

PREDIKSI KREDIT MACET MELALUI PERILAKU NASABAH PADA KOPERASI SIMPAN PINJAM DENGAN MENGGUNAKAN METODE ALGORITMA KLASIFIKASI C4.5

Adi Sucipto
Fakultas Sains dan Teknologi UNISNU Jepara
adismusa@yahoo.com

ABSTRACT

Bad loans at credit unions has become very worrying for the health cooperative itself, where the credit banyakanya jammed from time to time and is difficult to charged, while the simplicity of the cooperative administration has been unable to protect the cooperative from the failure to collect debts to customers or sell existing collateral. Prevention and limitation of bad loans in the cooperative have not found a way most suitable for the cooperative does not have such a reliable credit analyst in banking and high cost, as long as this credit analysis is done with a personal approach with filling fiber blank credit application field survey. Classification of mining data with C4.5 algorithm model was conducted by measurable tests using AUC test, ROC and T-Test by rapid miner. The result is after testing by using parameters of customer personal data together with the number of customers is 1312, in fact, gives an accuracy of all the validation value is 91.06%, precision in is 100.00% and recall is 78.00%. It means the testing accuracy by algorithm of C4.5 is still good and It was able to use as one of the guidelines for the detection of bad debts before any decision of the member or not prospective new customers. Test results using C4.5 algorithm is not solely can be used for detection of bad credit, there are many other algorithms that can be compared with C4.5 algorithm, thus the results might be different.

Keywords: *curve AUC, ROC, T-Test*

ABSTRAK

Kredit macet di komperasi simpan pinjam sudah menjadi hal yang sangat mengkhawatirkan bagi kesehatan koperasi itu sendiri, dimana semakin banyakanya kredit yang yang macet dari waktu ke waktu dan sulit untuk ditagihkan, sementara kesederhanaan administrasi koperasi belum mampu melindungi koperasi dari kegagalan menagih hutang kenasabah atau menjual jaminan yang ada. Penanggulangan dan pembatasan kredit macet di koperasi belum menemukan cara yang paling sesuai karena koperasi belum mempunyai analis kredit yang handal seperti di perbankan dan mahlnya biaya, selama ini analisa kredit dilakukan dengan melakukan pendekatan personal dengan mengisi blangko pengajuan kredit serat survey lapangan. Algoritma klasifikasi data mining dengan model algoritma C4.5 dilakukan dengan pengujian-pengujian yang terukur melalui uji *AUC, ROC dan T-Test* dengan bantuan *rapid miner*. Hasilnya, setelah dilakukan pengujian dengan menggunakan parameter biodata nasabah dengan jumlah nasabah sebanyak 1312 ternyata menghasilkan akurasi sebesar secara keseluruhan nilai hasil validasi adalah accuracy = 91,06%, precision = 100,00% dan recall = 78,00%, artinya akurasi pengujian dengan menggunakan algoritma C4.5 masih baik dan dapat dijadikan salah satu pedoman untuk deteksi kredit macet sebelum ada keputusan member atau tidak calon nasabah baru. Hasil pengujian dengan menggunakan algoritma C4.5 bukan satu-satu algoritma yang dapat digunakan untuk deteksi kredit macet, masih banyak algoritma lain yang dapat diperbandingkan dengan algoritma C4.5 dimana kemungkinan hasilnya akan berbeda.

Kata Kunci: *curve AUC, ROC, T-Test*

Pendahuluan

Lembaga finansial merupakan suatu lembaga yang bergerak dalam bidang keuangan dimana para nasabahnya merupakan masyarakat dari berbagai lapisan dengan berbagai perilaku. Masyarakat indonesia dalam melakukan transaksi peminjaman dana

dari bank masih banyak menggunakan nanalisa kebutuhan secara pribadi walaupun dalam pengajuaannya menggunakan analisa perusahaan, hal ini dikarenakan tuntutan yang harus dipenuhi karena persyaratan perbankan (*bankable*). Masih banyak lembaga finansial baik itu perbankan maupun koperasi simpan

pinjam yang masih mempunyai jumlah kredit
macet yang cukup besar,

hal ini akan sangat mengganggu kesehatan koperasi maupun perbankan.

Peta Perkreditan Indonesia di bawah bayang-bayang macet. Salah satu isu sentral yang mempengaruhi dunia perbankan saat ini adalah masalah kredit macet. Perbincangan belakangan berkembang tidak hanya membahas sebab-sebab terjadinya kredit macet. Yang lebih menarik adalah salah satu upaya penyelesaian kredit macet lewat penghapusbukuan (*write off*) kredit macet agar kinerja bank terlihat sehat. Polemic penghapusan semakin santer setelah Gubernur Bank Indonesia Soedrajat Djiwandono, membeberkan telah dihapus bukukannya kredit macet di bank-bank pelat merah senilai Rp.4.2 triliun, dimana Rp1 triliun diantaranya dilakukan oleh BRI. (Mudrajat Kuncoro, Suhardjono, 2012).

Pada koperasi simpan pinjam masih banyak para manajer (pengelola) koperasi mengambil keputusan untuk memberikan kredit pada nasabah menggunakan pendekatan kedekatan nasabah dengan pengambil keputusan (Pengurus dan atau Manajer), walaupun masih tetap mempertimbangkan persyaratan-persyaratan yang harus dipenuhi. Perencanaan koperasi menjadi hal yang sangat penting dalam mengelola koperasi agar dapat berjalan sesuai dengan harapan semua pihak terutama para anggota. Dalam kaitannya dengan perencanaan, maka misi bank harus dijabarkan dalam tujuan perencanaan yang terinci. Walaupun demikian, penyusunan perencanaan kredit secara umum harus memenuhi criteria dapat dilaksanakan (*feasible*), dapat memberikan arah yang kita kehendaki (*suitable*), dapat diterima (*acceptable*), mempunyai nilai yang berarti (*valuable*), mudah dicapai (*achievable*), dan hasilnya dapat dievaluasi/diukur (*measurable*), (Taswan, 2010).

Perilaku nasabah koperasi maupun perbankan sangat menentukan seberapa keberhasilannya koperasi maupun perbankan dalam menyalurkan kredit, pelaku merupakan penentu apakah nasabah akan melakukan kewajibannya sebagai nasabah kredit atau sebaliknya, ketika mereka sudah mendapatkan kredit sengaja tidak mau membayar. Memahami kepribadian tidaklah lengkap jika tidak memahami konsep gaya hidup. Gaya hidup adalah konsep yang lebih baru dan lebih mudah terukur dibandingkan dengan kepribadian. (Ujang Sumarwan, 2011). Pendekatan perilaku ini sangatlah penting dilakukan oleh para pemangku

kebijakan baik di koperasi maupun di perbankan. Apabila mendekati masalah dari perspektif pengaruh perilaku (*behavior influence perspective*), para peneliti mamfokuskan pada perilaku konsumen dan kemungkinan lengkungan yang mempengaruhi perilaku-perilaku tersebut, (John C. Mowen/Michael Minor, 2002).

Permohonan kredit pada koperasi simpan pinjam biasanya dilakukan secara sederhana tidak seperti di bank, namun masih tetap memenuhi persyaratan prinsip dari sebuah kredit. Penilaian yang terkait dengan jumlah pinjaman, waktu pengembalian, jenis usaha/pekerjaan, jumlah tanggungan, status perkawinan, jumlah anggota keluarga, agunan dan ketentuan-ketentuan lain yang mengikat diantaranya pembayaran pajak, pembayaran rekening listrik, telepon dll. Permohonan ini biasanya diisi oleh calon nasabah dalam blanko yang telah tersedia untuk berkas pengajuan kredit, selebihnya ditentukan oleh hasil analisis team yang telah ditentukan koperasi simpan pinjam.

Analisa yang dilakukan oleh para manajer koperasi simpan pinjam dilakukan secara sederhana karena keterbatasan komperasi dalam meng-huyer tenaga analis kredit yang cukup mahal dan memberatkan nasabah, karena kesederhanaan nalisis inilah yang memicu terjadinya kredi macet. Antisipasi yang biasa dilakukan hanyalah dengan melakukan pendekatan-pendekatan personal pada semua nasabah kredit, ini memang tidak efektif manakala jumlah nasabah kredit menunggak sangat banyak. KSP memang belum cukup dana untuk mendatangkan analis kredit seperti layaknya Bank mengingat kondisi keuangan yang belum memungkinkan, sehingga terjadi tunggakan kredit yang cukup besar diatas ambang batas NPL (*Net Present Loan*) yang ditetapkan Bank Indonesia sebesar 5% [9]. Berikut contoh kredit macet yang saya ambil dari koperasi simpan pinjam Artha HPKJ Jepara mulai tahun 2009-2014, yaitu:

Tahun	Jumlah Anggota	Kredit Macet
2009	513	9,33%
2010	762	11,61%
2011	931	12,46%
2012	1.022	12,57%
2013	1.246	11,2%
2014	1.312	10,7%

Tabel 1 Data Kredit Macet Tahun 2009-2014
(Sumber KSP Artha Abadi Jepara)

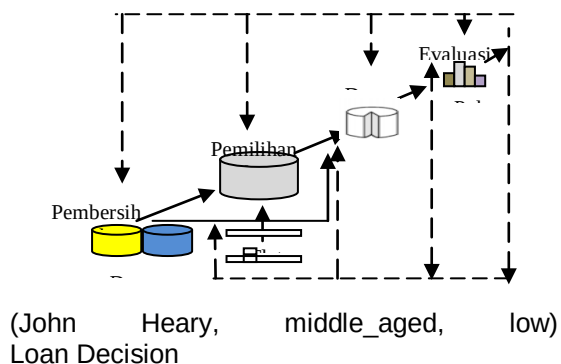
Tabel 1. menunjukkan kenaikan kredit macet pada koperasi simpan pinjam Artha Abadi Jepara, dimana setiap tahunnya selalu ada peningkatan kredit yang macet. Semakin banyak anggota kredit setiap tahunnya semakin bertambah pula prosentase kredit macetnya. Tujuan dari model penilaian kredit untuk menetapkan pemohon kredit "Baik" kelompok yang mungkin membayar kewajiban keuangan sepenuhnya atau "Macet" kelompok yang memiliki kemungkinan default tinggi pada kewajiban membayar [1]. Oleh karena itu, penilaian kredit bermasalah dalam lingkup masalah klasifikasi yang lebih umum dan banyak dibahas.

Tinjauan Pustaka

Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap yang diilustrasikan dalam Gambar 2.1. Tahap tersebut bersifat interaktif dimana pemakai terlibat langsung atau dengan perantara *knowledge base*. (Xindong Wu and Vipin Kumar, 2009)

1. Pembersihan Data.

Digunakan untuk membuang data yang tidak konsisten dan noise.



Gambar 2. 1 Langkah-langkah Data Mining
(Xindong Wu and Vipin Kumar, 2009)

1. Integrasi Data

Data yang diperlukan untuk data mining tidak hanya berasal dari suatu data base tetapi juga berasal dari beberapa database atau file teks. Hasil integrasi sering diwujudkan dalam sebuah data warehouse karena dengan data warehouse, data dikonsolidasikan dengan struktur khusus yang efisien. Selain itu data warehouse juga memungkinkan tipe analisa seperti OLAP.

2. Transformasi Data

Transformasi dan pemilihan data ini untuk menentukan kualitas dari data mining, sehingga data dirubah menjadi bentuk sesuai untuk di-Mining.

3. Aplikasi Teknik Data Mining

Aplikasi teknik data mining sendiri hanya merupakan salah satu bagian dari proses data mining. Ada beberapa teknik data mining yang sudah umum dipakai.

4. Interpretasi/Evaluasi pola yang ditemukan

Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang memang tercapai.

5. Pengetahuan

Presentasi pola yang ditemukan untuk menghasilkan aksi tahap terakhir dari proses data mining dan bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat.

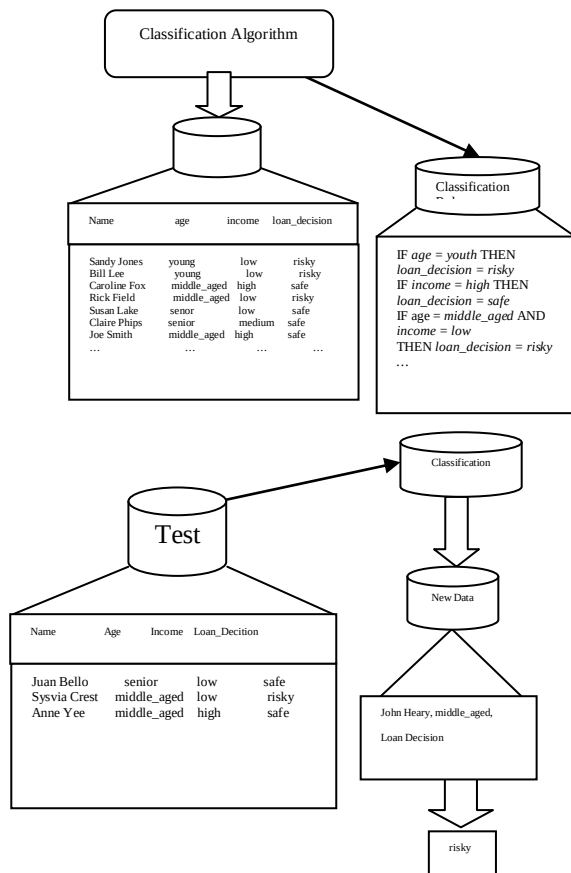
gambar 2. 2 Proses Klasifikasi(a) data dan Prediksi (b), (Budi Santosa,2007)

Algoritma Klasifikasi C4.5

Algoritma C4.5 adalah algoritma populer dalam klasifikasi data mining. Algoritma C4.5 merupakan pengembangan dari ID3, dalam ID3 kita gunakan criteria information gain untuk memilih atribut yang akan digunakan untuk pemisahan obyek [12]. Atribut yang mempunyai informasi gain paling tinggi dibandingkan dengan atribut yang data lain relatif terhadap set y dalam satu data, dipilih untuk melakukan pemecahan.

Pohon keputusan (*decision Tree*), dalam *decision tree* kita tidak menggunakan vector jarak untuk mengklasifikasikan obyek. Seringkali kita mempunyai data observasi dengan atribut-atribut yang bernilai nominal. *Decision Tree* sesuai digunakan untuk kasus-kasus dimana outptnya bernilai diskrit. Walaupun banyak variasi model *decision tree* dengan tingkat kemampuan dan syarat yang berbeda, pada umumnya beberapa ciri kasus berikut cocok untuk diterapkan pada *decision tree*:

1. *Data/example dinyatakan dengan pasangan atribut dan nilainya*. Misalnya atribut satu example adalah temperature dan nilainya adalah dingin.



Gambar 2. 1 Proses Klasifikasi(a) Data dan Prediksi (b) [10]

2. *Label/output data biasanya bernilai diskrit.* Output ini bias bernilai “ya” atau “tidak”, dalam beberapa kasus mungkin saja outputnya tidak hanya dua klas, tetapi penerapn decision tree lebih banyak untuk kasus binary.
3. *Data mempunyai missing value.* Misalkan untuk beberapa example, nialai dari satu atributnya tidak diketahui [12]. Dalam keadaan seperti ini decision tree masih mampu member solusi yang baik.
4. Tabel 8 menyajikan data set “*beli komputer*” yang dipraktekkan dengan algoritma C4.5 dengan tujuan untuk memprediksi apakah siswa mengambil keputusan membeli komputer. Data set tersebut dapat diilustrasikan dalam pohon keputusan, maka dapat digambarkan sebagai berikut:
5. Gambar 2.3 mengilustrasikan pohon keputusan yang disebabkan oleh algoritma C4.5 dengan menggunakan data dari tabel 8 sebagai data pelatihan.

6. Perhatikan bagaimana atribut pertama yang dipilih untuk uji keputusan atribut *Aged*. Untuk melihat mengapa, mari kita memperkirakan entropi dari variabel kelas secara acak (membeli komputer). Variabel ini membutuhkan dua nilai dengan probabilitas 9/14 untuk “Ya” dan 5/14 untuk “Tidak”.

Entropi dari suatu variabel kelas acak yang mengambil nilai-nilai c dengan probabilitas $p_1, p_2 \dots p_c$ diberikan oleh:

(2.1)

Dimana:

D : kumpulan masalah

c : jumlah partisi

p_i : proporsi dari D_1 terhadap D

dari contoh tabel (3) input data set, maka *entropy* membeli komputer dapat dihitung sebagai berikut:

$$-(9/14) \log_2 (5/14) \log_2 (5/14) = \mathbf{0,940}$$

Ini berarti bahwa pada rata-rata 0,940 bit harus ditransmisikan untuk informasi komunikasi membeli komputer? variabel acak. Tujuan dari induksi C4.5 untuk mengajukan pertanyaan yang tepat sehingga entropi ini adalah diputuskan.

Tabel 2 Contoh Table Input Training Data Set untuk C4.5 [12]

(sumber: All Electronic customer database

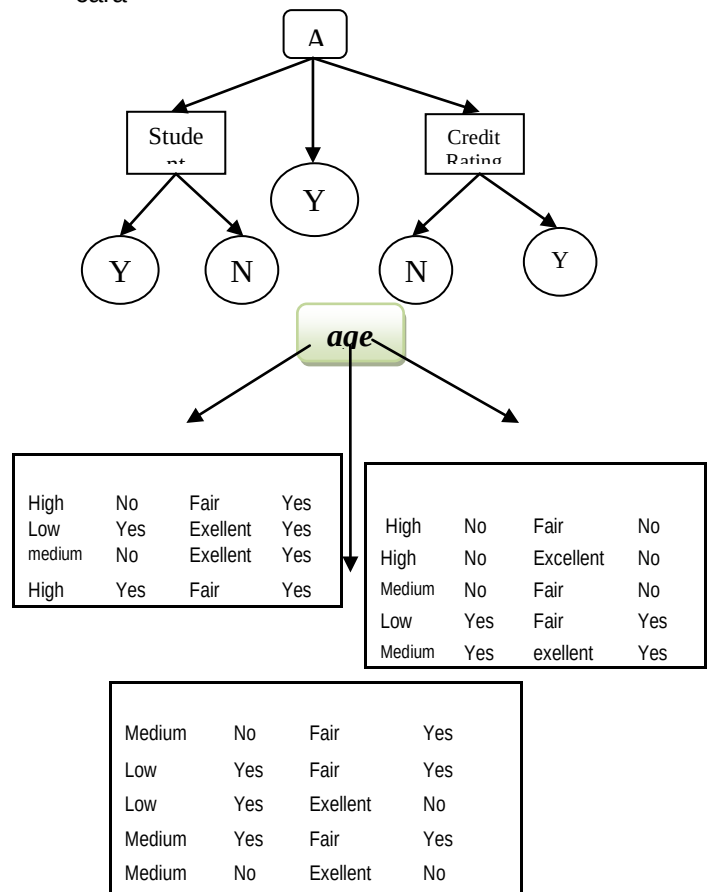
<i>R/I</i> <i>D</i>	<i>Age</i>	<i>Income</i>	<i>Student</i>	<i>Credit_rating</i>	<i>Class: buys_comput</i>
1	Youth	High	No	Fair	No
2	Youth	High	No	Excellent	No
3	Middle_aged	High	No	Fair	Yes
4	Senior	Medium	No	Fair	Yes
5	Senior	Low	Yes	Fair	Yes
6	Senior	Low	Yes	Excellent	No
7	Middle_aged	Low	Yes	Excellent	Yes
8	Youth	Medium	No	Fair	No
9	Youth	Low	Yes	Fair	Yes
10	Senior	Medium	Yes	Fair	Yes
11	Youth	Medium	Yes	Excellent	Yes
12	Middle_aged	Medium	No	Excellent	Yes
13	Middle_aged	High	Yes	Fair	Yes
14	Senior	Medium	No	Excellent	No

7. Pertimbangkan giliran masing-masing atribut untuk menilai peningkatan *entropi* untuk variabel acak yang diberikan, *aged* mengatakan, peningkatan *entropi* direpresentasikan sebagai *Gain* (*aged*), dapat dihitung sebagai berikut:

v : himpunan nilai yang mungkin (dalam kasus ini, tiga nilai aged)
D : keseluruhan data set
Dv : sub set dari data set yang memiliki nilai atribut aged
| · | : ukuran data set (dalam jumlah kasus)
Perhitungan ini akan menunjukkan bahwa:

Perhatikan bahwa tupel jatuh ke partisi untuk *aged=middle_aged* dengan memiliki kelas yang sama. Karena mereka mempunyai kelas

Tapi bagaimana kita bisa menghitung informasi *gain* dari suatu atribut yang terus dihargai tidak seperti diatas? "Misalkan, sebaliknya, bahwa kita memiliki atribut A yang terus dihargai, bukannya diskrit dihargai. (sebagai contoh, anggaplah bahwa bukan versi *discretized-valued* diatas, kita malah memiliki nilai baku untuk atribut ini). Untuk cara



Gambar 2.3 *aged*

seperti itu kita harus menentukan yang “terbaik” split titik A, di mana titik split ambang pada A. Kami pertama semacam nilai-nilai A dalam rangka meningkat. Biasanya, titik tengah antara setiap pasangan nilai yang berdekatan mungkin dianggap sebagai titik perpecahan. Oleh karena itu, mengingat nilai-nilai v A, maka v_1 mungkin perpecahan

dievaluasi. Sebagai contoh, titik tengah antara nilai-nilai a_i dan a_{i+1} dari A adalah:

Jika nilai-nilai A diurutkan terlebih dahulu, kemudian menentukan pemecahan terbaik untuk A hanya membutuhkan satu lulus melalui nilai-nilai. Untuk setiap titik dibagi mungkin bagi A, kita mengevaluasi Info A (D), di mana jumlah partisi adalah dua, yaitu $v = 2$ (atau $j = 1; 2$) dalam Persamaan (6.2). Titik dengan kebutuhan informasi minimum yang diharapkan untuk A dipilih sebagai titik split untuk A. D_1 adalah himpunan tupel di D memuaskan $A \leq \text{split_point}$, dan D_2 adalah himpunan tupel di D memuaskan $A > \text{split_point}$.

Ukuran mendapatkan informasi yang bias terhadap tes dengan hasil yang banyak. Artinya, ia lebih suka untuk memilih atribut memiliki sejumlah nilai-nilai besar. Sebagai contoh, pertimbangkan sebuah atribut yang bertindak sebagai pengenalan unik, seperti *product-ID* Sebuah perpecahan di *product-ID* akan mengakibatkan sejumlah besar partisi (sebanyak nilai-nilai yang ada), masing-masing berisi hanya satu tupel. Karena setiap partisi murni, informasi yang diperlukan untuk mengklasifikasikan kumpulan data D berdasarkan partisi ini akan Info *produk-ID* (D) = 0. Oleh karena itu, informasi yang didapatkan melalui partisi pada atribut ini maksimal. Jelas, seperti partisi tidak berguna untuk klasifikasi. C4.5, penerus dari ID3, menggunakan ekstensi untuk mendapatkan informasi yang dikenal sebagai *gain ratio*, yang mencoba untuk mengatasi bias ini. Ini berlaku semacam normalisasi untuk mendapatkan informasi menggunakan "informasi split" nilai yang didefinisikan secara analogi dengan Info (D) sebagai berikut:

Nilai ini me

wakili informasi potensi yang dihasilkan dengan memisahkan set data pelatihan, D, ke partisi v , sesuai dengan hasil uji v pada atribut A. Perhatikan bahwa, untuk setiap hasil, menganggap jumlah tupel memiliki hasil yang berkenaan dengan jumlah tuple dalam D. Ini berbeda dari mendapatkan informasi, yang mengukur informasi sehubungan dengan klasifikasi yang diperoleh didasarkan pada

partisi yang sama. Rasio gain didefinisikan sebagai.

Atribut dengan rasio keuntungan maksimum dipilih sebagai atribut membelah. Catatan, bagaimanapun, bahwa sebagai informasi perpecahan mendekati 0, rasio menjadi tidak stabil. Kendala A ditambahkan untuk menghindari hal ini, dimana mendapatkan informasi dari tes yang dipilih harus menjadi besar setidaknya sama besarnya dengan keuntungan rata-rata selama semua tes diperiksa.

Perhitungan rasio keuntungan untuk pendapatan atribut. Sebuah tes pada pendapatan membagi data pada Tabel 6.1 menjadi tiga partisi, yaitu rendah, sedang, dan tinggi, yang mengandung empat, enam, dan empat tupel, masing-masing. Untuk menghitung rasio keuntungan dari pendapatan, pertama-tama kita gunakan persamaan dibawah ini untuk mendapatkan

$$= 0,926$$

Dari contoh tersebut, kita memiliki
 $\text{Gain}(\text{income}) = 0.029$. oleh karena itu
 $\text{GainRatio}(\text{income}) = 0.029/0.926 = 0.031$.

Metode Penelitian

Metode penelitian yang dilakukan melalui tahapan penelitian sebagai berikut:

1. Pengumpulan Data
 Pengumpulan data diambil dari data kredit dari koperasi simpan pinjam (KSP) Artha Abadi Jepara mulai tahun 2009-2014, dengan jumlah data sebanyak 1.560
2. Pengolahan Awal Data (Pre Processing)
 Data yang diperoleh dari KSP ditransformasi untuk mendapatkan atribut yang benar-benar sesuai atau relevan dengan format input algoritma yang dikomparasi dan bersesuaian dengan rapid miner.
3. Metode yang diusulkan
 Metode yang diusulkan adalah metode analisa untuk mengetahui tingkat akurasi dari algoritma yang akan dikomparasi yaitu: Algoritma C4.5, k- hasil komparasi digunakan untuk menentukan prediksi kredit macet.
4. Eksperimen dan Pengujian Methode
 Algoritma yang akan dikomparasi dengan menggunakan data kredit dari KSP Artha

Abadi yang sesuai, dimana sebagian besar (80%) akan digunakan sebagai data training dan sisanya (20%) digunakan untuk data testing. Komparasi algoritma akan dilakukan beberapa kali agar mendapatkan besaran parameter yang terbaik.

5. Evaluasi dan Validasi Hasil

Evaluasi dapat dilakukan dengan melakukan pengamatan dan sekaligus menganalisa hasil komparasi keempat algoritma.

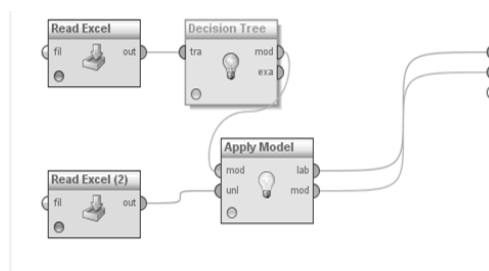
Sedangkan validasi dilakukan dengan mengukur hasil prediksi dengan membandingkan dengan data awal.

Pengukuran dilakukan dengan menggunakan ROC Curve dan Confusion Matrik sampai mendapatkan akurasi yang tinggi.

Hasil Dan Pembahasan

Pengujian menggunakan algoritma C4.5 didahului perhitungan klasifikasi dari data set yang digunakan, hasil yang diperoleh dari perhitungan klasifikasi seperti pada tabel 3.1 berikut:

Dengan bantuan *Rapid Miner* dapat dibuat model pohon keputusan dengan masukan data yang telah disesuaikan seperti pada perhitungan *entropy* di atas, adapun model yang terbentuk seperti pada gambar 4.1. Pertama-tama dimasukkan data training yang telah disiapkan dengan melalui *read excel*, berikut dicari model *decision tree*, setelah itu masukkan data testing dan *apply model*, kemudian disambungkan seperti tampak pada gambar 3.1.



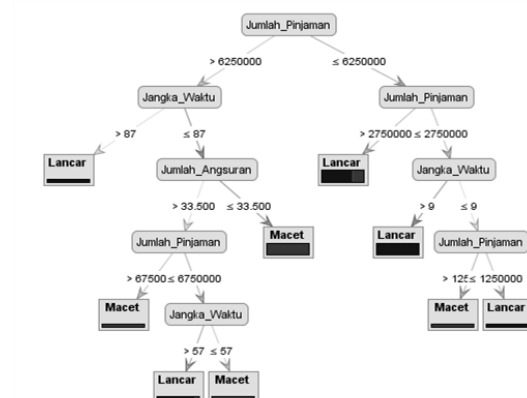
Gambar 4.1 Desain model C 4.5

Setelah model terbentuk maka selanjutnya akan dilakukan proses *running* pada *rapid miner*. Hasil pengujian dengan menggunakan *rapid miner* pada metode C4.5 dengan desain model seperti pada gambar 4.1 akan menghasilkan data *testing* pada tabel 4.1, dimana pada tabel ini memberikan informasi bahwa dalam tabel ini memunculkan prediksi lancar dan macet.

Tabel 4. 1 Data Hasil Pengujian C4.5

Row No.	Penghasilan	confidence	confidence	prediction	S_Jumlah_Pin	Jangka_Wa	Jasa_Pinja	Jumlah_An
1	3500000	0	1	Lancar	1000000	24	90000	10
2	1500000	0	1	Lancar	1000000	24	90000	10
3	1500000	0	1	Lancar	1000000	18	90000	10
4	1500000	0.268	0.732	Lancar	3000000	90	270000	10
5	1500000	0	1	Lancar	1500000	24	135000	10
6	1500000	0	1	Lancar	1500000	72	135000	10
7	2000000	0.268	0.732	Lancar	3500000	84	315000	10
8	2000000	0.268	0.732	Lancar	4000000	42	360000	10
9	2500000	0	1	Lancar	2500000	18	225000	10
10	2500000	0.268	0.732	Lancar	3000000	90	270000	10
11	2500000	0	1	Lancar	2000000	90	180000	10
12	3000000	0.268	0.732	Lancar	5000000	72	450000	10
13	3000000	0	1	Lancar	1000000	60	90000	20
14	4000000	0.268	0.732	Lancar	3000000	90	270000	20
15	1500000	0	1	Lancar	1500000	24	135000	20
16	2000000	0.268	0.732	Lancar	4500000	30	405000	20
17	3000000	0	1	Lancar	1500000	42	135000	20
18	3500000	0.268	0.732	Lancar	4000000	78	360000	20
19	3000000	0.268	0.732	Lancar	3000000	84	270000	20
20	4000000	0	1	Lancar	1500000	36	135000	25
21	2000000	0	1	Lancar	8500000	90	585000	25

Dari desain model tersebut, maka *Rapid Miner* akan bekerja berdasarkan input data yang telah disiapkan terlebih dahulu, maka hasil model pohon keputusan yang diperoleh seperti terlihat pada gambar 4.2:

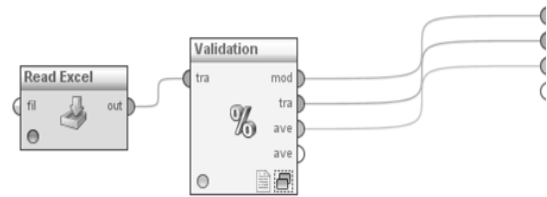


Gambar 4.2. pohon keputusan C4.5

Dari gambar 4.2. pohon keputusan C4.5 maka dapat dibuat aturan atau *rule* seperti berikut ini:

- R1: if Jumlah_Angsuran ≥ 12.500 then LANCAR
- R2: if Jumlah_Angsuran ≤ 12.500 AND Jumlah_Pinjaman $\geq 3.250.000$ = then MACET

- c. R3: if Jumlah_Angsuran $\leq$ 12.500 AND Jumlah_Pinjaman $\leq$ 3.250.000 AND Jumlah_Pinjaman $\geq$ 2.750.000 AND Jangka_Waktu $\geq$ 19.5 = then LANCAR
- d. R4: if Jumlah_Angsuran $\leq$ 12.500 AND Jumlah_Pinjaman $\leq$ 3.250.000 AND Jumlah_Pinjaman $\geq$ 2.750.000 AND Jangka_Waktu $\geq$ 19.500 = then MACET
- e. R5: if Jumlah_Angsuran $\leq$ 12.500 AND Jumlah_Pinjaman $\leq$ 3.250.000 AND Jumlah_Pinjaman $\leq$ 2.750.000 AND Jangka_Waktu $\geq$ 10.500 = then LANCAR
- f. R6: if Jumlah_Angsuran $\leq$ 12.500 AND Jumlah_Pinjaman $\leq$ 3.250.000 AND Jumlah_Pinjaman $\geq$ 2.750.000 AND Jangka_Waktu $\leq$ 10.500 AND Jumlah_Pinjaman $\geq$ 1.250.000 = then MACET
- g. R7: if Jumlah_Angsuran $\leq$ 12.500 AND Jumlah_Pinjaman $\leq$ 3.250.000 AND Jumlah_Pinjaman $\geq$ 2.750.000 AND Jangka_Waktu $\leq$ 10.500 AND Jumlah_Pinjaman $\leq$ 1.250.000 = then LANCAR



Gambar 4. 2 Validasi **C4.5**

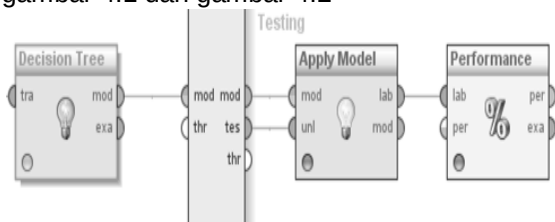
accuracy: 91.06% +/- 2.29% (mikro: 91.05%)			
	true Lancar	true Macet	class p
pred. Lancar	895	135	86.89%
pred. Macet	0	479	100.00%
class recall	100.00%	78.01%	

Gambar 4.3 Hasil Accuracy C4.5

Gambar 4.3 memberikan informasi pada kita bahwa hasil accuracy dari metode klasifikasi C4.5 sebesar = 91,06% ini menunjukkan bahwa hasil akurasi yang diperoleh masuk dalam katagori sangat baik.

Pengujian Model C4.5

Pengujian model C4.5 dengan menggunakan rapid miner kali akan menguji akurasi metode C4.5, seberapa besar akurasi yang akan dihasilkan yang nantinya hasil pengujian akurasi akan dibandingkan dengan metode yang lain. Adapun langkah-langkah yang akan dilakukan dalam pengujian ini menggunakan performa untuk menghasilkan tingkat akurasi dari metode yang digunakan. Pengujian model C4.5 dengan menggunakan testing dan validasi seperti ditunjukkan pada gambar 4.1 dan gambar 4.2



Gambar 4. 1 Model Testing C4.5

Gambar 4.2 menjelaskan teknik validasi dalam rapid miner dengan cara mengimport data yang telah disiapkan dan menyambungkannya dengan model validasi yang ada, setelah di-running maka akan muncul hasil *accuracy*, *presision* dan *recall* yang dikehendaki.

precision: 100.00% +/- 0.00% (mikro: 100.00%) (positive class: Macet)			
	true Lancar	true Macet	class p
pred. Lancar	895	135	86.89%
pred. Macet	0	479	100.00%
class recall	100.00%	78.01%	

Gambar 4. 4 Hasil Precision C4.5

Precision yang dihasilkan dari metode C4.5 yang diproses dengan rapid miner menghasilkan presisi yang sangat baik dengan nilai *precision* sebesar = 100,00%.

recall: 78.00% +/- 5.71% (mikro: 78.01%) (positive class: Macet)			
	true Lancar	true Macet	class p
pred. Lancar	895	135	86.89%
pred. Macet	0	479	100.00%
class recall	100.00%	78.01%	

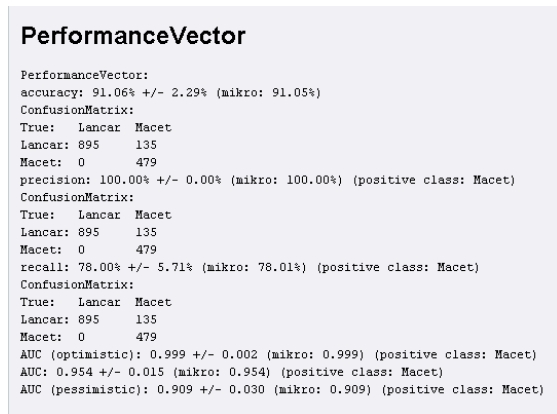
Gambar 4. 5 Hasil Recall dari C4.5

Gambar 4.5 memberikan informasi pada kita bahwa hasil recall C4.5 yang diperoleh melalui proses validasi sebesar = 78,00% nilai ini menunjukkan bahwa hasil yang diperoleh adalah baik.

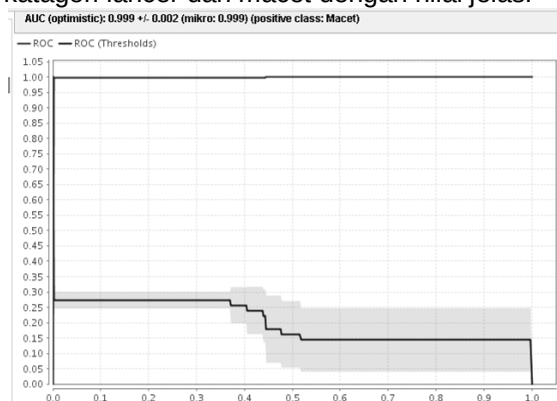
Secara keseluruhan nilai hasil validasi adalah *accuracy* = 91,06%, *precision* = 100,00% dan *recall* = 78,00%, hasil ini nantinya akan dibandingkan dengan algoritma yang lain

untuk menentukan algoritma mana yang mendapatkan nilai tertinggi yang akan digunakan untuk memprediksi kredit macet.

Tabel 4 2 *Performa Vector C4.5*

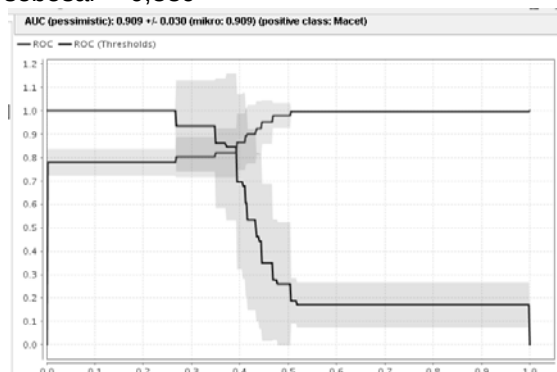


Tabel 4.2 memberikan informasi secara keseluruhan tentang hasil dari validasi dengan metode klasifikasi C4.5, disamping hasil *accuracy*, *precision* dan *recall*, masih ada informasi penting dalam *performance vector* ini yaitu *confusion matrix* pada masing-masing katagori lancar dan macet dengan nilai jelas.



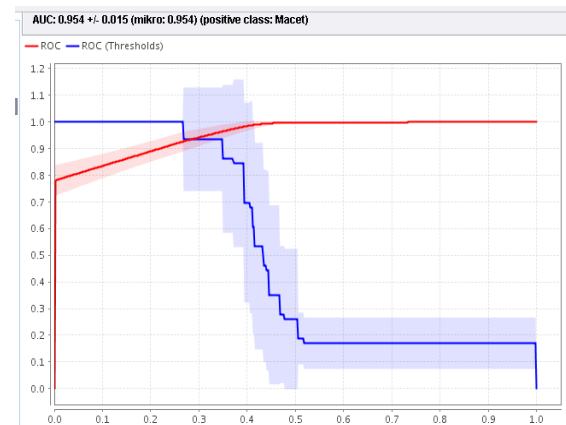
Gambar 4. 6 Hasil AUC *Optimistic* dari C4.5

Gambar 4.6 menggambarkan grafik *area under curve* (AUC) optimistic hasil validasi C4.5 dengan nilai yang tertera dalam gambar sebesar = 0,889



Gambar 4. 7 Hasil AUC *Pesimistic* dari C4.5

Gambar 4.7 menggambarkan grafik *area under curve* (AUC) pesimistic hasil validasi C4.5 dengan nilai yang tertera dalam gambar sebesar = 0,909

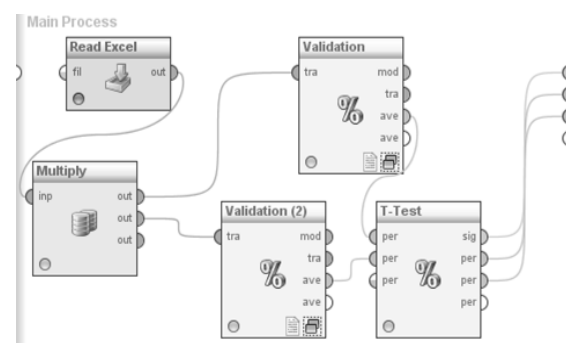


Gambar 4. 8 Hasil AUC dari C4.5

Gambar 4.8 menggambarkan grafik *area under curve* (AUC) hasil validasi C4.5 dengan nilai yang tertera dalam gambar sebesar = 0,854

Pengujian T-Test

Pengujian T-Test ini akan menguji algoritma 4.5 agar mendapatkan nilai yang terbaik, dimana dalam pengujian tersebut sampai mendapatkan nilai terkecil $\leq 0,05$ dinyatakan sebagai hasil uji yang terbaik (Santoso. S, 2010). Gambar 4.35 menggambarkan desain *T-Test* dengan menggunakan *rapid mier*, dimana penggunaan desain ini dilakukan berulang-ulang sampai menghasilkan nilai yang terkecil.



Hasil pengujian algoritma C 4.5 dengan menggunakan *T-Test* menghasilkan akurasi seperti pada tabel 5.1 berikut:

	C4.5
Accuracy	92,00%
AUC	0,833

Test	baik
------	------

Tabel 5.1. Hasil akurasi pengujian Algoritma C4.5 dengan *T-Test*

Kesimpulan

Hasil penelitian yang dilakukan dengan menggunakan metode klasifikasi data mining dengan menggunakan algoritma C4.5 untuk deteksi kredit macet telah menghasilkan nilai akurasi, dimana akurasi tersebut adalah algoritma C4.5 sebesar 92,00%, Dari hasil akurasi tersebut maka algoritma C4.5 merupakan algoritma yang baik keakuratannya digunakan untuk meprediksi kredit macet pada koperasi simpan pinjam. Akurasi sebesar 92,00% tersebut dirasa sudah baik menurut *witen* (2011) untuk mendeteksi perilaku nasabah apakah nasabah yang akan diberi kredit seberapa besar potensi untuk tidak membayar angsuran atau macet.

Setelah dilakukan pengujian menggunakan algoritma C 4.5 mendapatkan hasil perbandingan keseluruhan dengan uji statistic *T-Test*, maka dapat diketahui algoritma C 4.5 yang mempunyai nilai uji test cukup tinggi yaitu nilai probalitas sebesar 0,05, karena memiliki nilai probalitas kurang dari ≤ 0.05 .

Saran

Hasil atau kesimpulan dari pengujian algoritma C4.5 seperti yang ditunjukkan dalam hasil pengujian baik menggunakan pengujian *AUC*, *ROC* maupun *T-Test* dengan bantuan *rapid miner* ada banyak hal yang dapat disampaikan kepada para pengambil keputusan di manajemen koperasi simpan pinjam utamanya pada manajer dan pengurus. Penelitian di atas menunjukkan bahwa algoritma C4.5 mempunyai akurasi yang cukup tertinggi jika digunakan untuk melakukan prediksi tingkah laku nasabah di koperasi simpan pinjam' kesimpulan dapat dipergunakan oleh para pengambil keputusan di koperasi simpan pinjam terutama para surveyor dan analis yang memberikan data calon nasabah kredit baru, diantaranya:

1. Dari data yang disajikan para surveyor dalam menyajikan data calon nasabah kredit baru dapat diketahui atau dideteksi bahwa calon nasabah tersebut akan mengalami kemungkinan macet atau lacer.
2. Dalam pengujian tersebut dapat ditambah beberapa parameter yang dapat melengkapi keadaan calon nasabah, karena semakin banyak

parameter yang digunakan untuk pengujian ini akan semakin akurat hasilnya.

3. Akurasi algoritma C4.5 masih dapat ditingkatkan hasilnya dengan menambahkan parameter baru yang sesuai dengan keadaan calon nasabah, karena sifat pengujian dengan bantuan *rapid miner* ini masih tergantung dengan jumlah parameter dan jumlah nasabah yang dijadikan pengujian.
4. Pengujian dengan bantuan *rapid miner* juga dapat digunakan untuk membandingkan antara algoritma C4.5 dengan algoritma lainnya sehingga hasil pengujian dengan cara membandingkan akan mempunyai akurasi yang lebih baik hanya saja membutuhkan waktu yang lebih panjang.
5. Bagi pengambil keputusan di koperasi simpan pinjam dapat menerima atau menolak nasabah kredit baru sehubungan dengan hasil prediksi perilaku nasabah.

Daftar Pustaka

- Arifin Sitio and Halomoan Tamba, *Koperasi Teori dan Praktik*. Jakarta, Indonesia: Erlangga, 2001.
- Teguh Pudjo Muljono, *Manajemen Perkreditan Bagi Bank Komersial*, 4th ed. Yogyakarta, Indonesia: BPFE-Yogyakarta, 2001.
- Budi Santosa, *Data mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis Teori dan Aplikasi*, 1st ed. Surabaya, Indonesia: Graha Ilmu, 2007.
- Lan Yu, Guoqing Chen, Andy Koronios, Shiwu Zhu, and Xunhua Guo, "Application and Comparism of Classification Techniques in Controlling Credit Risk," *Recent Advances in Data Mining of Enterprise Data: Algorithms and Applications*, pp. 111-145, May 2007.
- K Niki Kunene and Jozef Zurada, "Comparism of the Perforomance of Computational Ittellegence Methods for Loan Granting Decisions," in *Hawaii International Confrence on System Sciences*, Louisville, Januari, 2011, pp. 1-10.
- Hassan Sabzevari, Mehdi Soleymani, and Eaman Noorbakhsh, "A comparism Between Statistical and Data Mining Method for Credit

- Scoring in Case of Limited Available Data," p. 8, June 2008.
- Hengwen Cai, Hong Yu, Xiaolei Huang, and Xiaorong Hu, "A Comparative Study on Data Mining Algorithm for Individual Credit Risk Evaluation," in *Management of e-Commerce and e-Government, International Conference on*, Nanchang, 330029, October, 2010, pp. 35-38.
- Yi Jiang, Qingshan Chen, Hongyi Huang, and Defu Zhang, "A Comparison Study of Credit Scoring Models," in *International Conference on Natural Computation*, Xiame, August, 2007, pp. 15-18.
- Ian H Witten, Eibe Frank, and Mark A Hall, *Data Mining Practical Machine Learning Tool and Techniques*, Third Edition ed. Burlington, United States of America: Elseiver Inc., 2011.
- Jiawei Hand and Micheline Kamber, *Data Mining Concept and Techniques*, Second Edition ed., Jim Gray, Ed. San Francisco, United States of America: Elseiver Inc., 2006.
- Xindong Wu and Vipin Kumar, *The Top Ten Algorithms in Data Mining*, 1st ed. Minnesota, USA: CRC Press Taylor & Francis Group, 2009.
- T Sutejo, Edy Mulyanto, and Vincent Suhartono, *Kecerdasan Buatan*, Perdana ed. Semarang: Andi Offset, 2011.
- Rong-Zhou Li, Su-Lin Pang, and Jian-Min Xu, "Neural Network Credit-Risk Evaluation Model Based On Back-Propagation Algorithm," *Prosedings of the First International Conference on Machine Learning and Chbernetic, Beijing, 4-5 November 2002*, p. 17021706, November 2002.
- Lin Lin and Nantian Huang, "Credit Risk Assessment Using BP Neural Network with Demspter-Shafer Theory," *International Conference on Artifial Intellegence*, pp. 397-401, 2009.
- Florin Gorunescu, *Data Mining Concept, Model and Technique*, 12th ed., Prof.Janusz Kecprzyk and Prof.Lakhmi C Jain, Eds. Verlag Berlin Heldelberg: Springer, 2011.
- Santoso S, *Statistik Parametrik-Konsep dan Aplikasi dengan SPSS*. Jakarta: PT. Eka Media Komputindo, 2010.
- Budi Warsito, *Kapita Selecta Statistika Neural Network*, Perdana, Ed. Semarang: BP Undip, 2009.
- Adel Lahsasna, Raja Noor Ainon, and The Ying Wah, "Credit Scoring Models Using Soft Computing Methods: A Survey," *The International Arab Journal of Information Technology*, vol. 7, no. 2, April 2010.
- John C. Mowen / Michael Minor, "Perilaku Konsumen" Jilid 2 Edisi Kelima, Erlangga, 2002
- Prof. Dr. Ir. Ujang Sumarwan, M.Sc, "Perilaku Konsumen" , Edisi Kedua, Ghalia Indonesia, 2011.
- Dr. (Cand) Taswan, SE, M.Si, "Manajemen Perbankan" Konsep, Teknik dan Aplikasi, Edisi Kedua UPP STIM YKPN, Yogyakarta, 2010.
- Mudrajat Kuncoro, Suhardjono, "Manajemen Perbankan" Teori dan Aplikasi, Edisi Kedua, Fakultas Ekonomi dan Bisnis UGM, 2012

