

PENERAPAN ALGORITMA C4.5 UNTUK MENGANALISIS KELAYAKAN PEMBERIAN KREDIT NASABAH

Susanto¹, Neneng Sri Uryani², Cepi Rahmat Hidayat³

STMIK Tasikmalaya

JL. RE Martadinata 272 A, Kota Tasikmalaya – Jawa Barat (0265) 310830

Email: [1susantoprutama@yahoo.co.id](mailto:susantoprutama@yahoo.co.id), [2neneng_sri_u@yahoo.com](mailto:neneng_sri_u@yahoo.com), [3ranvix14@gmail.com](mailto:ranvix14@gmail.com)

Abstrak

Dalam pemberian Kredit perlu menganalisa kebutuhan kreditur, maka yang harus diketahui terlebih dahulu adalah prinsip-prinsip yang perlu ditegakan dalam rangka pemberian Kredit. Hal-hal yang perlu diperhatikan dalam pemberian kredit bagi nasabah adalah Prinsip 6 C's Analysis. Perlu adanya solusi pemecahan masalah yang ada, dengan membuat suatu Sistem Pendukung Keputusan. Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Semakin kaya informasi atau pengetahuan yang dikandung oleh data training, maka akurasi pohon keputusan akan semakin meningkat. Dengan dihasilkannya klasifikasi kelayakan pemberian kredit nasabah dengan jumlah 8 atribut menghasilkan akurasi 87.36% merupakan tingkat akurasi yang baik dibandingkan dengan hasil penelitian klasifikasi kelayakan pemberian kredit nasabah yang menggunakan 4 atribut menghasilkan akurasi 79.50%.

Kata kunci: Analisis kredit, Prinsip 6 C's Analysis, Data Mining dan Algoritma C4.5

Abstract

In granting credit need to analyse the needs of creditors, then that should be noted beforehand is the principles that need to be ditegakan in order to grant credits. The things that need to be observed in granting credit for customers is the principle 6 C's Analysis. Need for troubleshooting existing solutions, by making a decision support system. The algorithm is the algorithm 4.5 C used to form a decision tree. Decision tree classification method and prediction is a very powerful and famous. The rich information or knowledge that is conceived by data training, then the accuracy of the decision tree will progressively increase. With the resultant classification feasibility of granting credit borrowers with a total of 8 attributes produces 87.36% accuracy is a good degree of accuracy compared with the results of the study the feasibility of granting credit customer classification that uses 4 attribute generates accuracy 79.50%.

Keywords: Credit Analysis, principles 6 C's Analysis, Data Mining and algorithms C4.5

1. PENDAHULUAN

Banyak penelitian membahas mengenai penentuan kelayakan pemberian kredit dengan berbagai algoritma data mining. Seperti penelitian yang dilakukan Abbas Heiat (2011) menyatakan bahwa resiko bagi lembaga keuangan untuk memberikan kredit yang diminta tergantung pada seberapa baik mereka membedakan pemohon kredit yang baik dari para pemohon kredit macet[1]. Di bawah ini adalah beberapa penelitian yang berkaitan dengan masalah kredit yaitu:

- a. Jiang (2009) membuat model untuk memprediksi nasabah yang bermasalah dan tidak bermasalah dalam pembayaran kredit dengan menggunakan model algoritma C4.5. Data yang digunakan diambil dari perusahaan German credit yang merupakan perusahaan pembiayaan. Jiang mengambil beberapa atribut dan kemudian dimasukkan ke dalam model untuk memprediksi persentase nasabah yang bermasalah[2]. Pada penelitian ini, peneliti

menyatakan hasil penelitiannya yaitu Statistik menunjukkan bahwa biaya *misclassifying* kredit lancar dan kredit macet adalah 5~20kali dari *misclassifying* kredit lancar dan kredit macet. Ketika jatah dan adalah mendekati atau lebih besar dari 2, SAA mengidentifikasi semua nasabah kredit lancar dan kredit macet karena keterbatasan fungsi diskriminan, ketika C4.5 dikombinasikan dengan SAA, aturan keputusan evaluasi relative stabil. Ini tidak mengklasifikasikan semua nasabah kredit lancar dan kredit macet bahkan dengan diskriminan fungsi, tingkat kesalahan keseluruhan untuk SAA adalah 34% yang menunjukkan bahwa bank-bank kehilangan beberapa pelanggan yang baik.

- b. S. Satchidananda and J. B. Simha (2006) Penelitian ini membandingkan dua model algoritma untuk analisa resiko kredit, yaitu Pohon Keputusan dan Regresi Logistik. Data diambil dari dua bank yang berbeda, kemudian untuk mengelompokkan kasus positif dan negatif maka dilakukan klustering data dengan menggunakan k-means. Hasil analisa dari masing-masing model dikomparasi dan kemudian diukur, kemudian didapatkan bahwa algoritma pohon keputusan mempunyai tingkat akurasi yang tinggi dibandingkan algoritma regresi logistik[3]. Penelitian ini masih dalam proses untuk menyelidiki kinerja yang diusulkan, Pendekatan dibandingkan dengan teknik klasifikasi lainnya untuk *credit scoring* sehingga tingkat akurasi belum dapat diketahui.
- c. Firmansyah (2011), juga melakukan penelitian dengan judul “Penerapan Algoritma Klasifikasi C4.5 untuk Penentuan Kelayakan Pemberian Kredit Koperasi” Nilai *accuracy*, *precision*, dan *recall* nyadari data training dapat dihitung dengan menggunakan Rapid Miner. Setelah diuji coba dengan metode *crossvalidation*, didapatkan hasil pengukuran terhadap data trainingnya yaitu hanya mencapai *accuracy* = 79.50%, *precision* = 86.50% dan *recall* = 91.00%[4]. Hasil pengujian tersebut berdasarkan 5 parameter, merupakan masalah penelitian Teknik Informatika yang masih bisa di tingkatkan akurasi. mengapa Hasil penelitian sebelumnya “Penerapan Algoritma Klasifikasi C4.5 untuk Penentuan Kelayakan Pemberian Kredit Koperasi” Nilai *accuracy* nya hanya 79.50 % ? oleh karena itu penulis berkesempatan untuk meningkatkan hasil penelitian yang dilakukan oleh Firmansyah dengan cara menambahkan 1 parameter sehingga jumlah parameter nya menjadi 6 parameter dalam mengklasifikasikan Nasabah atau *Debitur*”

2. METODE PENELITIAN

- a. Tahap *data understanding*.

Data diperoleh dari salah satu Bank BPR di Tasikmalaya pada tahun 2012. Data tersebut sebanyak 700 record memiliki atribut Nama_Nasabah, Jenis_Kelamin, Umur, Jumlah_Pinjaman, Jangka_Waktu, Jumlah_Angsuran_Perbulan, Type_Pinjaman, Jenis_Pinjaman, Bi_Sektor_Ekonomi, Col, Bi_Golongan_Debitur, Bi_Golongan_Penjamin, Saldo_Nominatif, Tunggakan_Pokok, Tunggakan_Bunga, Status_Kredit. Nilai dari semua atribut yang ada di tabel, merupakan nilai kategorikal dan bukan nilai angka, misalnya seperti atribut umur, yaitu debitur yang berusia 17 th sampai 40 tahun termasuk dalam kategori muda, sedangkan debitur yang berusia 41 tahun sampai 55 tahun termasuk kategori paruh baya, dan kategori ketiga adalah debitur yang berusia diatas 55 tahun termasuk kategori tua. Tabel 6 di bawah ini ditampilkan nama atribut, kategori, dan nilai angka (rangnya). Berikut rule nilainya :

Tabel 1 Kategori atribut

Atribut	Nilai Angka	Kategori
jenis_kelamin	L	Laki-laki
	P	Perempuan
Umur	17 - 40	Muda
	41- 55	paruh baya
	> 55	Tua
jml_pinjaman	Rp. 1.000.000- Rp.	Kecil

	5.000.000	
	Rp. 5.000.001- Rp. 10.000.000	Sedang
	> Rp. 10.000.000	Besar
jangka waktu	6-12 bulan	Cepat
	12-18 bulan	Sedang
	> 18 bulan	Lambat
jml_angsuran_per_bulan	<= Rp. 100.000	Kecil
	<= Rp. 1000.000	Sedang
	> Rp. 1000.000	Besar
bi_gol_penjamin	000	tanpa penjamin
	835	perusahaan2 lainnya
	875	Perorangan
saldo_nominatif	Rp. 1.000.000- Rp. 5.000.000	Kecil
	Rp. 5.000.001- Rp. 10.000.000	Sedang
	> Rp. 10.000.000	Besar
tunggakan_pokok	<= Rp. 500.000	Kecil
	<= Rp. 1000.000	Sedang
	> Rp. 1000.000	Besar
tunggakan_bunga	<= Rp. 100.000	Kecil
	<= Rp. 1000.000	Sedang
	> Rp. 1000.000	Besar

b. Tahap *data preparation*

Tabel di bawah ini menunjukkan data transaksi kredit baik yang bermasalah maupun yang tidak bermasalah. Pada tabel ini terdapat 210 record data nasabah yang terdiri dari 8 atribut yaitu : Jenis_Kelamin, Umur, Jumlah_Pinjaman, Jangka_Waktu, Jumlah_Angsuran_Perbulan, Bi_Golongan_Penjamin, Saldo_Nominatif, Tunggakan_Pokok, Status_Kredit. Tabel terlampir.

Tabel data set merupakan data murni yang belum diolah menjadi data training. Jika kita lihat secara seksama pada tabel data set ini, maka akan didapati beberapa atribut yang tidak digunakan dan record yang duplikasi. Untuk itu maka diperlukan teknik dalam preprocessing yaitu :

- a. Data cleaning bekerja untuk membersihkan nilai yang kosong ,tidak konsisten atau mungkin tupel yang kosong (missing values dan noisy).
 - b. 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.
 - c. Data *reduction*. Jumlah atribut dan tupel yang digunakan untuk data training mungkin terlalu besar, hanya beberapa atribut yang diperlukan sehingga atribut yang tidak diperlukan akan dihapus. Tupel dalam data set mungkin terjadi duplikasi atau terdapat tupel yang sama, sehingga untuk memperkecil jumlah tupel, tupel yang sama akan dijadikan dalam satu tupel untuk mewakili tupel tersebut akan terlihat pada tabel 7, untuk lebih lengkapnya tabel dapat dilihat di lampiran.
- Telah disebutkan sebelumnya bahwa kredit macet berjumlah 138 nasabah atau debitur, data ini telah dipisahkan dari data pembayar yang tidakbermasalah. Dengan proses training, data

debitur bermasalah menjadi 53 tupel karena telah melalui tahap training. Berikut datanya dalam bentuk tabel 8, untuk lebih lengkapnya tabel terlampir.

c. Tahap *Modeling*

Computing approach pada penelitian ini dipilih berdasarkan studi literatur tentang algoritma yaitu C 4.5 yang mampu mengklasifikasikan status kredit lancar dan kredit macet. Tahap ini akan dibahas di bab III

d. Analisa dan *evaluation pattern*

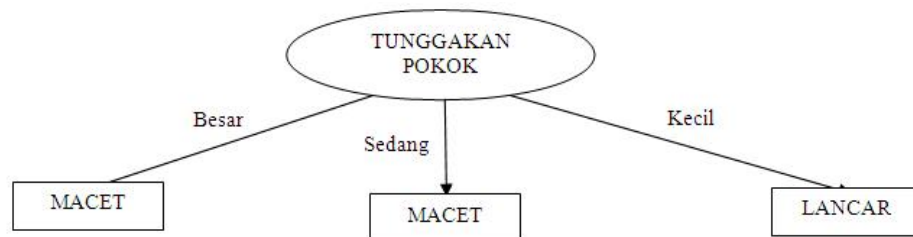
Algoritma yang telah dikembangkan dalam penelitian ini akan diterapkan pada data pembayaran kredit melalui suatu model simulasi. 80% data akan dipergunakan sebagai data training dan 20% data akan digunakan sebagai data checking. Evaluasi dilakukan dengan mengamati hasil klasifikasi dari penerapan C4.5. Pengukuran tingkat akurasi dilakukan dengan menggunakan evaluasi model *confusion matrix* dan kurva ROC (*Receiver Operating Characteristic*). Dengan demikian dapat diketahui berapa tinggi akurasi dari algoritma tersebut.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Pemodelan dan Pembahasan

Pada tabel 11 di atas, pada node 1.3 untuk *leaf* tunggakan pokok kategori besar, tidak dilakukan perhitungan, karena nilai entropy untuk Tunggakan Pokok besar adalah 0.00000 sehingga jika dilakukan perhitungan maka hasilnya akan 0.00000

Dari hasil perhitungan di atas, maka didapatkan model pohon keputusan seperti berikut:



Gambar 1 Model pohon keputusan menggunakan algoritma C 4.5

Dari model pohon keputusan tersebut didapat aturan/ pola sebagai berikut :

tunggakan_pokok = Besar: MACET {LANCAR=0, MACET=18}

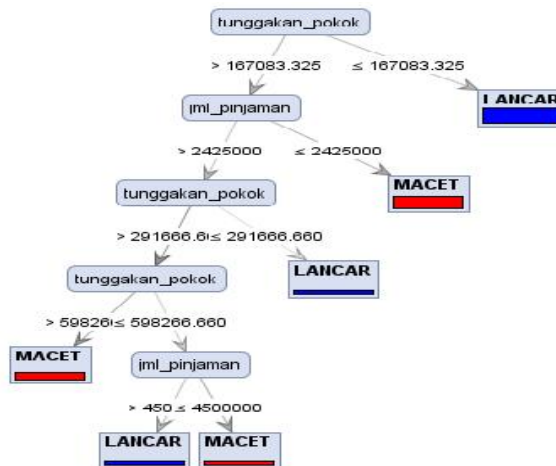
tunggakan_pokok = Kecil: LANCAR {LANCAR=49, MACET=24}

tunggakan_pokok = Sedang: MACET {LANCAR=1, MACET=11}

3.2. Hasil Pengujian

Tahapan selanjutnya adalah melakukan eksperimen 8 atribut data training dengan menggunakan Framework RapidMiner. Hal yang pertama dilakukan adalah mengimpor data set yang sudah disiapkan sebelumnya. Untuk menganalisa, dibutuhkan data training. Data training yang akan dimasukkan ke dalam Rapid Miner bisa dalam format .csv, .xls, .mdb dan lain-lain. Data yang penulis gunakan adalah dalam format .xlsx. dari dataset yang didapat dari tempat penelitian yang sudah diolah menjadi data training dengan menggunakan teknik preproccesing. Kemudian dilakukan pengujian model dengan menggunakan RapidMiner. Algoritma yang pertama diterapkan adalah algoritma C 4.5 dan berikut hasil dari pengujian menggunakan Rapid Miner. Dalam hal ini penulis membandingkan hasil pengujian dengan jumlah atribut yang berbeda yaitu experiment pertama penulis menggunakan 9 atribut dan experiment kedua penulis menggunakan 8 atribut atau menghilangkan salah satu atribut pada data training yang sama sehingga dapat diketahui tingkat akurasi yang lebih tinggi.

Pada eksperimen ini penulis sengaja mengurangi salah satu atribut sehingga menjadi 8 atribut pada data training dengan tujuan untuk membandingkan tingkat akurasi yang didapatkan, berikut adalah hasil pengujian yang dilakukan dengan menggunakan 8 atribut.



Gambar 2 gambar pohon keputusan dengan 8 atribut

accuracy: 87.36% +/- 17.79% (mikro: 87.38%)

	true LANCAR	true MACET	class precision
pred LANCAR	46	9	83.64%
pred MACET	4	44	91.57%
class recall	92.00%	83.02%	

Gambar 3 nilai akurasi dengan 8 atribut

precision: 89.67% +/- 20.14% (mikro: 91.57%) (positive class: MACET)

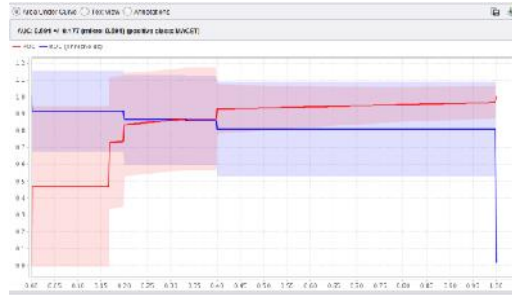
	true LANCAR	true MACET	class precision
pred LANCAR	46	9	83.64%
pred MACET	4	44	91.57%
class recall	92.00%	83.02%	

Gambar 4 Precision 89.67% dengan 8 atribut

recall: 83.33% +/- 24.99% (mikro: 83.02%) (positive class: MACET)

	true LANCAR	true MACET	class precision
pred LANCAR	46	9	83.64%
pred MACET	4	44	91.57%
class recall	92.00%	83.02%	

Gambar 5 nilai Recall 83.33 dengan 8 atribut



Gambar 6 Kurva AUC dengan 8 atribut menunjukkan nilai 0.691

3.3. Evaluasi Model *Confusion Matrix*

TrueMACET merupakan tupel positif di data set yang diklasifikasikan positif berjumlah 46, sedangkan *trueLANCAR* merupakan tupel negatif di data set yang diklasifikasikan negatif berjumlah 44. *FalseMACET* merupakan tupel positif di data set yang diklasifikasikan negatif berjumlah 4, sedangkan *falseLANCAR* merupakan tupel negatif di data set yang diklasifikasikan positif berjumlah 9. Kemudian dari data diatas dapat dihitung beberapa persamaan sebagai berikut:

Tabel Nilai *Precision*, *Accuracy* dan *Recall* untuk algoritma C 4.5

Hasil pengujian menggunakan 8 atribut		Hasil pengujian menggunakan 9 atribut	
<i>Accuracy</i>	87.36	<i>Accuracy</i>	84.64 %
<i>Precision</i>	89.67	<i>Precision</i>	86.33%
<i>Recall</i>	83.33	<i>Recall</i>	81.67 %

Dari hasil perbandingan pada tabel 14 ternyata terbukti bahwa penggunaan 8 atribut lebih tinggi akurasiya dibandingkan dengan penggunaan 9 atribut.

4. KESIMPULAN

Dengan dihasilkannya klasifikasi kelayakan pemberian kredit nasabah dengan jumlah 8 atribut menghasilkan akurasi 87.36% merupakan tingkat akurasi yang baik dibandingkan dengan hasil penelitian klasifikasi kelayakan pemberian kredit nasabah yang menggunakan 4 atribut menghasilkan akurasi 79.50 % , sehingga kelancaran proses penilaian kelayakan kredit dapat tercipta dan pembayaran terlambat (menunggak) sudah terprediksi dari awal untuk dapat diwaspadai agar dapat meminimalisir meningkatnya kredit macet.

Tingkat akurasi algoritma C 4.5 eksperimen ini sudah mencapai tingkat baik, sehingga dapat meningkatkan ketelitian dalam proses klasifikasi dan prediksi dengan cara menambahkan beberapa atribut dari histori pembayaran kredit nasabah yang ada sehingga dihasilkan pola klasifikasi yang lebih akurat.

5. SARAN

1. Untuk memperlancar proses penilaian kelayakan pemberian kredit, sebaiknya pola klasifikasi hasil dari penelitian ini diwujudkan kedalam suatu perangkat lunak aplikasi.
2. Untuk meningkatkan tingkat akurasi, sebaiknya menggunakan data dalam jumlah yang banyak dan bersumber dari beberapa perusahaan Bank lainnya.
3. Untuk mempermudah proses klasifikasi, sebaiknya dapat ditemukan cara untuk menentukan pembobotan pada algoritma C 4.5 secara otomatis.

DAFTAR PUSTAKA

- [1] A. Heiat, 2011, *Modeling Consumer Credit Scoring Through Bayes Network Heiat, Social Sciences*, vol. 1, no. 3, pp. 132–141
 - [2] Y. Jiang, 2009, *Credit Scoring Model Based on the Decision Tree and the Simulated Annealing Algorithm, Learning*, no. 2007, pp. 18–22
 - [3] J. B. Simha, 2006, *Comparing decision trees with logistic regression for credit risk analysis*
 - [4] Firmansyah, M. Ilmu, S. Tinggi, M. Informatika, D. A. N. Komputer, and N. Mandiri, 2011, *Penerapan algoritma klasifikasi c4.5 untuk penentuan kelayakan pemberian kredit koperasi tesis*
-