

## PENERAPAN ALGORITMA KLASIFIKASI DATA MINING DALAM PENENTUAN PEMBERIAN PINJAMAN KOPERASI

**Nandang Iriadi**

Program Studi Manajemen Informatika  
Akademik Manajemen Informatika dan Komputer Bina Sarana Informatika  
Jl. Rs.Fatmawati No.24 Jakarta Selatan  
[nandang.ndi@bsi.ac.id](mailto:nandang.ndi@bsi.ac.id)

### **ABSTRACT**

*The loan is the delivery of goods, services, or money from one party (the creditor / lender) on the basis of trust in the other party (customer or debtor / borrower) to the appointment of a receiver to pay credit to the lender on the date agreed upon by both parties. Research each loan application is highly dependent on factors such as type of business, the economy, the intended use of the credit, the amount of credit that do not kredit. Analisa carefully and thoroughly resulted in some borrowers who have no credit given the ability to make payments so that there was non-performing loans. In this study, using the method of C4.5 algorithm, which was applied to the data members Multipurpose Cooperative Enterprises "Ceger Jaya" that get good credit loans are problematic in the installment payment or not. From the test results to measure the performance of the algorithm testing using Cross Validation, Confusion Matrix and the ROC curve, it is known that the C4.5 algorithm has the highest accuracy value, which is 90.67 which is very good classification.*

**Keywords:** C 4.5, Credit Analysis, Decision Tree.

### **1. PENDAHULUAN**

Koperasi sebagai salah satu bentuk organisasi yang penting dalam meningkatkan pertumbuhan ekonomi. Koperasi simpan pinjam menjadi salah satu alternative bagi masyarakat untuk mendapatkan dana dalam upaya memperbaiki taraf kehidupan, pemenuhan kebutuhan sehari-hari dan mengembangkan usaha. Tidak dipungkiri, memberikan pinjaman dana kepada nasabah pasti akan muncul permasalahan-permasalahan, seperti nasabah terlambat membayarkan cicilan dana, penyalahgunaan dana untuk keperluan lain, nasabah gagal mengembangkan usahanya sehingga dapat mengakibatkan dana di koperasi tidak mengalir atau dapat mengakibatkan kredit macet. Oleh karena itu, penelitian ini dilakukan untuk membantu menyelesaikan permasalahan tersebut dengan merancang sebuah aplikasi *data mining* yang berfungsi untuk memprediksi perilaku anggota koperasi yang berpotensi melakukan peminjaman (kredit) terhadap Koperasi Serba Usaha (KSU) Ceger Jaya yang berlokasi di Ceger-Cipayung Jakarta Timur, merupakan tempat yang dipilih oleh peneliti sebagai studi kasus. Jumlah anggota yang mengajukan pinjaman dana kredit pada tahun 2010 sebanyak 762 orang dengan nasabah bermasalah sebanyak 50 orang dan pada tahun 2011 jumlah nasabahnya sebanyak 974 orang

dengan nasabah bermasalah sebanyak 22 orang. Data yang dizinkan untuk diteliti adalah data pada tahun 2011. Untuk memprediksi perilaku nasabah kredit ini diperlukan suatu metode atau teknik yang dapat mengolah data-data yang sudah ada di koperasi tersebut. Salah satu metodenya dapat menggunakan teknik *data mining*. Dari fungsi *data mining* yang sudah disebutkan di atas akan digunakan dalam penelitian ini. Teknik *data mining* yang diterapkan pada aplikasi yang dibangun adalah klasifikasi sedangkan metode klasifikasi yang digunakan adalah *Decision Tree* (pohon keputusan). Algoritma yang dipakai sebagai algoritma pembentuk pohon keputusannya adalah *Algoritma C4.5*. Adapun data yang di proses dalam penelitian ini adalah data angsuran anggota pada Koperasi Serba Usaha Ceger Jaya pada tahun 2011 dalam format *Microsoft Excel*. Analisis kredit merupakan hal yang penting dalam lingkup resiko keuangan (Lai, Yu, Zhou, & Wang, 2006), oleh karena itu perlunya dilakukan analisa. Mengidentifikasi data debitur yang bermasalah merupakan hal yang sulit (Odeh, Featherstone, & Das, 2010). Data yang besar dan banyaknya parameter tentunya membutuhkan alat yang efektif dan efisien untuk melakukan analisa kredit dan menilai debitur yang mempunyai resiko gagal bayar dan yang tidak beresiko.

## 2. LANDASAN TEORI

### A. Koperasi

Koperasi adalah suatu kumpulan orang - orang yang bekerjasama demi kesejahteraan bersama. Badan usaha yang beranggotakan orang - orang atau badan hukum koperasi dengan melandaskan kegiatannya berdasarkan atas asas kekeluargaan. Berdasarkan pengertian tersebut, yang dapat menjadi anggota koperasi adalah perorangan (orang yang secara sukarela menjadi anggota koperasi), badan hukum koperasi (suatu koperasi yang menjadi anggota koperasi yang memiliki lingkup lebih luas).

### B. Kredit

Pengertian kredit dalam perusahaan dagang sering disebut kredit perdagangan. Arti kredit perdagangan menurut Kasmir (2011:272) adalah "penjualan barang jasa kepada pembeli secara kredit. Atau dengan kata lain, penjualan barang yang pembayarannya dilakukan secara angsuran (dicicil) utuk beberapa waktu." Melalui hasil analisis kreditnya, dapat diketahui apakah usaha nasabah layak (*feasible*) dan *marketable* (hasil usaha dapat dipasarkan), dan *profitable* (menguntungkan), serta dapat dilunasi tepat waktu. Untuk mewujudkan hal tersebut, perlu dilakukan persiapan kredit, yaitu dengan mengumpulkan informasi dan data untuk bahan analisis. Kualitas hasil analisis tergantung pada kualitas SDM, data yang diperoleh, dan teknik analisis. Analisis kredit merupakan hal yang penting dalam lingkup resiko keuangan (Lai, Yu, Zhou, & Wang, 2006), oleh karena itu perlunya dilakukan analisa. Namun, melakukan proses analisa kredit membutuhkan waktu lama (Kotsiantis, Kanellopoulos, Karioti, & Tampakas, 2009) dan mengidentifikasi data debitur yang bermasalah merupakan hal yang sulit (Odeh, Featherstone, & Das, 2010). Data yang besar dan banyaknya parameter tentunya membutuhkan alat yang efektif dan efisien untuk melakukan analisa kredit dan menilai debitur yang mempunyai resiko gagal bayar dan yang tidak beresiko. Penelitian setiap permohonan kredit sangat tergantung pada faktor-faktor seperti jenis usaha, sektor ekonomi, tujuan penggunaan kredit, jumlah kredit diharapkan dapat meminimalkan masuknya debitur yang bermasalah, karena semakin banyak debitur bermasalah maka akan semakin memperbesar tingkat kredit macet yang pada akhirnya dapat menimbulkan kebangkrutan (Zurada, 2010:19). bahwa ada beberapa atribut yang menyertai data debitur yaitu seperti umur, jumlah kredit, *checking*, penjamin, jangka kredit, lama bekerja, jumlah akun di bank, status pekerjaan, *histori* kredit,

status rumah, dana aman, status pernikahan, alasan pinjaman, dan lain-lain.

### C. Data Mining

Menurut Vercellis (2009:81) *Data mining* merupakan proses berulang-ulang yang ditujukan pada analisis *database*, dengan tujuan menyaring informasi dan pengetahuan yang dapat membuktikan keakuratan data dan berpotensi baik dan berguna untuk para professional dibidang ilmu pengetahuan yang terlibat dalam pengambilan keputusan dan pemecahan masalah. Menurut Maimon (2005:) *Data mining* merupakan inti dari proses *Knowledge Discovery in Database* (KDD). Menurut Maimon (2005) KDD adalah proses terorganisir untuk mengidentifikasi pola yang valid, baru, berguna, dan dapat dimengerti dari sebuah *data set* yang besar dan kompleks. KDD terdiri dari sembilan langkah. Berikut ini penjelasan singkat langkah-langkah dalam KDD (Maimon, 2005:2-5):

1. Pembentukan pemahaman domain aplikasi  
Pada tahap ini menentukan tujuan dari *end-user* dan bagian terkait dimana KDD dilakukan.
2. Memilih dan membuat *data set* dimana proses penemuan *knowledge* akan dilakukan  
Penentuan data yang akan digunakan untuk proses KDD dilakukan pada tahap ini.
3. *Preprocessing* dan *cleansing*  
Dalam tahap ini kehandalan data ditingkatkan. Termasuk *data clearing*, seperti menangani data yang tidak lengkap, menghilangkan gangguan atau *outlier*.
4. Transformasi data  
Pada tahap ini, membuat data menjadi lebih baik menggunakan metode reduksi dimensi dan transformasi atribut.
5. Memilih tugas *data mining* yang cocok  
Pada tahap ini ditentukan tipe *data mining* yang akan digunakan, apakah klasifikasi, regresi, atau *clustering*, tergantung pada tujuan KDD dan tahap sebelumnya.
6. Memilih algoritma *data mining*  
Pada tahap ini dilakukan pemilihan algoritma yang paling tepat untuk menemukan pola.
7. Penggunaan algoritma *data mining*  
Pada tahap ini dilakukan implementasi dari algoritma *data mining* yang telah ditentukan pada tahap sebelumnya.
8. Evaluasi  
Pada tahap ini dilakukan evaluasi dan penerjemahan dari pola yang didapat.
9. Penggunaan pengetahuan yang didapat  
Pada tahap ini, pengetahuan dimasukkan ke sistem lain dan mengaktifkan sistem

tersebut serta mengukur hasilnya. Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

Proses dalam tahap *data mining* terdiri dari tiga langkah utama, yaitu (Sumathi dan Sivandam, 2006:40):

#### 1. *Data Preparation*

Pada langkah ini, data dipilih, dibersihkan, dan dilakukan *preprocessed* mengikuti pedoman dan *knowledge* dari ahli domain yang menangkap dan mengintegrasikan data internal dan eksternal ke dalam tinjauan organisasi secara menyeluruh.

#### 2. Algoritma *data mining*

Penggunaan algoritma *data mining* dilakukan pada langkah ini untuk menggali data yang terintegrasi untuk memudahkan identifikasi informasi bernilai.

#### 3. Fase analisa data

Keluaran dari data mining dievaluasi untuk melihat apakah *knowledge* domain ditemukan dalam bentuk *rule* yang telah diekstrak dari jaringan.

### D. Model *Data Mining*

#### 1. *Decision Tree*

Pohon keputusan merupakan salah satu metode klasifikasi dan prediksi yang sangat kuat dan terkenal dalam penerapan *data mining*. Pada dasarnya *Decision Tree* mengubah data menjadi pohon keputusan (*decision tree*) dan aturan-aturan keputusan (*rule*).

Menurut Larose (2005:116) Hal yg menarik dari *Decision Tree* yaitu:

- Decision Tree* memiliki konsep yang mudah dimengerti.
- Decision Tree* membagi data pada setiap *leaf* tanpa kehilangan data sedikitpun.
- Decision Tree* cocok untuk digunakan dalam bidang bisnis ataupun pemasaran.

Menurut Gorunescu (2011:161) Pohon-pohon keputusan memiliki tiga pendekatan klasik:

#### a. Klasifikasi pohon

Istilah yang digunakan ketika hasil prediksi adalah klasifikasi kelas data

#### b. Regresi Pohon

Ketika hasil diprediksi dapat dianggap sebagai nilai yang sebenarnya (misalnya: harga minyak, nilai rumah, harga saham)

#### c. Klasifikasi (*C and RT*)

Klasifikasi dan pohon regresi ketika kita mengambil dan mempertimbangkan kedua kasus diatas.

#### 2. Klasifikasi

Menurut Bramer (2007:13) Klasifikasi adalah salah satu aplikasi yang paling umum untuk *data mining*.

### E. Tahap persiapan data (*Data Preparation*)

Tahap persiapan data mencakup semua kegiatan untuk mencari data yang akan dimasukkan ke dalam alat pemodelan, dari data mentah awal. Untuk itu maka diperlukan tehnik dalam *preprocessing* (Han & Kamber, 2007:50) yaitu:

#### 1. *Data Cleaning*

Bekerja untuk membersihkan nilai yang kosong, tidak konsisten atau mungkin tupel yang kosong (*missing values* dan *noisy*).

#### 2. *Data Integration*

Berfungsi menyatukan tempat penyimpanan (arsip) yang berbeda ke dalam satu data. Dalam hal ini, ada dua arsip yang diambil sebagai *data warehouse* yaitu data anggota dan data kredit.

#### 3. *Data size reduction and discretization*

Untuk memperoleh data set dengan jumlah atribut dan *record* yang lebih sedikit tetapi bersifat informatif.

### 3. METODE PENELITIAN

#### A. Algoritma C4.5

Pohon keputusan mirip sebuah struktur pohon dimana terdapat node internal (bukan daun) yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas. Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5 (Kusrini, 2009), yaitu:

#### 1. Menyiapkan data *training*.

Data *training* biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.

#### 2. Menentukan akar dari pohon.

Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung dahulu nilai *entropy*. Untuk menghitung nilai *entropy* digunakan rumus:

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

Keterangan:  
 S = himpunan kasus  
 n = jumlah partisi S  
 p<sub>i</sub> = proporsi S<sub>i</sub> terhadap S

- Kemudian hitung nilai *gain* menggunakan rumus:

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

Keterangan:  
 S = himpunan kasus  
 A = fitur  
 n = jumlah partisi atribut A  
 S<sub>i</sub> = proporsi S<sub>i</sub> terhadap S  
 S<sub>i</sub> = jumlah kasus dalam S

- Ulangi langkah ke-2 hingga semua *record* terpartisi.
- Proses partisi pohon keputusan akan berhenti saat :
  - Semua *record* dalam simpul N endap kelas yang sama.
  - Tidak ada atribut di dalam *record* yang dipartisi lagi.
  - Tidak ada *record* di dalam cabang yang kosong.

Menurut Larose (2005:116) Algoritma C4.5 merupakan pengembangan dari algoritma ID3. Algoritma C4.5 mengkonstruksi pohon keputusan dari sebuah set *data training*, yang berupa kasus-kasus atau *record-record (tupel)* dalam basis data, dimana tiap *record* memiliki atribut *kontinyu* diskrit atau keduanya.

**B. Klasifikasi Rule Based**

Aturan-aturan dalam *rule based* dapat diturunkan dari pohon keputusan yang telah terbentuk. Karena pohon keputusan yang besar, terkadang sulit untuk menginterpretasikan pohon bentuk keputusan (Han & Kamber, 2006:321).

**C. Evaluasi dan Validasi Metode Klasifikasi Data Mining**

Untuk menguji model, pada penelitian ini, digunakan metode *Cross Validation*, *Confusion Matrix*, dan kurva ROC (*Receiver Operating Characteristic*).

1. *Cross Validation*

*Cross validation* adalah pengujian standar yang dilakukan untuk memprediksi *error rate*. *Data training* dibagi secara random ke dalam

beberapa bagian dengan perbandingan yang sama kemudian *error rate* dihitung bagian demi bagian, selanjutnya hitung rata-rata seluruh *error rate* untuk mendapatkan *error rate* secara keseluruhan.

2. *Confusion matrix*

Metode ini menggunakan tabel matriks seperti pada tabel 1, jika data set hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negatif (Bramer, 2007:90).

Tabel 1. Model *Confusion Matrix*

Klasifikasi yang benar	Diklasifikasikan sebagai	
	+	-
+	<i>true positives</i>	<i>false negatives</i>
-	<i>false positives</i>	<i>true negatives</i>

Sumber: (Bramer,2007:90)

*True positives* adalah jumlah *record* positif yang diklasifikasikan sebagai positif, *false positives* adalah jumlah *record* negatif yang diklasifikasikan sebagai positif, *false negatives* adalah jumlah *record* positif yang diklasifikasikan sebagai negatif, *true negatives* adalah jumlah *record* negatif yang diklasifikasikan sebagai negatif, kemudian masukkan data uji. Setelah data uji dimasukkan ke dalam *confusion matrix*, hitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung jumlah *sensitivity (recall)*, *specificity*, *precision* dan *accuracy* (Han dan Kamber, 2006:361):

- Sensitivity* = t-pos/pos
- Specificity* = t-neg/neg
- Precision* = t-pos / (t-pos+f-pos)
- Accuracy*= *sensitivity*.pos / (pos+neg) + *Specificity* . neg / pos + neg
- Accuracy*=(TP+TN) / (TP+TN+FP+FN)

Keterangan:

TP = jumlah *true positives*  
 TN = jumlah *true negatives*  
 P = jumlah *record* positif  
 N = jumlah *tupel* negatif  
 FP = jumlah *false positives*

3. Kurva ROC

Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan *false positives* sebagai garis horisontal dan *true*

*positives* sebagai garis vertikal (Vecellis, 2009:233).

**4. PEMBAHASAN DAN HASIL**

**a. Jenis Penelitian**

Penelitian yang dilakukan adalah penelitian yang berbentuk Eksperimen. Dimana Penelitian Eksperimen ini menekankan pada percobaan dari teori-teori yang sudah ada.

**b. Ekperimen**

Metode penelitian yang digunakan dalam eksperimen ini menggunakan model *Cross-Standard Industry for Data Mining* (CRISP-DM) yang terdiri dari 6 fase, yaitu: (Sumathi, 2006:662-663)

1. Pemahaman tujuan proyek dari bisnis (*Business Understanding*)  
Berdasarkan laporan peminjaman tahun 2011 ditemukan kredit bermasalah sebanyak 20 anggota, sehingga ini menjadi permasalahan KSU “Ceger Jaya” yang merupakan imbas dari analisa analisis peminjaman dana yang kurang akurat.

Agar akurat, maka dikembangkan model klasifikasi algoritma C4.5 untuk menentukan kelayakan pemberian peminjaman dengan tujuan analisa yang dilakukan lebih akurat.

2. pemahaman data (*Data Understanding*)  
Untuk menentukan kelayakan Peminjaman, delapan atribut *predictor* dan satu atribut kelas. Pada tabel 2, ada beberapa atribut yang digunakan di dalam data anggota peminjam yaitu Pendapatan, Tanggungan, JenisUsaha, Plapon pinjaman, jangka waktu (lama pinjaman dalam tahun), Tingkat Bunga (Nisbah), Kepemilikan Rumah Tinggal, *Class* (label yang diberikan kepada nasabah, label Lancar untuk yang pembayarannya tidak bermasalah, dan bermasalah jika peminjaman macet). Tugas termasuk tabel dan catatan, sementara pemilihan atribut meliputi transformasi dan pembersihan data untuk alat pemodelan. Berikut ini kategori nilai atributnya :

Tabel 2. Katagori Nilai Atribut

No	Atribut	Nilai Angka
1.	Pendapatan	1. 2-3 Juta
		2. 3-4 Juta
		3. 5-9 Juta
2.	Tanggungan	a. Kosong
		b. Sedikit
		c. Sedang
		d. Banyak
3.	Jenis Usaha	a. Bengkel
		b. Toko
		c. Jasa
4.	Plafon Pinjaman	a. <5000000
		b. 5000000-10000000
5.	Jangka Waktu	a. 6 Bulan
		b. 12 Bulan
6.	Tingkat Bunga	a. 3%
		b. 4%
7.	Kepemilikan Rumah Tinggal	a. Rumah Sendiri
		b. Rumah Sewa
		c. Menumpang
8.	Status Pembayaran	a. Lancar
		b. Bermasalah

Sumber: Penulis

3. Tahap persiapan data (*data preparation*)  
Di dalam data *training* yang digunakan dalam penelitian ini, dilakukan seleksi atribut dan penghapusan data duplikasi menggunakan *software Rapid Miner*.

Setelah dilakukan *preprocessing* data yang didapat dari koperasi sebanyak 400 *record* direduksi dengan menghilangkan duplikasi menjadi 300 *record* untuk data *training* seperti pada tabel 3 yang merupakan *Sampel Data training*.

Tabel 3 Sampel Data *Training*

Pendapatan	Tanggungan	Jenis Usaha	Plapon pinjaman	Jangka Waktu Pinjaman	Tingkat Bunga	Kepemilikan Rumah Tinggal	Status Pembayaran(Class)
5-9 jt	Kosong	Bengkel	5000000-10000000	6 bln	3%	Rumah Sendiri	Bermasalah
2-3 jt	Banyak	Toko	<5000000	6 bln	3%	Rumah Sewa	Lancar
5-9 jt	Sedikit	Jasa	5000000-10000000	6 bln	4%	Rumah Sendiri	Lancar
2-3 jt	Banyak	Bengkel	<5000000	12 bln	4%	Menumpang	Lancar
5-9 jt	Sedang	Jasa	5000000-10000000	12 bln	4%	Rumah Sendiri	Lancar
5-9 jt	Kosong	Toko	5000000-10000000	6 bln	3%	Rumah Sewa	Bermasalah
5-9 jt	Sedikit	Bengkel	<5000000	12 bln	4%	Rumah Sendiri	Lancar
5-9 jt	Sedang	Jasa	5000000-10000000	6 bln	3%	Menumpang	Lancar
2-3 jt	Banyak	Toko	<5000000	6 bln	3%	Rumah Sendiri	Lancar
5-9 jt	Sedikit	Bengkel	5000000-10000000	12 bln	3%	Rumah Sewa	Lancar
5-9 jt	Banyak	Toko	<5000000	12 bln	3%	Rumah Sendiri	Lancar
5-9 jt	Sedang	Bengkel	5000000-10000000	6 bln	4%	Menumpang	Lancar
5-9 jt	Sedikit	Jasa	5000000-10000000	6 bln	4%	Rumah Sendiri	Lancar

Sumber: Penulis

a. Modelan yang dipilih (*Modeling*)

Tahap ini juga dapat disebut tahap *learning* karena pada tahap ini data training diklasifikasikan oleh model dan kemudian menghasilkan sejumlah aturan. Untuk membuat pohon keputusan, dibutuhkan data *training* yang diambil dari tabel 3. Pada tabel 3 adalah data *training* untuk menentukan nasabah kredit yang bermasalah atau tidak. Langkah-langkah untuk membuat algoritma C.45, yaitu :

- a) Siapkan data *training*
- b) Hitung nilai *entropy*.

Jika diketahui dalam data *training*, jumlah data 300, nasabah yang diterima kredit dalam status lancar 272 *record* dan yang status bermasalah 28 *record*. Setelah itu, hitung nilai *gain information* untuk setiap atribut selanjutnya tentukan simpul berikutnya, yaitu simpul 1.1, dan dilakukan perhitungan *gain information* masing-masing atribut dari nilai Pendapatan. Jumlah kasus yang dihitung adalah nilai dari simpul Pendapatan dan seterusnya sampai semua *record* dalam simpul, mendapatkan kelas yang sama.

b. Mengevaluasi mode (*Evaluation*)

Pada tahap ini dilakukan pengujian terhadap model untuk mendapatkan informasi model yang paling akurat. Evaluasi dan validasi menggunakan metode *cross validation*, *confusion matrix*, dan, kurva ROC.

Model dengan akurasi yang baik selanjutnya akan di test dengan data tunggal pada tahap penerapannya.

c. Penyebaran hasil (*Deployment*)

Setelah pembentukan model dan dilakukan analisa dan pengukuran pada tahap sebelumnya, selanjutnya pada tahap ini diterapkan model yang paling akurat ke Koperasi untuk penentuan kelayakan pemberian Pinjaman.

C. Hasil Penelitian

Tujuan dari penelitian ini menguji keakuratan analisa kredit dengan menggunakan algoritma C4.5. Data yang dianalisa adalah data pinjaman dalam bentuk kredit, yaitu semua data kredit yang telah disetujui oleh pihak koperasi.

D. Algoritma C4.5

Data *training* pada tabel 3 adalah untuk menentukan apakah seorang peminjam bermasalah atau tidak. Berikut akan dibahas prediksi apakah nasabah bermasalah atau tidak, menggunakan metode klasifikasi.

Langkah untuk membuat pohon keputusan, yaitu:

1. Tabel 3 adalah data *training* beserta kelasnya
2. Hitung nilai *entropy*. Dari data training diketahui jumlah kasus ada 300, peminjam yang termasuk kelas Lancar 272 *record* dan bermasalah 28 *record* sehingga didapat *entropy*:  

$$= (-272/300 \cdot \log_2 (272/300)) + (- 28/300 \cdot \log_2 (28/300)) = 0.4475$$
3. Hitung nilai *gain* untuk tiap atribut, lalu tentukan nilai *gain* tertinggi. Yang mempunyai nilai *gain* tertinggi itulah yang akan dijadikan akar dari pohon. Misalkan untuk atribut Pendapatan, didapat nilai *gain*:

$$Gain (S,A) = 0.4475 - (272/300*0.9986) + (28/300*0.4475) = 0.0786$$

Perhitungan *entropy* dan *gain* untuk semua atribut dilakukan, untuk mendapatkan nilai *gain* tertinggi. Hasil perhitungan seluruh atribut terlihat pada tabel 4.

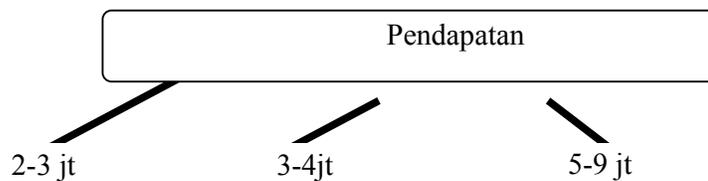
Tabel 4 Nilai *Entropy* dan *Gain* Untuk Menentukan Simpul Akar

Simpul	Kasus	Lancar	Bermasalah	Entropy	Gain
	300	272	28	0,4475	
<b>Root</b>					<b>0,0786</b>
<b>Pendapatan</b>					
2-3 jt	23	11	12	0,9986	
3-4 jt	110	102	8	0,3760	
5-9 jt	167	159	8	0,2774	
	300	272	28		
<b>Tanggungan</b>					<b>0,0740</b>
Kosong	77	66	11	0,5917	
Sedikit	82	76	6	0,3776	
Sedang	68	60	8	0,5226	
Banyak	73	70	3	0,2473	
	300	272	28		
<b>Jenis usaha</b>					<b>0,0016</b>
Toko	103	92	11	0,4902	
Jasa	94	87	7	0,3824	
Bengkel	103	93	10	0,4597	
	300	272	28		
<b>Plafon Pinjaman</b>					<b>0,0009</b>
<5000000	139	124	15	0,4936	
5000000-10000000	161	148	13	0,4048	
	300	277	23		
<b>Jangka Waktu Pinjaman</b>					<b>0,0011</b>
6 bln	243	219	24	0,4651	
12 bln	57	53	4	0,3666	
	300	272	28		
<b>Tingkat Bunga</b>					<b>0,0014</b>
3%	140	125	15	0,4912	
4%	160	147	13	0,4066	
	300	272	28		
<b>Kepemilikan rumah</b>					<b>0,0006</b>
Rumah Sendiri	150	137	13	0,4252	
Rumah Sewa	75	67	8	0,4898	
Menumpang	75	68	7	0,4475	
	300	272	28		

Sumber: Penulis (hasil pengolahan data)

Dari hasil perhitungan *entropy* dan *gain* yang didapat pada tabel 4, terlihat bahwa atribut Pendapatan mempunyai nilai *gain* tertinggi yaitu 0.0786, Oleh karena itu maka Pendapatan merupakan simpul akar pada pohon keputusan. Untuk menentukan simpul

berikutnya, yaitu simpul 1.1 dilakukan lagi perhitungan *entropy* dan *gain* berdasarkan atribut Pendapatan. Jumlah kasus yang dihitung adalah sejumlah kasus dengan nilai dari simpul akar (Pendapatan).. Sehingga bentuk *Tree* dari perhitungan tersebut seperti gambar berikut ini.



Gambar 1. *Tree* simpul akar  
Sumber : Penulis

Pada gambar 1, yaitu pohon keputusan yang dihasilkan dari perhitungan *entropy* dan *gain* untuk keseluruhan atribut. Berdasarkan hasil perhitungan *entropy* dan *gain* pada tabel 4, Pendapatan mempunyai nilai *gain* tertinggi dan menghasilkan tiga *path* sesuai dengan atribut

yang dimilikinya, yaitu : 2-3 juta, 3-4 juta , 5-9 juta. Langkah selanjutnya adalah memperhitungkan kembali *entropy* dan *gain* dari akar pohon yang sudah ada untuk menentukan simpul 1.1 untuk jalur 2-3 juta, lihat pada tabel 5.

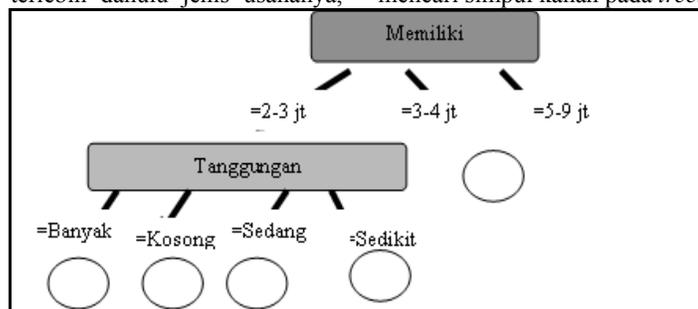
Tabel 5. Nilai *Entropy* dan *Gain* Untuk Menentukan Simpul 1.1

Simpul	Kasus	Lancar	Bermasalah	Entropy	Gain
	300	272	28	0,4475	
Pendapatan					0,0464
2-3 jt	19	11	8	0,9819	
3-4 jt	114	102	12	0,4855	
5-9 jt	167	159	8	0,2774	
	300	272	28		
<b>1.1 Tanggungan</b>					<b>0,3290</b>
Kosong	3	1	2	0,9183	
Sedikit	3	1	2	0,9183	
Sedang	7	3	4	0,9852	
Banyak	6	6	0	0,0000	
	19	11	8		
Jenis usaha					0,0132
Toko	6	4	2	0,9183	
Jasa	7	4	3	0,9852	
Bengkel	6	3	3	1,0000	
	19	11	8		
Plafon Pinjaman					0,1516
<5000000	12	9	3	0,8113	
5000000-10000000	7	2	5	0,8631	
	19	11	8		
Jangka Waktu Pinjaman					0,0894
6 bln	17	9	8	0,9975	
12 bln	2	2	0	0,0000	
	19	11	8		
Tingkat Bunga					0,0046
3%	11	6	5	0,9940	
4%	8	5	3	0,9544	
	19	11	8		
kepemilikan rumah					0,0670
Rumah Sendiri	9	4	5	0,9911	
Rumah Sewa	5	4	1	0,7219	
Menumpang	5	3	2	0,9710	
	19	11	8		

Sumber: Penulis (hasil pengolahan data)

Dari tabel 5, menghasilkan atribut Tanggungan sebagai simpul 1.1 memiliki nilai tertinggi 0.3290. Atribut tanggungan memiliki 4 *path* yaitu Banyak, Kosong, Sedang dan Sedikit. Pada *path* Banyak bisa dikatakan Lancar karena datanya menunjukkan sepenuhnya lancar, sedangkan path Kosong Bermasalah sedangkan untuk *path* sedang dapat dilihat terlebih dahulu jenis usahanya,

sedangkan *path* sedikit bermasalah. Setelah pohon cabang kiri terbentuk, maka langkah selanjutnya membentuk cabang kanan dari pohon. Dengan memperhitungkan *entropy* dan *gain* dari akar untuk simpul yang memiliki Pendapatan 3-4 Juta Lancar dan begitupun dengan pendapatan 5-9 juta datanya juga didapat lancar sehingga tidak diperlukan untuk mencari simpul kanan pada *tree*.



Gambar 2 *Tree* Cabang Kiri Simpul 1.1

Sumber : Penulis

Pada Atribut tanggungan yang perlu di cari *path* berikutnya adalah sedang dan sedangkan yang lain sudah dominan pada posisi yang benar. Langkah berikutnya

menghitung *gain ratio* atribut yang berada di path sedang untuk menemukan simpul 1.1 Lihat pada tabel 6 untuk mengetahui simpul tersebut.

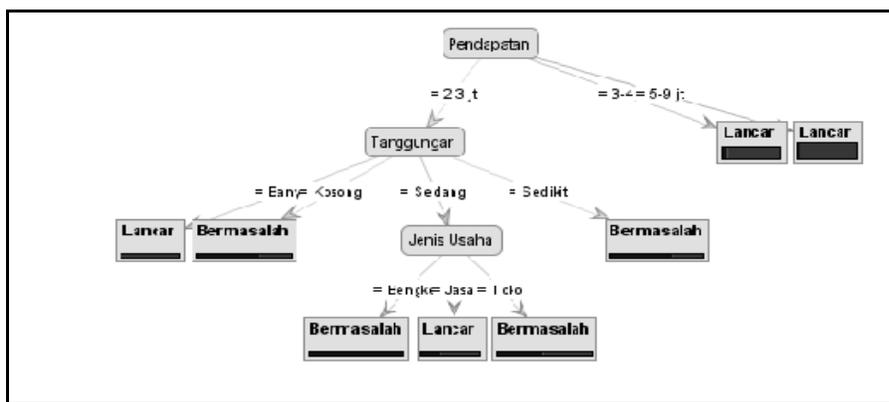
Tabel 6. Nilai *entropy* dan *gain* untuk menentukan simpul

Simpul	Kasus	Lancar	Bermasalah	Entropy	Gain
	300	272	28	0,4475	
<b>Pendapatan</b>					<b>0,0464</b>
2-3 jt	19	11	8	0,9819	
3-4 jt	114	102	12	0,4855	
5-9 jt	167	159	8	0,2774	
	<b>300</b>	<b>272</b>	<b>28</b>		
<b>1.1 Tanggungan</b>					<b>0,3290</b>
Kosong	3	1	2	0,9183	
Sedikit	3	1	2	0,9183	
Sedang	7	3	4	0,9852	
Banyak	6	6	0	0,0000	
	<b>19</b>	<b>11</b>	<b>8</b>		
<b>Jenis usaha</b>					<b>0,7317</b>
Toko	2	1	1	1,0000	
Jasa	3	2	1	0,9183	
Bengkel	2	0	2	0,0000	
	<b>7</b>	<b>3</b>	<b>4</b>		
<b>Plafon Pinjaman</b>					<b>0,6264</b>
<5000000	3	1	2	0,9183	
5000000-10000000	4	2	2	1,0000	
	<b>7</b>	<b>3</b>	<b>4</b>		
<b>Jangka Waktu Pinjaman</b>					<b>0,6920</b>
6 bln	6	2	4	0,9183	
12 bln	1	1	0	0,0000	
	<b>7</b>	<b>3</b>	<b>4</b>		
<b>Tingkat Bunga</b>					<b>0,6662</b>
3%	4	1	3	0,8113	
4%	3	2	1	0,9183	
	<b>7</b>	<b>3</b>	<b>4</b>		
<b>kepemilikan rumah</b>					<b>0,6190</b>
Rumah Sendiri	7	3	4	0,9852	
Rumah Sewa	0	0	0	0,0000	
Menumpang	0	0	0	0,0000	
	<b>7</b>	<b>3</b>	<b>4</b>		

Sumber : Penulis (hasil pengolahan data).

Sedangkan *decesion tree* hasil pengolahan data dengan menggunakan *Rapidminer* dapat dilihat pada gambar 3, dimana atribut pendapatan sebagai akar (*root*), atribut tanggungan dan

atribut jenis usaha,dimana masing-masing atribut tersebut menentukan anggota koperasi dalam hal pembayaran cicilan kreditnya bermasalah atau lancar.



Gambar 3. *Decesion tree* hasil pengolahan data dengan menggunakan *Rapidminer*  
 Sumber : Penulis (hasil pengolahan data)

Berdasarkan pohon keputusan pada gambar 3, dapat dibentuk *rule classification*, yaitu sebagai berikut:

a. R1: *IF* Memiliki pendapatan = 2-3 Juta

*AND* Tanggungan = Banyak *THEN* status pembayaran = Bermasalah

b. R2: *IF* Memiliki pendapatan = 2-3 Juta *AND* Tanggungan = sedang *THEN* status

- pembayaran = Lancar
- c. R3: *IF* Memiliki pendapatan = 2-3 Juta *AND* Tanggungan = Sedang *THEN* status pembayaran = Bermasalah
- d. R4: *IF* Memiliki pendapatan = 2-3 Juta *AND* Tanggungan = Sedikit *THEN* status pembayaran = Bermasalah
- e. R4: *IF* Memiliki pendapatan = 3-4 Juta *THEN* status pembayaran = Lancar
- f. R5: *IF* Memiliki pendapatan = 5-9 Juta *THEN* status pembayaran = Lancar

**E. Evaluasi dan Validasi**

**1. Evaluasi Pengujian**

Untuk membuat model klasifikasi, bisa digunakan banyak metode. Dalam penelitian ini hanya menggunakan metode klasifikasi dengan algoritma C4.5. Metode klasifikasi bisa dievaluasi berdasarkan beberapa kriteria seperti tingkat akurasi, kecepatan, kehandalan. Nilai

*accuracy*, *precision*, dan *recall* dari data training dapat dihitung dengan menggunakan *RapidMiner*. Hasil dari pengujian model yang telah dilakukan, dilakukan pengujian tingkat akurasi dengan menggunakan *confussion matrix* dan kurva ROC/AUC (*Area Under Cover*).

**a) Confussion Matrix**

Gambar 4. adalah perhitungan akurasi data training menggunakan algoritma C4.5. Diketahui dari 300 data *training* dan 7 atribut (pendapatan, tanggungan, jenis usaha, plafon pinjaman, jangka waktu pinjam, tingkat bunga, kepemilikan rumah tinggal), dengan menggunakan metode algoritma C4.5 didapat 270 data prediksi Lancar sesuai dengan Lancar, 2 prediksi Lancar ternyata Bermasalah, 2 data prediksi Bermasalah ternyata Lancar, dan 26 data prediksi Bermasalah sesuai dengan bermasalah.

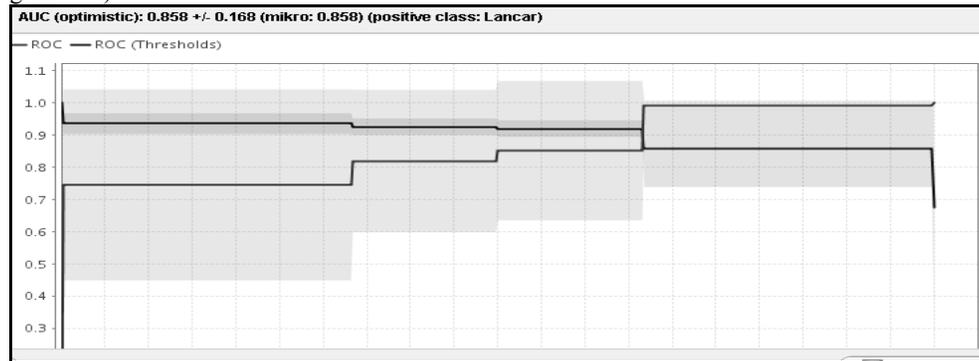
**Accuracy: 90.67 +/- 2.00% (mikro: 90.67%)**

	true Bermasalah	true Lancar	Class pred
Pred Bermasalah	2	2	50.00%
Pred Lancar	26	270	91.22%
Class recall	7.14%	99.26%	

Gambar 4. *Confussion Matrix(accuracy)* data training  
 Sumber : Penulis (hasil pengolahan data)

Setelah dilakukan pengujian data *training* maka didapatkan hasil pengukuran terhadap data *training* yaitu yaitu *accuracy* = 90.67% (lihat gambar 4)

**b) Kurva ROC (Receiver Operating Characteristic)**



Gambar 5. Menunjukkan grafik *ROC* dengan nilai *AUC* (*Area Under Curve*).  
 Sumber : Penulis (hasil pengolahan data)

Dengan metode *information gain* sebesar 0.858, metode *gain ratio* juga menunjukkan hasil yang sama. Akurasi *AUC* dikatakan sempurna apabila nilai *AUC* mencapai 1.000 dan akurasi buruk jika nilai *AUC* dibawah 0.500. Untuk klasifikasi *data mining*, nilai *AUC* dapat dibagi menjadi beberapa kelompok (Gorunescu, 2011:325-326).

- A. 0.90-1.00 = klasifikasi sangat baik
- B. 0.80-0.90 = klasifikasi baik

- C. 0.70-0.80 = klasifikasi cukup
- D. 0.60-0.70 = klasifikasi buruk
- E. 0.50-0.60 = klasifikasi salah

Berdasarkan pengelompokan klasifikasi data serta gambar 5, maka dapat disimpulkan bahwa metode C4.5 termasuk klasifikasi sangat baik karena memiliki nilai *AUC* antara 0.90-1.00 yaitu memiliki klasifikasi sangat baik.