PENERAPAN ALGORITMA KLASIFIKASI C4.5 UNTUK MENGHASILKAN POLA KELAYAKAN KREDIT

Abdussomad

Program Studi Sistem Informasi

Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri

ABSTRAK

Kredit merupakan suatu kepercayaan seseorang yang diberikan kepada seseorang atau badan lainnya dimana yang bersangkutan pada masa yang akan datang akan memenuhi segala sesuatu kewajiban yang telah disepakati sebelumnya. beberapa masalah yang sering terjadi pada lembaga pengkreditan misalkan tunggakan konsumen yang sebelumnya dianggap layak menerima kredit, macetnya status kredit. Munculnya masalah tersebut diakibatkan kurangnya pertimbangan atau kemantapan analisis kredit dalam menentukan kelayakan kredit pada saat konsumen mengajukan pengkreditan. Oleh karena itu perlu dilakukan analisis kredit sehingga dapat mengetahui kelayakan dari suatu permasalahan kredit, Melalui hasil analisis kreditnya, dapat diketahui apakah nasabah layak atau tidak. dari permasalahan yang ada digunakan metode klasifikasi untuk memprediksi kelayakan kredit yaitu dengan menggunakan model algoritma klasifikasi C4.5 Setalah dilakukan pengujian dengan model tersebut didapatkan hasil yaitu 90,99% dan nilai AUC sebesar 0,911 dengan tingkat diagnosa Excellent Classification. Dan dapat disimpulkan bahwa penerapan algoritma klasifikasi C4.5 mampu menghasilkan pola kelayakan kredit dengan tingkat akurasi dan diagnosa yang baik

Kata Kunci: Kelayakan kredit, algoritma klasifikasi, algoritma C4.5.

ABSTRACT

Credit is a belief that one is given to a person or other entity which is concerned in the future will fulfill all the obligations previously agreed. some of the problems that often occur in the crediting institutions eg consumer arrears were previously deemed worthy of receiving credit, the breakdown of credit status. The emergence of the problem as a lack of consideration or the stability of credit analysis in determining credit worthiness when consumers apply for crediting. Therefore it is necessary to do credit analysis so as to determine the feasibility of a credit crunch, through credit analysis results, it can be seen whether the customer is feasible or not. Some of the existing problems of classification method is used to predict credit is by using models classification algorithm C4.5 After testing with the model showed that 90.99% and AUC value of 0.911 to the level of diagnosis Excellent Classification. and it can be concluded that the application of the C4.5 classification algorithm capable of generating patterns of creditworthiness with pinpoint accuracy and a good diagnosis.

Keyword*: Credit analysis, classification algorithms, C4.5 algorithm.*

1. PENDAHULUAN

Surat Keputusan Bersama Menteri Keuangan, Perindustrian dan Perdagangan No.1169 /KMK.01/ 1991 tanggal 21 Nopember 1991 tentang kegiatan Sewa Guna Usaha, Leasing adalah setiap kegiatan pembiayaan perusahaan dalam bentuk penyediaan barang-barang modal untuk digunakan oleh suatu perusahaan untuk jangka waktu tertentu, berdasarkan pembayaran-pembayaran berkala disertai dengan hak pilih (opsi) bagi perusahaan tersebut.

Leasing memiliki permasalahan yang erat dengan hal kredit, beberapa masalah yang sering terjadi pada lembaga pengkreditan disebabkan ulah konsumen, misalkan tunggakan konsumen yang sebelumnya dianggap layak menerima kredit, macetnya status kredit yang dikarenakan faktor ekonomi konsumen dalam melakukan pembayaran

angsuran yang mengakibatkan ditariknya barang atau motor yang dikredit tersebut. Munculnya masalah tersebut diakibatkan kurangnya pertimbangan atau kemantapan analisis kredit dalam menentukan kelayakan kredit pada saat konsumen mengajukan pengkreditan.

Analisis kredit merupakan hal yang penting dalam lingkup resiko keuangan (Lai, Yu, Zhou, & Wang, 2006), Sebagai tolak ukur bahwa debitur disetujui atau ditolak, dapat digunakan data histori debitur yang telah disetujui oleh perusahaan. Namun, perlu diperhatikan juga bahwa debitur yang telah disetujui tidak semuanya termasuk kategori pembayar kredit yang baik, artinya ada beberapa debitur yang telah disetujui tapi beberapa bulan kemudian pembayarannya menunggak.Berdasarkan laporan sebuah *leasing* di daerah Karawang, menunjukkan tingkat nasabah bermasalah pada tahun 2015 pada angka 60% dibandingkan nasabah yang yang memiliki status lancar menunjukan pada angka 40%.

Terdapat beberapa penelitian dan teknik analisa kredit yang dilakukan oleh para peneliti seperti Yi Jiang (2009) Membuat model untuk memprediksi nasabah yang bermasalah dan tidak bermasalah dalam pembayaran kredit dengan menggunakan model Pohon Keputusan dan C4.5 dan Simulated Annealing Algoritma. Firmansyah (2011) menerapkan algoritma klasifikasi C4.5 untuk penentuan kelayakan pemberian kredit koperasi. Jozef Zurada (2010) melakukan penelitian untuk membandingkan beberapa algoritma seperti *Regresi Linier, Neural Network, Support Vector Machine, Case Base Reasoning, Rule Based Fuzzy Neural Network* dan *Decision Tree*. Semua model algoritma diatas digunakan untuk menganalisa persetujuan pinjaman dalam bentuk kredit. Dari hasil penelitian didapatkan bahwa *Decision Tree* terbukti mempunyai akurasi tertinggi dalam menentukan keputusan dibandingkan algoritma lain.

Untuk mengatasi permasalahan diatas, pada penelitian ini menggunakan model pohon keputusan algoritma C4.5 untuk membentuk model klasifikasi kelayakan kredit.

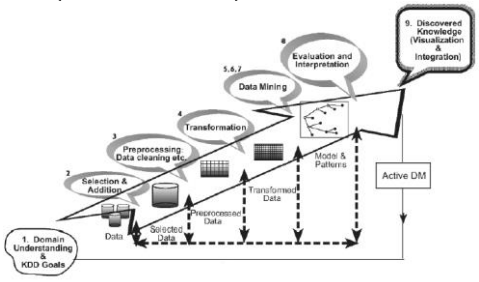
2. Landasan Teori

A. Data mining

Data mining adalah proses pengambilan pengetahuan dari volume data yang besar yang disimpan dalam basis data, data warehouse, atau informasi yang disimpan dalam repositori (Han & Kamber, 2012).

Data Mining (DM) adalah inti dari proses *Knowledge Discovery in Database* (KDD), yang melibatkan algoritma dalam mengeksplorasi data, mengembangkan model dan menemukan pola yang sebelumnya tidak diketahui (Maimon, 2010). Model ini digunakan untuk memahami fenomena dari data, analisis dan prediksi. KDD adalah proses terorganisir untuk mengidentifikasi pola yang valid, baru, berguna, dan dapat dimengerti dari sebuah *dataset* yang besar dan kompleks.

Knowledge Discovery in Database (KDD) memiliki langkah-langkah seperti terlihat pada gambar 1 (Maimon, 2010) :



Gambar 1: Proses KDD

proses data mining paling populer yaitu proses *Cross-Industry Standard Process for Data Mining* (CRISP-DM), berikut tahapan-tahapan dari proses CRISP-DM (Brown, 2014) :

1. *Business Understanding*
2. *Data Understanding*
3. *Data Preparation*
4. *Modeling*
5. *Evaluation*
6. *Deployment (using models in everyday business)*

Beberapa teknik dan sifat data mining (Hermawati, 2013) adalah sebagai berikut:

1. *Classification [Predictive]*
2. *Clustering [Descriptive]*
3. *Association Rule [Descriptive]*
4. *Sequential Pattern Discovery [Descriptive]*
5. *Regression [Predictive]*
6. *Deviation Detection [Predictive]*

B. Algoritma Klasifikasi Data Mining

Klasifikasi adalah menentukan sebuah record data baru ke salah satu dari beberapa kategori yang telah didefinisikan sebelumnya

atau yang disebut dengan *supervised learning* (Hermawati, 2009) Proses klasifikasi didasarkan pada empat komponen (Gorunescu, 2011) :

a. Kelas
   Variabel dependen yang berupa kategorikal yang merepresentasikan "label" yang terdapat pada objek. Contohnya: resiko kredit, *customer loyalty*, jenis gempa.
b. *Predictor*
   Variabel independen yang direpresentasikan oleh karakteristik (attribut) data. Contohnya: tabungan, aset, gaji.
c. *Training dataset*
   Satu *set* data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan *predictor*.
d. *Testing dataset*
   Berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi

Ada banyak algoritma klasifikasi yang dapat digunakan dalam data mining. Mulai dari *k-nearest neighbor, decision tree, neural network*, dan lainnya. Dalam penelitian terkait, disimpulkan bahwa algoritma pohon keputusan dengan C4.5 memiliki hasil yang baik untuk proses klasifikasi kelayakan pelanggan.

C. Algoritma Klasifikasi C4.5

Salah satu teknik klasifikasi yang paling populer digunakan dalam proses data mining adalah *classification* and *decision trees* (pohon keputusan). Pohon keputusan digunakan untuk memprediksi keanggotaan objek untuk kategori yang berbeda (kelas), dengan mempertimbangkan nilai-nilai yang sesuai dengan attribut mereka atau variabel prediktor (Gorunescu, 2011).

Algoritma C4.5 atau pohon keputusan mirip sebuah pohon dimana terdapat node internal (bukan daun) yang mendeskripsikan attribut-attribut, setiap cabang menggambarkan hasil dari attribut yang diuji, dan setiap daun menggambarkan kelas. Pohon keputusan dengan mudah dapat dikonversi ke aturan klasifikasi

Dalam proses pengujian attribut, cabang baru yang terbentuk akan diperhatikan dari tipe attribut (Han & Kamber, 2012). terdapat 3 jenis cabang yang mungkin muncul dalam pohon keputusan, yaitu :

1. Jika attribut bernilai diskrit, maka cabang yang terbentuk akan selalu sama dengan jumlah variasi nilai yang terdapat pada attribut tersebut.
2. Jika cabang bernilai kontinyu, maka akan dipecahkan menurut titik perpecahan, sedangkan titik perpecahan dikalkulasi dengan masing masing algoritma penyusun pohon keputusan. Cabang perpecahan yang terbentuk akan berpola seperti ≤ attribut, dan satu cabang lagi > attribut
3. Jika attribut yang diuji bernilai biner, maka cabang yang terbentuk pasti dua dan melibatkan nilai ya atau tidak.

Tahapan dalam membuat sebuah pohon keputusan dengan algoritma C4.5 (Gorunescu, 2011) yaitu:

1. Mempersiapkan *data training*, dapat diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokan dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon dengan menghitung nilai gain yang tertinggi dari masing-masing attribut atau berdasarkan nilai *index entropy* terendah. Sebelumnya dihitung terlebih dahulu nilai *index entropy*, dengan rumus:

$$Entropy(i) = -\sum_{j=1}^{m} f(i,j).\log 2\, f[(i,j)]$$

Keterangan:
i = himpunan kasus
m = jumlah partisi i
f(i,j) = proposi j terhadap i

3. Hitung nilai gain dengan rumus berikut :

$$Entropy(i) = -\sum_{i=1}^{p} \frac{n1}{n} . IE(i)$$

Keterangan:
p = jumlah partisi attribut
ni = proporsi ni terhadap i
n = jumlah kasus dalam n

4. Ulangi langkah ke-2 hingga semua *record* terpartisi.
   Adapun proses partisi pada pohon keputusan akan berhenti jika:

a. Semua tupel pada *record* dalam simpul m mendapat kelas yang sama.
b. Tidak ada attribut dalam *record* yang dipartisi lagi
c. Tidak ada *record* di dalam cabang yang kosong.

D. Pengujian *K-Fold Cross Validation*

Cross Validation adalah teknik validasi dengan membagi data secara acak kedalam k bagian dan masing-masing bagian akan dilakukan proses klasifikasi (Han & Kamber, 2012).

Dengan menggunakan *cross validation* akan dilakukan percobaan sebanyak k. Tiap percobaan akan menggunakan satu data *testing* dan k-1 bagian akan menjadi data *training*, kemudian data *testing* itu akan ditukar dengan satu buah data *training* sehingga untuk tiap percobaan akan didapatkan data *testing* yang berbeda-beda. *Data training* adalah data yang akan dipakai dalam melakukan pembelajaran sedangkan data testing adalah data yang belum pernah dipakai sebagai pembelajaran dan akan berfungsi sebagai data pengujian kebenaran atau keakurasian hasil pembelajaran (Witten & Frank, 2011).

Data yang digunakan dalam percobaan ini adalah data *training* untuk mencari nilai *error rate* secara keseluruhan. Secara umum pengujian nilai k dilakukan sebanyak 10 kali untuk memperkirakan akurasi estimasi. Dalam penelitian ini nilai k yang digunakan berjumlah 10 atau 10-*fold Cross Validation*.

E. *Confusion Matrix*

Confusion Matrix adalah alat (*tools*) visualisasi yang biasa digunakan pada *supervised learning*. Tiap kolom pada matriks adalah contoh kelas prediksi, sedangkan tiap baris mewakili kejadian di kelas yang sebenarnya (Gorunescu, 2010).

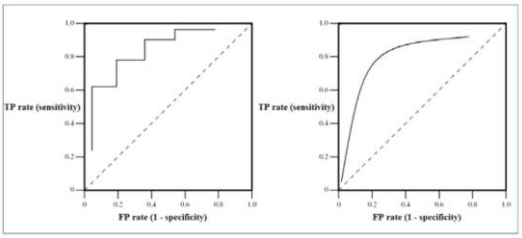
Confusion matrix berisi informasi aktual (*actual*) dan prediksi (*predicted*) pada sisitem klasifikasi.

F. *ROC Curve*

Dalam masalah klasifikasi, kurva ROC merupakan teknik untuk memvisualisasikan, mengatur dan memilih pengklasifikasi, berdasarkan kinerja mereka kurva ROC, banyak digunakan dalam menilai hasil prediksi (Gorunescu, 2011).

Kurva ROC (*Receiver Operating Characteristic*) adalah cara lain untuk mengevaluasi akurasi dari klasifikasi secara visual (Vercellis, 2009).

Kurva ROC menunjukan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan false positives sebagai garis horizontal dan true positives untuk mengukur perbedaaan performasi metode yang digunakan. ROC Curve adalah cara lain untuk menguji kinerja pengklasifikasian (Gorunescu, 2011).



Gambar 2: Grafik ROC (*discrete* dan *continous*)

Tingkat akurasi dapat di diagnosa sebagai berikut (Gournescu, 2011):
Akurasi 0.90 – 1.00 = *Excellent classification*
Akurasi 0.80 – 0.90 = *Good classification*
Akurasi 0.70 – 0.80 = *Fair classification*
Akurasi 0.60 – 0.70 = *Poor classification*
Akurasi 0.50 – 0.60 = *Failure*

G. Kerangka Pemikiran

Berdasarkan tinjauan pustaka dan tinjauan studi diatas, penulis membuat sebuah kerangka pemikiran yang berguna sebagai acuan penelitian ini, sehingga penelitian dapat dilakukan secara konsisten. Penelitian ini terdiri dari beberapa tahap yaitu: *Problem*, *Approach* (Algoritma Klasifikasi C4.5), *Development* (Rapidminer), *Implementation* (Eksperimen dengan Model CRISP-DM), *Measurement* (*Confusion Matrix, ROC Curve*), Result.

Permasalahan pada penelitian ini adalah belum adanya metode yang paling akurat untuk mengetahui prediksi kelayakan kredit (macet atau lancar). atas dasar permasalahan tersebut penulis menggunakan metode klasifikasi algoritma C4.5 untuk memecahkan masalah penelitian ini. Pengujian metode dilakukan dengan cara *confusion matrix* dan kurva ROC. Untuk mengembangkan aplikasi berdasarkan metode, digunakan tools RapidMiner 7.2.

3. METODE PENELITIAN

Metode penelitian yang digunakan pada eksperimen ini adalah model *Cross-Standard Industry for Data Mining* (CRISP-DM) yang terdiri dari 6 fase (Brown, 2014), yaitu :

A. Bussiness Understanding

Pada tahap business understanding bisa disebut juga tahap pemahaman penelitian, menentukan tujuan proyek penelitian dalam perumusan mendefinisikan masalah data maining. Berdasarkan kondisi nasabah kredit periode desember 2015, menunjukkan status kredit macet lebih tinggi tingkat persentasenya dengan kondisi nasabah yang berstatus lancar

Lebih tingginya data nasabah yang bermasalah menunjukkan kurangnya analisa yang akurat dalam menentukan kelayakan pemberian kredit bagi konsumen. teknik klasifikasi algoritma C4.5 banyak digunakan dalam analisa kredit dengan tujuan analisa yang lebih akurat.

B. Data Understanding

Pada tahap Data Understanding, dilakukan pengumpulan data, melakukan analisis penyelidikan data (data kredit) untuk mengenali lebih lanjut data dan pencarian pengetahuan awal kemudian mengevaluasi kualitas dari data tersebut.

C. Data Preparation

Data preparation merupakan kelanjutan dari data understanding, untuk mendapatkan data yang berkualitas, beberapa teknik *preprocessing* digunakan (Vecellis, 2009), yaitu:

1. *Data validation*, untuk mengidentifikasi dan menghapus data yang ganjil (*outlier/noise*), data yang tidak konsisten, dan data yang tidak lengkap (missing value)
2. *Data integration and transformation*, untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penulisan ini bernilai kategorikal dan kontinyu.
3. *Data size reduction and dicretization*, untuk memperoleh data set dengan jumlah atribut dan *record* yang lebih sedikit tetapi bersifat informatif. Di dalam data training yang digunakan dalam penelitian ini, dilakukan seleksi atribut dan penghapusan data duplikasi menggunakan software RapidMiner

D. Modelling

Pada tahap ini dilakukan pemrosesan data training sehingga akan menghasilkan beberapa aturan dan akan membentuk sebuah pohon keputusan. Model yang akan digunakan ada dua yaitu algoritma klasifikasi C4.5.

E. Evaluation

Pada tahap evaluasi, disebut tahap klasifikasi karena pada tahap ini akan ditentukan pengujian untuk akurasi. Tahap pengujiannya adalah melihat hasil akurasi pada proses klasifikasi Algoritma C4.5 serta evaluasi dengan ROC Curve.

F. Deployment

Pada tahapan deployment, dilakukan penerapan model algoritma klasifikasi C4.5 untuk menghasilkan pola kelayakan kredit.

4. HASIL DAN PEMBAHASAN

A. Pengumpulan data

Data yang didapat dari sebuah leasing di Kota Karawang pada periode Desember 2015, Jumlah data yang digunakan sebanyak 1044 record, dengan 8 attibut Prediktor yaitu Status, Pekerjaan, Penghasilan, Object, Dp_Net, Otr, Tenor, Angsuran, dan attribut Kondisi sebagai kelas atau label. Data diatas menunjukkan pada angka 620 data nasabah bermasalah (Macet) dan 424 data nasabah tidak bermasalah (Lancar).

Kemudian dilakukan teknik *preprocesing* menggunakan aplikasi Rapidminer sehingga menghasilkan *candidate split* seperti tabel 1 dibawah ini :

Tabel 1 *Candidate Split*

Cand. Split	***Child Nodes***	
1	**OTR** ≤ 12862500 ≤ 13387500 ≤ 20965000 ≤ 21115000	**OTR** > 12862500 > 13387500 > 20965000 > 21115000
2	**Angsuran** ≤ 1216500 ≤ 1376000 ≤ 2117000	**Angsuran** > 1216500 > 1376000 > 2117000
3	**Pekerjaan** Peg.swasta formal Wiraswasta formal Wiraswasta non formal Peg negeri Peg. Swasta non formal	

4	**Penghasilan (Juta)** 0-5 5-10 >10	
4	**Tenor** ≤ 20 ≤ 34.500	**Tenor** > 20 > 34.500
5	**Status** Penjamin Pemohon Tunggal	
6	**DP NET** ≤ 2250000 ≤ 2400000 ≤ 2725000 ≤ 2775000 ≤ 3800000 ≤ 3813100 ≤ 4113100	 > 2250000 > 2400000 > 2725000 > 2775000 > 3800000 > 3813100 > 4113100
7	**Object** Motor Bekas Motor Baru	

B. Hasil Pengujian Model Algoritma C4.5

Pada tahap ini dilakukan pemrosesan data training sehingga akan menghasilkan beberapa aturan dan akan membentuk sebuah pohon keputusan.

Berikut langkah-langkah model algoritma klasifikasi C4.5 yang dilakukan:

1) Menghitung jumlah kasus dengan kondisi Lancar dan kondisi Macet serta *Entropy* dari semua kasus

$$\text{Entropy(total)} = \left(-\frac{424}{1044} * log2\left(\frac{424}{1044}\right)\right) + \left(-\frac{620}{1044} * log2\left(\frac{620}{1044}\right)\right)$$

= 0,9744

2) Kemudian hitung nilai *Entropy* dan *gain* pada masing-masing atribut, sebagai contoh dibawah ini menghitung nilai *entropy* dan *gain* untuk atribut OTR :

OTR :

≤ 12.862.500 = $\frac{592}{1044}$ dan >12.862.500 = $\frac{452}{1044}$

Jumlah record OTR ≤ 12.862.500 terdiri dari 92 Lancar dan 500 Macet, sedangkan OTR > 12.862.500 terdiri dari 332 Lancar dan 120 Macet. Kemudian dapat dihitung entropynya sebagai berikut :

$$\text{Entropy}(i) = -\sum_{j=1}^{m} f(i,j).\text{Log2}\, f[(i,j)]$$

$$\text{OTR} \leq 12.862.500 = \left(-\frac{92}{592} * log2\left(\frac{92}{592}\right)\right) + \left(-\frac{500}{592} * log2\left(\frac{500}{592}\right)\right)$$

$$\text{OTR} > 12.862.500 = \left(-\frac{332}{452} * log2\left(\frac{332}{452}\right)\right) + \left(-\frac{120}{452} * log2\left(\frac{120}{452}\right)\right)$$

$$E\ split \leq 12.862.500(\text{T}) = \frac{592}{1044} * 0{,}6232 + \frac{452}{1044} * 0{,}8349$$

Gain OTR = 0,9744 - 0,7149 = 0.2595

Hasil perhitungan *entropy* dan *gain* selengkapnya dapat dilihat pada Tabel 4.3

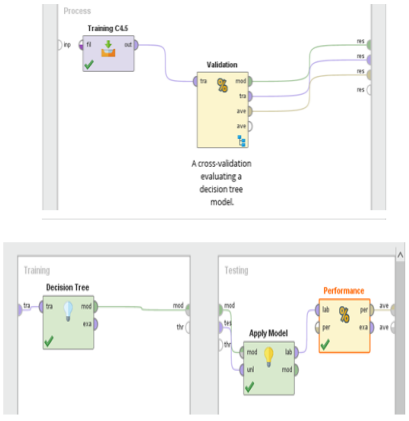
Tabel 2 *Information Gain* untuk Algoritma C4.5

SIMPUL	***ENTROPY***	***GAIN***
Jumlah Kasus	0,9744	
Candidate Split	***ENTROPY***	***GAIN***
OTR ≤ 12862500 > 12862500	 0,6232 0,8349	0,259560324
> 13387500 ≤ 13387500	0,8425 0,7162	0,208610141
≤ 20965000 > 20965000	0,9678 0,9996	0,00275032
> 21115000 ≤ 21115000	0,9863 0,9629	0,009039814
Angsuran > 1216500 ≤ 1216500	 0,8951 0,9784	0,002187399
> 1376000 ≤ 1376000	0,9457 0,9760	0,000368713
> 2117000 ≤ 2117000	0,8113 0,9722	0,004089345
Pekerjaan PEG,SWASTA FORMAL WIRASWASTA FORMAL WIRASWASTA NON FORMAL PEG NEGERI PEG. SWASTA NON FORMAL	 0,9772 0,9403 0,9703 0,0000 0,8610	0,015889231
penghasilan 0-5 5-10 10 - ...	 0,9710 0,9703 0,9403	0,0045211124
Tenor > 20 ≤ 20	 0,9919 0,9073	0,010395086
> 34,500 ≤ 34,500	0,5933 0,9985	0,062442456

Status Penjamin Pemohon Tunggal	 0,8405 0,9673	0,082424478
Dp Net ≤ 2250000 > 2250000	 0,8869 0,9816	0,004432489
≤ 2400000 > 2400000	0,7518 0,9916	0,024391324
≤ 2725000 > 2725000	0,8352 0,9978	0,028792892
≤ 2775000 > 2775000	0,8229 0,9995	0,037323171
≤ 3800000 > 3800000	0,9675 0,9879	0,001020909
≤ 3813100 > 3813100	0,9675 0,9879	0,001020909
≤ 4113100 > 4113100	0,9598 0,9989	0,005163587
Object Motor Bekas Motor Baru	 0,6448 0,8448	 0,243809265 7

Data diatas menunjukkan attribut OTR dengan split ≤ 12862500 dan > 12862500 memiliki nilai gain tertinggi yaitu 0,259560324, sehingga attribut OTR akan menjadi akar utama dari model tersebut. Lakukan perhitungan *entropy* dan *gain* sampai pembentukan akar terakhir.

Hasil pengujian dengan *K-Fold Cross Validation* Algoritma C4.5



Gambar 3: Pengujian *K-Fold Cross Validation* Algoritma C4.5

Perhitungan nilai akurasi dilakukan dengan menggunakan aplikasi rapidminer. Adapun hasil tes menggunakan algoritma C4.5 ditunjukkan pada Tabel 3.

Hasil dari pengujian model yang telah dilakukan adalah untuk mengukur tingkat akurasi dan AUC (Area Under Curve).

1) Confusion Matrix

Jumlah *True Positive* (TP) adalah 332 *record* diklasifikasikan sebagai LANCAR dan *False Negative* (FN) sebanyak 2 record diklasifikasikan sebagai LANCAR tetapi MACET. Berikutnya 618 *record* untuk *True Negative* (TN) diklasifikasikan sebagai MACET, dan 92 *record False Positive* (FP) diklasifikasin sebagai MACET ternyata LANCAR.

Tabel 3 Konversi *confusion matrix* algoritma klasifikasi C4.5

***accuracy*: 90.99% +/- 2.54% (mikro: 91.00%)**

	true LANCAR	*true* MACET	*class precision*
pred. LANCAR	332	2	99.40%
pred. MACET	92	618	87.04%
class recall	78.30%	99.68%	

Berdasarkan Tabel 3 tersebut menunjukan bahwa, tingkat akurasi dengan menggunakan algoritma C4.5 adalah sebesar 90,99%, dan dapat dihitung untuk mencari nilai *accuracy*, *sensitivity*, *specificity*, ppv, dan npv pada persamaan dibawah ini:

$$acc = \frac{tp+tn}{tp+tn+fp+fn} \qquad acc = \frac{332+618}{332+618+92+2}$$

$$sensitivity = \frac{tp}{tp+fn} \qquad sensitivity = \frac{332}{332+2}$$

$$specitivity = \frac{tn}{tn+fp} \qquad specitivity = \frac{618}{618+92}$$

$$ppv = \frac{tp}{tp+fp} \qquad ppv = \frac{332}{332+92}$$

$$npv = \frac{tn}{tn+fn} \qquad npv = \frac{618}{618+2}$$

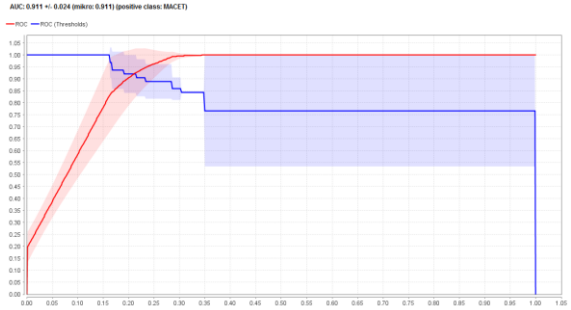
Kesimpulan Hasil perhitungan persamaan di atas ditunjukkan pada Tabel 4 di bawah ini:

Tabel 4 Hasil perhitungan algoritma C4.5

	Nilai (%)
Accuracy	90,99
Sensitivity	99,40
Specitivity	87,04
PPV	78,30
NPV	99,68

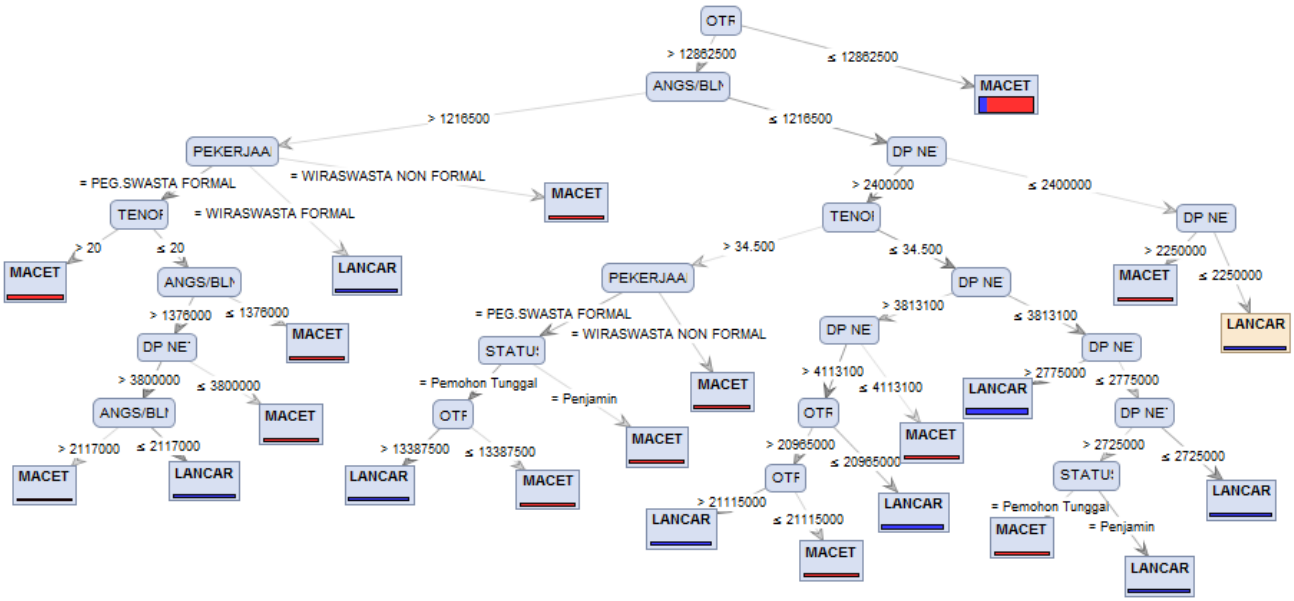
2) ROC Curve

Pada gambar 4 menunjukkan grafik ROC dengan nilai AUC (*Area Under Curve*) sebesar 0,911 dengan tingkat diagnosa *Excellent Classification*



Gambar 4: Nilai AUC dalam Grafik ROC algoritma C4.5

Hasil pohon keputusan algoritma klasifikasi C4.5 dengan pengujian K-fold cross Validation adalah sebagai berikut:



Gambar 7: Pohon Keputusan Klasifikasi Kedit dengan Algoritma C4.5

C. Implikasi Penelitian

Dari hasil evaluasi yang dilakukan diatas, baik secara *confusion matrix* maupun *ROC curve* menunjukkan bahwa algoritma klasifikasi C4.5 memiliki nilai akurasi yang tinggi yaitu sebesar 90,99% dan memiliki nilai uji AUC yang tinggi yaitu 0,911 (*Excellent Classification*).

5. KESIMPULAN

Dari hasil penerapan algoritma klasifikasi C4.5 diatas dapat disimpulkan bahwa nilai akurasi yang dihasilkan oleh model Algoritma C4.5 sebesar 90,99%. Sedangkan untuk evaluasi menggunakan ROC curve untuk model algoritma C4.5 menghasilkan nilai AUC sebesar 0,911 dengan tingkat diagnosa *Excellent Classification*. Sehingga dapat disimpulkan bahwa penerapan algoritma klasifikasi C4.5 mampu menghasilkan pola kelayakan kredit dengan tingkat akurasi dan diagnosa yang baik

DAFTAR PUSTAKA

[1] Brown, Meta S. (2014). Data Mining for Dummies a Wiley Brand. Hoboken: John Wiley & Sons, Inc.

[2] Firmansyah. (2011). Penerapan Algoritma Klasifikasi C4.5 untuk Penentuan Kelayakan Pemberian Kredit Koperasi. Tesis STMIK Nusa Mandiri.

[3] Gorunescu, Florin. (2011). Data Mining : Concepts, Models and Techniques. Chennai: Springer.

[4] Han, J., Kamber, M & Jian Pei. (2012). Data Mining: Concepts and Techniques (Third Edition ed.). San Francisco: Elsevier Inc.

[5] Hermawati, Fajar Astuti. (2013). Data Mining. Yogyakarta:ANDI.

[6] Jiang, Yi. (2009). Credit Scoring Model Based on the Decision Tree and the Simulated Annealing Algorithm. 978-0-7695-3507-4/08 $25.00 © 2008 IEEE. DOI 10.1109/CSIE.2009.481

[7] Lai, Kin Keung, Lean Yu, Ligang Zhou and Shouyang Wang. (2006). Credit Risk Evaluation with Least Square Support Vector Machine. G. Wang et al. (Eds.): RSKT 2006, LNAI 4062, pp. 490–495, 2006. Springer-Verlag Berlin Heidelberg 2006.

[8] Maimon, Oded & Lior Rokach. (2010). Data Mining and Knowledge Discovery Handbook. New York: Springer.

[9] Vercellis, C. (2009). Business Intelligence Data Mining And Optimization For Decision Making . United Kingdom: A John Wiley And Sons, Ltd., Publication.

[10] Witten, H. I., Frank, E., & Hall, M. A. (2011). Data Mining Pratical Mechine Learning Tools And Technique. Burlington: Elsevier Inc.

[11] Zurada, Josef. (2010). Could Decision Trees Improve the Classification Accuracy and Interpretability of Loan Granting Decisions?. Proceedings of the 43rd Hawaii International Conference on System Sciences -2010. 978-0-7695-3869-3/10 $26.00 © 2010 IEEE.