, merupakan *Node* terakhir, pada *Node* ini hanya terdapat satu masukan dan tidak mempunyai luaran.

Dalam induksi *Decision Tree* adalah bagaimana menyatakan syarat pengujian pada *Node*. Ada 3 kelompok penting syarat pengujian *Node*:

1. Fitur biner  
Memiliki 2 nilai berbeda disebut dengan fitur biner. Syarat pengujian ketika fitur ini menjadi *Node* (akar maupun internal) hanya punya dua pilihan cabang.
2. Fitur bertipe kategorikal  
Fitur yang nilainya bertipe kategorikal (nominal atau ordinal) bisa mempunyai beberapa nilai berbeda.
3. Fitur bertipe *numeric*  
Fitur bertipe numerik, syarat pengujian dalam *Node* (akar maupun internal) dinyatakan dengan perbandingan pengujian ( $A < v$ ) atau ( $A \geq v$ ) dengan hasil biner, atau untuk multi dengan hasil berupa jangkauan nilai dalam bentuk  $v_i \leq A < v_{i+1}$ , untuk  $i=1,2,\dots,k$ .

## 2.4. Algoritma C4.5

Algoritma C4.5 adalah satu algoritma untuk klasifikasi penggolongan data. Dalam algoritma C4.5 sebelum membangun pohon keputusan hal yang paling perlu dilakukan adalah menentukan atribut sebagai akar. Kemudian dibuat cabang untuk setiap nilai dalam akar tersebut. Langkah selanjutnya adalah membagi kasus dalam cabang. Kemudian mengulangi proses pada setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama [10]. Untuk memilih atribut dengan akar, didasarkan pada nilai *Gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung *Gain* digunakan persamaan:

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

Keterangan:

S: Himpunan kasus; A: Atribut ; N: Jumlah partisi atribut A  
 |Si|: Jumlah kasus pada partisi ke-i ; |S|: Jumlah kasus dalam S

Perhitungan nilai entropi dapat dilihat pada persamaan:

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

Keterangan :

S: Himpunan kasus; A: Atribut ; n: Jumlah partisi S ; pi : Proporsi dari Si terhadap S

## 2.5. Klasifikasi

Klasifikasi merupakan proses untuk menemukan fungsi dan model yang dapat membedakan atau menjelaskan konsep atau kelas data dengan tujuan memperkirakan kelas yang tidak diketahui dari suatu objek. Dalam proses pengklasifikasian biasa terdapat dua proses yang harus dilakukan, yaitu :

1. Proses *Training*. Pada proses ini akan digunakan data *training set* atau data sampel yang telah diketahui label-label atau atribut dari data sampel tersebut untuk membangun model.
2. Proses *Testing*. Pada proses *testing* ini dilakukan untuk mengetahui keakuratan model yang telah dibuat pada proses *training* maka dibangun data yang disebut dengan data *testing* untuk mengklasifikasikan label-labelnya.

## 3. Hasil dan Pembahasan

Untuk menentukan akar dari pohon keputusan ditentukan oleh *Gain* yang tertinggi, sebelum menemukan *Gain* terlebih dahulu menghitung entropi keseluruhan dan entropi dari setiap atribut.

### 3.1. Pengolahan Data

Pada penelitian ini menggunakan data primer dengan data nasabah kredit dengan atribut *nik\_nasabah*, umur, satus, status, penghasilan, angsuran, jangka waktu, dan keterangan. Dari data tersebut diambil 16 data sample yang digunakan sebagai contoh perhitungan manual entropi dan *Gain*, data *sample* memiliki 7 data nasabah lancar dan 9 data nasabah macet, atribut nasabah kredit terdapat pada Tabel 1, data *testing* pada Tabel 2.

Tabel 1. Atribut nasabah kredit

Atribut	Detail pengguna
Umur	18-55
Status	Belum/menikah
Penghasilan	Rendah, tinggi, sedang
Angsuran	<=210rb-1500rb
Jangka waktu	1-14,15-30,31-60,60>
Keterangan	Lancar/macet

Tabel 2. Data *testing*

Umur	Status	Penghasilan	Angsuran	Jangka waktu	Keterangan
18-25 <sup>th</sup>	Menikah	Rendah	<=700000	1-14	Macet
26-29 <sup>th</sup>	Belum menikah	Tinggi	>1500000	15-30	Lancar
18-25 <sup>th</sup>	Menikah	Sedang	>=800000-1500000	15-30	Macet
26-29 <sup>th</sup>	Belum menikah	Rendah	<=700000	>61	Macet
30-65 <sup>th</sup>	Menikah	Rendah	<=700000	31-60	Macet
26-29 <sup>th</sup>	Belum menikah	Sedang	>=800000-1500000	15-30	Lancar
26-29 <sup>th</sup>	Menikah	Tinggi	>1500000	15-30	Lancar
30-65 <sup>th</sup>	Menikah	Rendah	<=700000	>61	Macet
30-65 <sup>th</sup>	menikah	Tinggi	>=800000-1500000	15-30	Lancar
26-29 <sup>th</sup>	Belum menikah	Sedang	<=700000	15-30	Lancar
18-25 <sup>th</sup>	Menikah	Rendah	<=700000	31-60	Macet
18-25 <sup>th</sup>	Menikah	Tinggi	>1500000	1-14	Lancar
26-29 <sup>th</sup>	Belum menikah	Sedang	>=800000-1500000	31-60	Macet
26-29 <sup>th</sup>	Belum menikah	Rendah	<=700000	15-30	Macet
18-25 <sup>th</sup>	Menikah	Rendah	<=700000	1-14	Macet
30-65 <sup>th</sup>	Menikah	Sedang	>=800000-1500000	1-14	Lancar

### 3.2. Pemodelan *Data Mining* Menggunakan Algoritma C4.5

Perhitungan nilai entropi dan *Gain* pada setiap atribut yang memiliki label Lancar dan Macet :

1. Perhitungan nilai entropi  
Langkah awal Algoritma C4.5 adalah mencari nilai *entropi*. Pertama, tentukan terlebih dahulu nilai entropi total dalam kasus.
2. Perhitungan nilai *Gain*  
Setelah semua perhitungan entropi pada masing-masing subset atribut selesai, kemudian lakukan perhitungan nilai *Gain*.

Perincian algoritma:

- 1) Menghitung jumlah kasus seluruhnya, jumlah berkeputusan “Lancar” maupun “Macet”.
- 2) Menghitung entropi dari semua kasus yang terbagi berdasarkan atribut “umur”, “status”, “penghasilan”, “angsuran”, dan “jangka waktu”.
- 3) Penghitungan *Gain* untuk setiap atributnya
- 4) Entropi total (S) dengan persamaan :

$$\left(-\left(\frac{jml.lancar}{totalkasus}\right) * \log_2 \left(\frac{jml.lancar}{totalkasus}\right) + \left(-\left(\frac{jml.macet}{totalkasus}\right) * \log_2 \left(\frac{jml.macet}{totalkasus}\right)\right) \right) \quad (3)$$

$$E_{total} (7,9) = \left(-\left(\frac{7}{16}\right) * \log_2 \left(\frac{7}{16}\right) + \left(-\left(\frac{9}{16}\right) * \log_2 \left(\frac{9}{16}\right)\right)\right) \\ = \left(-\left(0,4375\right) * \left(\frac{-0,3590}{0,3010}\right) + \left(-\left(0,5625\right) * \left(\frac{-0,2498}{0,3010}\right)\right)\right) = 0,9886$$

Selanjutnya perhitungan entropi tiap atribut berdasarkan jumlah kasus per subset atribut:

a. Perhitungan nilai subset pada atribut umur

E [umur:18-25] (1,4)

$$= \left(-\left(\frac{1}{5}\right) * \log_2 \left(\frac{1}{5}\right) + \left(-\left(\frac{4}{5}\right) * \log_2 \left(\frac{4}{5}\right)\right)\right) = \left(-\left(0,2\right) * \left(\frac{-0,6909}{0,3010}\right) + \left(-\left(0,8\right) * \left(\frac{-0,3997}{0,3010}\right)\right)\right) = 0,7218$$

E [umur:26-29] (4,3)

$$= \left(-\left(\frac{4}{7}\right) * \log_2 \left(\frac{4}{7}\right) + \left(-\left(\frac{3}{7}\right) * \log_2 \left(\frac{3}{7}\right)\right)\right) = \left(-\left(0,5714\right) * \left(\frac{-0,2430}{0,3010}\right) + \left(-\left(0,4285\right) * \left(\frac{-0,3679}{0,3010}\right)\right)\right) = 1,0059$$

E [umur:30-65] (2,2)

$$= (-2/4) * \log_2(2/4) + (-2/4) * \log_2(2/4) = -(0,5) * (\frac{-0,3010}{0,3010}) + -(0,4285) * (\frac{-0,3010}{0,3010}) = 1$$

Gain umur

$$= 0,9886 - (5/16 * 0,7218) + (7/16 * 1,0059) + (4/16 * 1) \\ = 0,9886 - (0,3125 * 0,7218) + (0,4375 * 1,0059) + (0,25 * 1) = 0,6394$$

b. Perhitungan nilai subset pada atribut status E [belum menikah] (3,3)

$$= (-3/6) * \log_2(3/6) + (-3/6) * \log_2(3/6) = -(0,5) * (\frac{-0,3010}{0,3010}) + -(0,5) * (\frac{-0,3010}{0,3010}) = 1$$

E [menikah] (4,6)

$$= (-4/10) * \log_2(4/10) + (-6/10) * \log_2(6/10) = -(0,4) * (\frac{-0,3979}{0,3010}) + -(0,6) * (\frac{-0,2218}{0,3010}) \\ = 0,9707$$

Gain status

$$= 0,9886 - (10/16 * 0,9707) + (6/16 * 1) = 0,007$$

c. Perhitungan nilai subset pada atribut penghasilan

E [sedang] (3,2)

$$= (-3/5) * \log_2(3/5) + (-2/5) * \log_2(2/5) = -(0,6) * (\frac{-0,2218}{0,3010}) + -(0,4) * (\frac{-0,3979}{0,3010}) = \\ 0,9707$$

E [rendah] (7) = 0 ; E [tinggi] (4) = 0

Gain penghasilan

$$= 0,9886 - (4/16 * 0) + (5/16 * 0,9707) + (7/16 * 0) = 0,9886 - (0) + (0,3125 * 0,9707) + (0) \\ = 0,6853$$

d. Perhitungan nilai subset pada atribut angsuran

E <= 700000 (1,7)

$$= (-1/8) * \log_2(1/8) + (-7/8) * \log_2(7/8) = -(0,125) * (\frac{-0,9030}{0,3010}) + -(0,875) * (\frac{-0,0579}{0,3010}) \\ = 0,5432$$

E >= 800000-1500000 (3,2)

$$= (-3/5) * \log_2(3/5) + (-2/5) * \log_2(2/5) = -(0,6) * (\frac{-0,2218}{0,3010}) + -(0,4) * (\frac{-0,3997}{0,3010}) = \\ 0,9707$$

E > 1500000 (3) = 0

Gain angsuran

$$= 0,9886 - (8/16 * 0,5432) + (5/16 * 0,9707) + (3/16 * 0) \\ = 0,9886 - (0,5 * 0,5432) + (0,3125 * 0,9707) + (0) = 0,4137$$

e. Perhitungan nilai subset pada atribut jangka waktu

E [1-14] (2, 2)

$$= (-2/4) * \log_2(2/4) + (-2/4) * \log_2(2/4) = -(0,5) * (\frac{-0,3010}{0,3010}) + -(0,5) * (\frac{-0,3010}{0,3010}) = -(0,5) * \\ (-1) + -(0,5) * (-1) = 1$$

E [31-60] (3) = 0

E [15-30] (5,2)

$$= (-4/7) * \log_2(4/7) + (-2/7) * \log_2(2/7) = -(0,5714) * (\frac{-0,2430}{0,3010}) + -(0,2857) * (\frac{-0,5440}{0,3010}) \\ = -(0,5714) * (-0,8073) + -(0,2857) * (-1,8073) = 0,9775$$

E [>61] (2) = 0

Gain jangka waktu

$$= 0,9886 - (4/16 * 1) + (3/16 * 0) + (7/16 * 0,9775) + (2/16 * 0) = 0,311$$

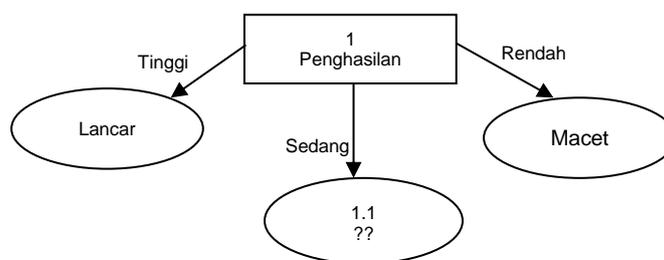
Berikut adalah tampilan hasil sementara dalam Tabel 3:

Tabel 3. Perhitungan Node 1

Node	Atribut	Simpul	Jumlah Kasus	Kolektibilitas		Entropi	Gain
				Lancar	Macet		
1	Total		16	7	9	0,9886	
	Umur						0,6394
		18-25 <sup>th</sup>	5	1	4	0,7218	
		26-29 <sup>th</sup>	7	4	3	1,0059	
		30-65 <sup>th</sup>	4	2	2	1	

Status					0,007
	Menikah	10	4	6	0,9707
	Belum menikah	6	3	3	1
Penghasilan					0,6853
	Tinggi	4	4	0	0
	Sedang	5	3	2	0,9707
	Rendah	7	0	7	0
Angsuran					0,4137
	<=26000	8	1	7	0,5432
	>=126000-500000	5	3	2	0,9707
	500000	3	3	0	0
Jangka waktu					0,311
	1-14	4	2	2	1
	31-60	3	0	3	0
	15-30	7	5	2	0,9775
	>61	2	0	2	0

Dari hasil Tabel 3. diketahui bahwa atribut dengan *Gain* tertinggi adalah Penghasilan, sebesar 0,6853. dengan demikian, penghasilan sebagai *Node akar*. Ada tiga nilai dari atribut penghasilan yaitu tinggi, sedang dan rendah. Nilai atribut tinggi dan rendah sudah mengklasifikasikan kasus menjadi satu yaitu keputusannya “lancar dan macet”, sehingga tidak perlu dilakukan perhitungan lebih lanjut. Tetapi untuk nilai atribut sedang masih perlu dilakukan perhitungan lagi, karena masih terdapat “lancar” dan “macet”. Gambar pohon keputusan sementara, Gambar 1:



Gambar 1. Pohon keputusan *Node 1*

Berdasarkan pembentukan pohon keputusan di atas, *Node 1.1* akan dianalisis lebih lanjut. Berikut pada Tabel 4, data yang memiliki penghasilan = sedang

Tabel 4. Data yang memiliki atribut penghasilan = sedang

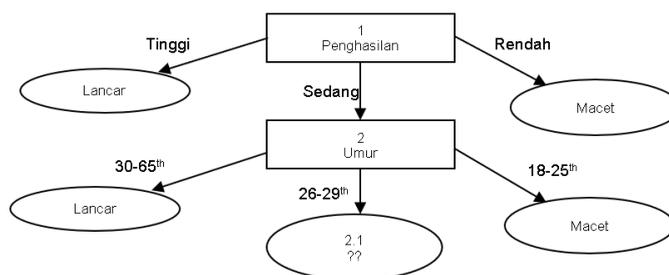
Umur	Status	Penghasilan	Angsuran	Jangka waktu	Keterangan
18-25 <sup>th</sup>	Menikah	Sedang	>=126000-500000	15-30	Macet
26-29 <sup>th</sup>	Belum menikah	Sedang	>=800000-1500000	15-30	Lancar
26-29 <sup>th</sup>	Belum Menikah	Sedang	<=26000	15-30	Lancar
26-29 <sup>th</sup>	Belum menikah	Sedang	>=800000-1500000	31-60	Macet
30-65 <sup>th</sup>	Menikah	Sedang	>=800000-1500000	1-14	Lancar

Kemudian data di Tabel 4. dihitung lagi entropi Atribut Penghasilan – sedang dan entropi setiap atribut serta *Gain*, diperoleh hasilnya pada Tabel 5.

Tabel 5. Hasil perhitungan *Node* 1.1

Node	Atribut	Simpul	Jumlah Kasus	Kolektibilitas		Entropi	Gain	
				Lancar	Macet			
1.1	Penghasilan	Sedang	5	3	2	0,9707	0,6215	
		Umur						
		18-25 <sup>th</sup>	1	0	1	0		
		26-29 <sup>th</sup>	3	2	1	0,9179		
		30-65 <sup>th</sup>	1	1	0	0		
	Status							0,02
		Menikah	2	1	1	1		
		Belum Menikah	3	2	1	0,9179		
	Angsuran							0,1707
		<=26000	1	1	0	0		
		>=126000-500000	4	2	2	1		
	Jangka Waktu							0,42
		1-14	1	1	0	0		
		31-60	1	0	1	0		
	15-30	3	2	1	0,9179			

Dari hasil Tabel 5 diketahui bahwa Atribut dengan *Gain* tertinggi adalah **umur** = 0,6215. dengan demikian, umur sebagai *Node* akar berikutnya. Ada tiga nilai dari atribut umur yaitu 18-25<sup>th</sup>, 26-29<sup>th</sup> dan 30-65<sup>th</sup>. Nilai atribut 18-25<sup>th</sup> dan 30-65<sup>th</sup> sudah mengklasifikasikan kasus menjadi satu yaitu keputusannya “macet dan lancar”, sehingga tidak perlu dilakukan perhitungan lebih lanjut. Untuk nilai atribut 26-29<sup>th</sup> masih perlu dilakukan perhitungan lagi (Gambar 2), karena masih terdapat “lancar” dan “macet”.



Gambar 2. Pohon keputusan *Node* 2

Berdasarkan pembentukan pohon keputusan di atas, *Node* 2.1 akan dianalisis lebih lanjut. Berikut data yang memiliki Umur = 26-29<sup>th</sup> Tabel 6.

Tabel 6. Data yang memiliki atribut umur = 26-29<sup>th</sup>

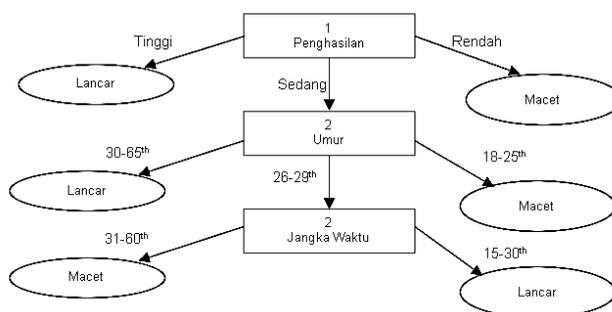
No	Umur	Status	Penghasilan	Angsuran	Jangka waktu	Keterangan
2	26-29 <sup>th</sup>	Belum menikah	Sedang	>=800000-1500000	15-30	Lancar
3	26-29 <sup>th</sup>	Belum Menikah	Sedang	<=26000	15-30	Lancar
4	26-29 <sup>th</sup>	Belum menikah	Sedang	>=800000-1500000	31-60	Macet

Kemudian data di Tabel 4 dihitung lagi entropi Atribut Penghasilan – sedang dan entropi setiap Atribut serta *Gain*, sehingga hasilnya seperti pada Tabel 7:

Tabel 7. Hasil perhitungan *Node 2.1*

Node	Atribut	Simpul	Jumlah Kasus	Kolektibilitas		Entropi	Gain
				Lancar	Macet		
2.1	Umur	26-29 <sup>th</sup>	3	2	1	0,9179	0
	Status	Belum Menikah	3	2	1	0,9179	
	Penghasilan	Sedang	3	2	1	0,9179	0
	Angsuran	<=26000	1	1	0	0	0,2513
	>=126000-50000	2	1	1	1		
	Jangka Waktu	31-60	1	0	1	0	0,9179
		15-30	2	2	0	0	

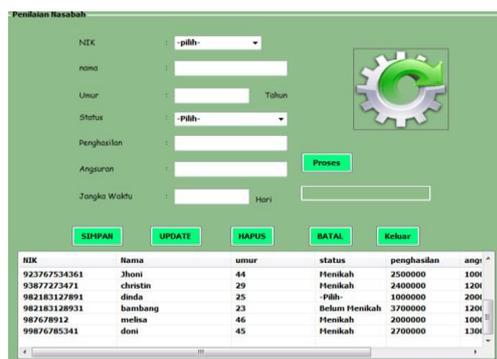
Selanjutnya menentukan Atribut yang memiliki *Gain* tertinggi untuk dibuatkan *Node* berikutnya. *Gain* tertinggi adalah jangka waktu pada Tabel 7 sebesar 0,9179. Ada tinggal dua nilai dari atribut jangka waktu, 31-60, 15-30 hari. nilai atribut 31-60 sudah mengklasifikasikan kasus menjadi satu yaitu keputusannya “macet” demikian juga dengan 15-30 hari sudah mengklasifikasi kasus menjadi satu yaitu keputusannya “lancar” sehingga tidak perlu melakukan perhitungan lebih lanjut, Gambar 3.



Gambar 3. Pohon keputusan *Node 2.1*

### 3.3. Implementasi Sistem

Implementasi sistem merupakan tahapan dalam pengembangan sistem, untuk melakukan sebuah implementasi maka diperlukan perancangan *interface* guna untuk interaksi antara *user* dengan sistem dan penulisan kode program yang sesuai dengan kebutuhan sistem aplikasi *Data Mining* yang dirancang.



(a) Form data training

(b) Form data penilaian



NIK	Namasabah	Penghasilan	Umur	Jangka waktu (hari)	Keterangan
93877273471	christin	2400000	29	45	Macet
91918888121	ribka simorangkir	1500000	23	12	Lancar
99876785341	doni	2700000	45	39	Lancar
912121756771	Adi	1700000	29	34	Macet
923767534361	Jhoni	2500000	44	30	Lancar
982183128931	bambang	3700000	23	140	Lancar
1434555351	julika	2000000	32	60	Lancar
9172763811	Ana	2000000	27	14	Lancar
982183127891	dinda	1000000	25	180	Macet
9113318383	denny	3000000	25	14	Lancar
92113311	Rusdi	2900000	36	23	Lancar
987678912	melisa	2000000	46	120	Lancar

(c) Form data training

(d) Hasil laporan

Gambar 4. Implementasi sistem

Sistem yang dibangun (Gambar 4) terdiri dari *Form data nasabah* Gambar 4(a) *user* memasukkan data nasabah berdasarkan *database* yang sudah tersedia sebelumnya. *Form data training* Gambar 4(b) dihasilkan nilai entropi dan *Gain* dan *rule* berdasarkan *Data Mining* algoritma C4.5 hasil matriks dengan pembayaran berdasarkan pembayaran lancar atau macet. *Form data penilaian* Gambar 4(c) dihasilkan pembayaran macet atau lancar berdasarkan *rule* atau aturan perhitungan akhir *Data Mining* algoritma C4.5 penghasilan, umur, dan jangka waktu hasil matriks. Laporan hasil penilaian Gambar 4(d).

#### 4. Kesimpulan

Berdasarkan pohon keputusan terakhir maka didapat *rule* dengan aturan jika penghasilan “tinggi” maka “lancar”, jika penghasilan “rendah” maka “macet”, jika penghasilan “sedang” dan umur di atas “30 tahun” maka “lancar”, jika penghasilan “sedang” dan umur di bawah “26 tahun” maka “macet”, jika penghasilan “sedang” dan umur “26-29 tahun” dan jangka waktu diatas “30 hari” maka “macet”, jika penghasilan “sedang” dan umur “26-29 tahun” dan jangka waktu dibawah “30 hari” maka “lancar”. *Data Mining* dengan *Decision Tree* mampu menghasilkan sebuah pengambilan keputusan yang berpotensi untuk dipertimbangkan dari setiap penerimaan nasabah dengan menghitung *Data Mining* keterangan pembayaran nasabah lancar atau macet berdasarkan metode algoritma C4.5.

#### Referensi

- [1] H. Sholihah, S. I. Hidayat, N. Yulianti, “Persepsi Dan Sikap Nasabah Dalam Memperoleh Kredit Usaha Agribisnis Pada Bank Konvensional Dan Bank Syariah”, *Jurnal Sosial Ekonomi Pertanian*, vol. 7, no. 1, hal. 24-31, 2014.
- [2] A. S. R. M. Sinaga, “Data Mining Informasi Tingkat Kelulusan Mahasiswa STMIK Pelita Nusantara Medan”, *Jurnal Mantik Penusa*, vol.1, no. 1, hal. 27-35, 2017.
- [3] F. Hadi, “Penerapan *Data Mining* Dalam Menganalisa Pemberian Pinjaman Dengan Menggunakan Metode Algoritma C5.0 (Studi Kasus : Koperasi Jasa Keuangan Syariah Kelurahan Lambung Bukik)”, *Jurnal KOMTEKINFO*, vol. 4, no. 2, hal. 214-223, 2017.
- [4] Eka Pandu C., Edi I., “Metode *Decision Tree* Algoritma C.45 Dalam Mengklasifikasi Data Penjualan Bisnis Gerai Makanan Cepat Saji”, *Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)*, vo.3, pp 1-13, 2018.
- [5] Siska H., A. Sudarsono, E. Suryana, “Implementasi *Data Mining* Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu)”, *Jurnal Media Infotama*, vol. 11. no. 2, hal. 130-138, 2015.
- [6] P. Meilina., “Penerapan *Data Mining* Dengan Metode Kalsifikasi Menggunakan *Decision Tree* Dan Regresi”, *Jurnal Teknologi*, vol 7, no. 1, hal. 11-20, 2015.
- [7] S Al Syahdan, A Sinda, “*Data Mining* Penjualan Produk Dengan Metode Apriori Pada Indomaret Galang”, *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)*, vol. 1, no. 2, hal. 56 – 63, 2018.
- [8] Durrotul M., Ana K., “Implementasi Data Mining Dalam Prediksi *Performance Software Engineer* Pt. Emesrio Menggunakan Decision Tree”, *Jurnal Ilmiah Informatika dan Komputer*, vol. 22, no. 1, hal. 31-43, 2017.

- [9] Beti N., Tedy R., Syamsul B., "Implementasi *Data Mining* Dengan Algoritma C4.5 Untuk Penjurusan Siswa (Studi Kasus: SMA Negeri 1 Pontianak)", *Coding Jurnal Komputer dan Aplikasi*, vol. 4, no.3, hal. 75-84, 2016.
- [10] Alfannisa Annurullah F., Algifanri M., "Penerapan *Data Mining* Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma Fpgrowth Pada Data Transaksi Penjualan Spare Part Motor", *Kumpulan jurnaL Ilmu Komputer (KLIK)*, vol. 5, no. 1, hal. 27-36, 2018.