

PENERAPAN ALGORITMA DECISION TREE C4.5 UNTUK PENILAIAN RUMAH TINGGAL

Budi Setiadi
Fakultas Teknik Uniska MAAB Banjarmasin
Email : budibtc75@yahoo.com

ABSTRACT

There is still a possibility of assessment error homes as a reference value of credit, which will open opportunities for NPL. So we need a way of assessment (predictive value) is quite proportional, credible and accurate. Inaccurate predictions led to the planning of improper credit management. Prediction value of collateral house has attracted the interest of many researchers because of its importance both in theoretical and empirical.

Namely C4.5 decision tree algorithm, CART and CHAID that can be used for credit risk status. The third tree algorithm produces different models for the same data set. Therefore, this study aims to implement a C4.5 decision tree algorithm for the assessment of the residence. Evaluation results will be processing using precision and recall, and then compared and analyzed the results between assessors using other analysis methods (Naive Bayes, K-NN) with the results predicted by the method of classification algorithm C4.5. From here will look the accuracy of the implementation of C4.5.

Keywords : Classification Algorithm, Decision Tree, C4.5, Assessment

1. PENDAHULUAN

Kredit bermasalah di perbankan dapat disebabkan oleh berbagai hal apakah itu kesengajaan pihak yang terlibat dalam proses kredit atau kesalahan prosedur atau mungkin faktor makro ekonomi.

Untuk dapat mencari permasalahan tersebut yang pertama dilihat tentu saja adalah prosedur atau proses kredit tersebut dilaksanakan dengan benar atau ada penyimpangan. Apabila prosedur telah dijalankan dengan benar maka perlu dikaji apakah ada kesengajaan pihak-pihak yang terlibat dalam memberikan rekomendasi, opini dan informasi yang tidak sebenarnya. Atau mungkin perlu

dikaji apakah ada faktor makro ekonomi yang menyebabkan penyimpangan terhadap proyeksi atau rencana yang telah ditetapkan.

Pada umumnya pihak perbankan atau penilai (jasa pihak ketiga) menggunakan instrumen analisis yang dikenal dengan the five of credit atau the 5 C, yaitu *character* (kepribadian) yaitu penilaian atas karakter atau watak dari calon debitornya, *capacity* (kemampuan) yaitu prediksi tentang kemampuan bisnis dan kinerja bisnis debitur untuk melunasi hutangnya, *capital* (modal) yaitu penilaian kemampuan keuangan debitur yang mempunyai korelasi langsung dengan tingkat kemampuan bayar kreditor, *condition of economy* (kondisi ekonomi) yaitu analisis terhadap kondisi perekonomian debitur secara mikro maupun makro dan *collateral* (agunan) yaitu harta kekayaan debitur sebagai jaminan bagi pelunasan hutangnya jika kredit dalam keadaan macet.

Kesengajaan pihak-pihak yang terlibat dalam proses kredit yang mungkin terjadi baik dari pihak calon debitur, pihak kreditor maupun konsultan independen antara lain :

- Pemberian informasi dan dokumen yang tidak sebenarnya oleh calon debitur baik dokumen administrasi, legal maupun laporan keuangan.
- Rekayasa terhadap feasibility yang dilakukan oleh pihak konsultan akibat pengaruh calon debitur atau oknum kreditor.
- Rekayasa terhadap penilaian yang dilakukan oleh penilai independen pengaruh dari calon debitur atau oknum kreditor.
- Rekayasa terhadap analisa kelayakan kredit dan analisa keuangan yang dilakukan oleh oknum divisi risk management akibat pengaruh dari calon debitur atau oknum kreditor.
- Rekayasa terhadap laporan pengawasan yang dilakukan oleh konsultan pengawas akibat pengaruh dari calon debitur atau oknum kreditor.

Apabila prosedur telah berjalan dengan benar dan tidak terdapat rekayasa seperti di atas maka kemungkinan kredit bermasalah/NPL (non performing loan) terjadi akibat faktor luar atau faktor makro ekonomi yang mungkin di luar kemampuan pihak debitur maupun pihak kreditor. Sehingga kemungkinan terjadi NPL yang terdapat kecenderungan adanya tindak kolusi dan korupsi adalah

akibat penyimpangan prosedur dan rekayasa dari pihak pihak yang terlibat dalam proses kredit.

Dengan demikian, meskipun telah menggunakan jasa pihak ketiga, masih terdapat kemungkinan kesalahan penilaian agunan sebagai acuan nilai kredit, yang akan membuka peluang terjadinya NPL. Jadi diperlukan suatu cara penilaian (prediksi nilai) yang cukup proporsional, kredibel dan akurat.

Prediksi yang tidak akurat menyebabkan perencanaan manajemen kredit yang tidak tepat. Prediksi nilai agunan telah menarik minat banyak peneliti karena nilai pentingnya baik di teoritis dan empiris. Banyak model dan metode telah disajikan di masa lalu. Perlu algoritma yang efektif untuk prediksi nilai agunan untuk pengajuan kredit.

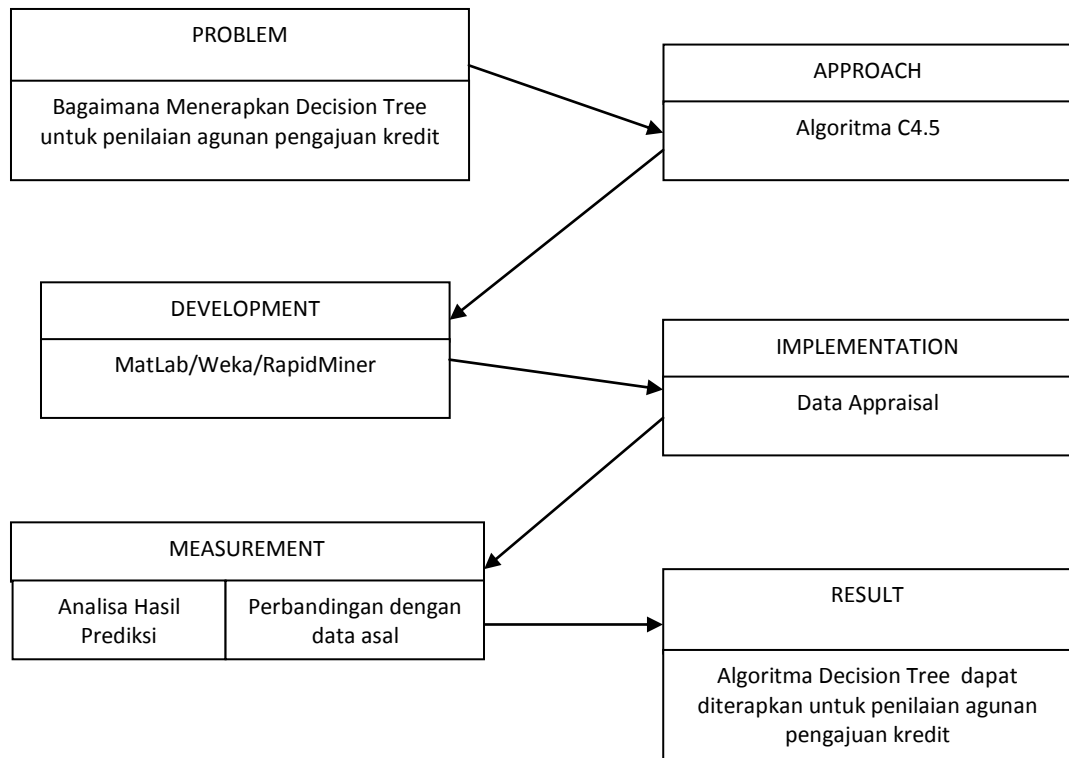
Josef Zurada dan K. Niki Kunene [1] menegaskan bahwa penting bagi institusi keuangan untuk secara akurat mengevaluasi risiko kredit yang mereka timbulkan tiap kali memutuskan sebuah pemberian pinjaman, sesuatu yang tidak dapat diabaikan; Dengan demikian, kebutuhan model penilaian yang akurat sangat penting dalam ekonomi yang mengandalkan pada ketersediaan kredit untuk kegiatan ekonomi sehari-hari. Penelitian tentang model penilaian kredit terus berkembang dan mengeksplorasi berbagai metode termasuk survival analysis, linear discriminant analysis (LDA), logistic regression (LR), k -nearest neighbor (k NN), classification trees (CT), neural networks (NN), radial basis function neural networks (RBFNN), support vector machines (SVM), decision trees (DT), ensemble techniques, genetic programming.

W Yogi Yusuf [3] menggunakan algoritma decision tree yaitu C4.5, CART, dan CHAID yang dapat digunakan untuk status resiko kredit. Ketiga algoritma tersebut menghasilkan model tree yang berbeda untuk set data yang sama. Model yang berbeda dapat memberikan keakuratan yang berbeda pula. Algoritma C4.5 memberikan rata-rata tingkat keakuratan sebesar 87,72%, CART 87,27%, dan CHAID 87,15%.

2. METODE PENELITIAN

Metode penelitian pada penelitian ini adalah penelitian eksperimen dengan urutan sebagai berikut, masalah pada penelitian ini adalah penilaian rumah tinggal pengajuan kredit khususnya Rumah tinggal berupa tanah plus bangunan. Dari masalah di atas, setelah dilakukan studi literatur dari buku dan jurnal, ditemukan bahwa untuk penilaian rumah tinggal dapat menggunakan metode *data mining* berupa algoritma klasifikasi untuk meningkatkan akurasi. Beberapa algoritma klasifikasi: Mean-vector, K-Nearest Neighbor, ID3, C4.5, C5.0, CHAID dan CART. Dalam penelitian ini akan menerapkan algoritma C4.5 untuk penilaian rumah tinggal. Pengembangan yang Menerapkan Computing approach dalam Penelitian ini akan mengembangkan penilaian rumah tinggal dengan algoritma C4.5. Penerapan Software Rapidminer pada Obyek Penelitian ini akan diterapkan pada satu set data model simulasi berdasarkan data nyata yang di dapatkan dari lembaga penilai kredit. Selanjutnya evaluasi dan validasi hasil penerapan, Hasil pengolahan akan dibandingkan dan dianalisa hasilnya antara penilai yang menggunakan metode analisis kredit biasa dengan hasil prediksi dengan metode klasifikasi algoritma C4.5.

Kerangka Pemikiran



3. HASIL DAN PEMBAHASAN PENELITIAN

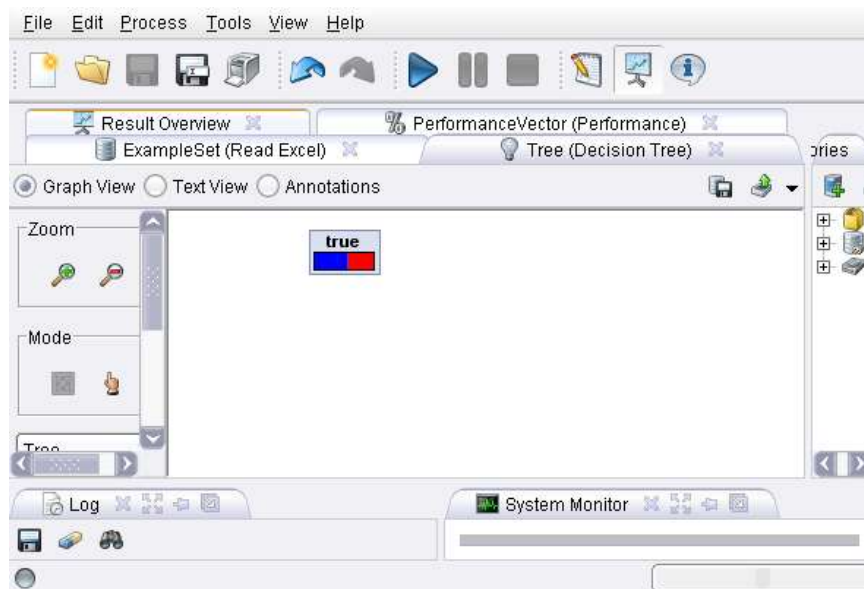
Pertama kali pengujian dilakukan dengan data yang belum diolah/dimodifikasi (gambar 1 dan 2). Ketika diimplementasi dengan rapidminer menghasilkan data:

Table View | Plot View

accuracy: 50.00% +/- 19.79% (mikro: 57.69%)

	true true	true false	class preci
pred. true	30	22	57.69%
pred. false	0	0	0.00%
class recall	100.00%	0.00%	

Gambar 1 Akurasi hasil C4.5 menggunakan data original



Gambar 2 Diagram pohon hasil C4.5 menggunakan data awal

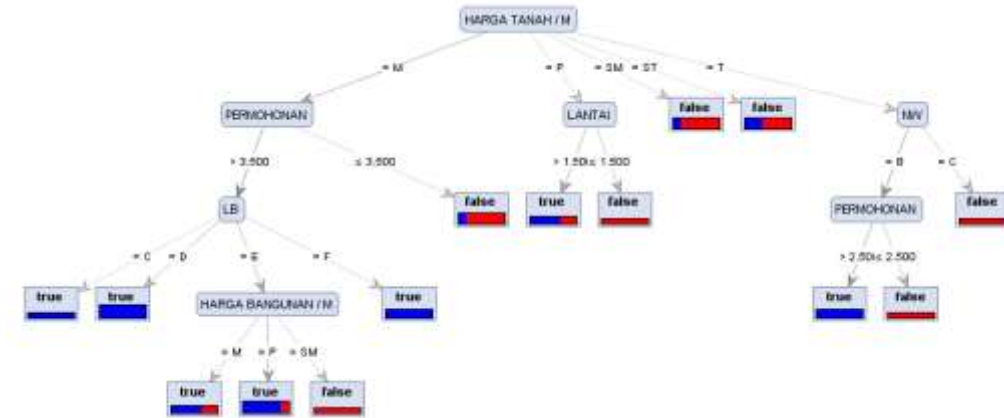
Karena hasil yang didapatkan belum sesuai, terutama pohon keputusan belum terbentuk, maka data kemudian diolah dengan membuat kategori-kategori:

1. Kategori wilayah (A, B, C)
2. Luas tanah (A, B, C, D, E)
3. Harga tanah (SM, M, P, T, ST)
4. Luas bangunan (A, B, C, D, E)
5. Harga Bangunan (SM, M, P, T, ST)

Ketika data ini diimplementasikan, hasilnya:

☒ Multiclass Classification Performance ☐ Annotations			
☒ Table View ☐ Plot View			
accuracy: 51.67% +/- 20.62% (mikro: 51.92%)			
	true true	true false	class precision
pred. true	19	14	57.58%
pred. false	11	8	42.11%
class recall	63.33%	36.36%	

Gambar 3 Akurasi hasil C4.5 menggunakan data Tabel 1



Gambar 4 Diagram pohon hasil C4.5 menggunakan data Tabel 1

Hasil akurasi memang menurun menjadi 51,67% tetapi, pohon keputusan terbentuk.

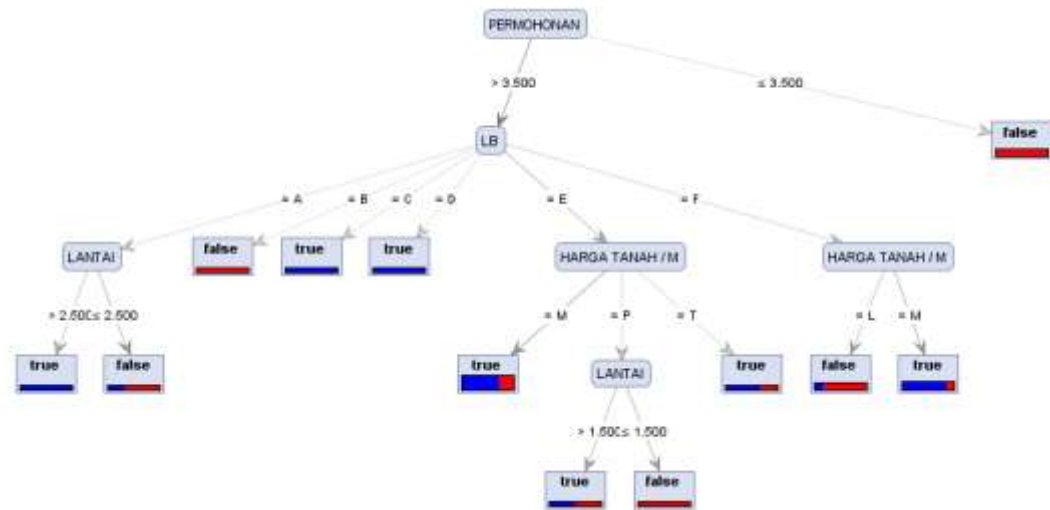
Dengan sedikit perbaikan, akurasi dapat ditingkatkan. Perbaikan yang dilakukan hanyalah membuat kategori untuk total harga tanah plus bangunan dan nilai permohonan, kemudian dikalikan dengan nilai wilayah,

Kategori Wilayah

A = 3, B = 2, C = 1

☒ Multiclass Classification Performance ☐ Annotations			
☒ Table View ☐ Plot View			
accuracy: 65.33% +/- 16.61% (mikro: 65.38%)			
	true true	true false	class precision
pred. true	25	13	65.79%
pred. false	5	9	64.29%
class recall	83.33%	40.91%	

Gambar 5 Akurasi hasil C4.5 menggunakan data Tabel 2



Gambar 6 Diagram pohon menggunakan data Tabel 2

B. Evaluasi dan Validasi Hasil

Hasil C4.5 dievaluasi dan dibandingkan dengan beberapa metode lain.

1. Random Tree

☒ Table View ☐ Plot View			
accuracy: 60.00% +/- 16.60% (mikro: 59.62%)			
	true true	true false	class preci
pred. true	27	18	60.00%
pred. false	3	4	57.14%
class recall	90.00%	18.18%	

Gambar 7 Akurasi Hasil Random Tree

2. k-NN

☒ Table View ☐ Plot View			
accuracy: 63.67% +/- 16.50% (mikro: 63.46%)			
	true true	true false	class preci
pred. true	23	12	65.71%
pred. false	7	10	58.82%
class recall	76.67%	45.45%	

Gambar 8 Akurasi Hasil k-NN

3. Naive Bayes

☒ Table View ☐ Plot View			
accuracy: 61.67% +/- 23.06% (mikro: 61.54%)			
	true true	true false	class
pred. true	20	10	66.67
pred. false	10	12	54.55
class recall	66.67%	54.55%	

Gambar 9 Akurasi Hasil Naive Bayes

4. C4.5

☒ Multiclass Classification Performance ☐ Annotations			
☒ Table View ☐ Plot View			
accuracy: 67.67% +/- 19.67% (mikro: 67.31%)			
	true true	true false	class pr
pred. true	27	14	65.85%
pred. false	3	8	72.73%
class recall	90.00%	36.36%	

Gambar 10 Akurasi hasil C4.5

Secara umum C4.5 lebih baik dibandingkan dengan k-NN, Random Tree, Naive Bayes dan CART, walaupun pada tingkat akurasi masih unggul hasil CART

4. KESIMPULAN DAN SARAN

Kesimpulan

- Decision Tree C4.5 dapat diterapkan dalam penilaian agunan kredit dengan akurasi 72,73%.
- Decision Tree C4.5 memprediksi lebih akurat dari pada k-NN, Random Tree, Naive Bayes dan CART, dalam penilaian agunan kredit, sehingga mampu memberikan solusi bagi kreditur, petugas penilai (estimator) dan dapat membantu pihak terkait, yang dalam hal ini adalah appraisal kredit dan perbankan dalam membuat penilaian agunan pengajuan kredit dengan hasil yang lebih baik.
- Proses klasifikasi dalam penelitian ini dilakukan dengan menentukan jumlah kategori yang terbentuk di awal proses sesuai dengan jumlah kelompok yang diinginkan. Dengan demikian, tidak dapat dipastikan

berapa sesungguhnya jumlah kategori ideal yang terbentuk dari data yang ada, sehingga akurasi hasil klasifikasi sulit ditingkatkan di atas 75%.

Saran

- a. Untuk kota Banjarmasin belum ada zonasi yang jelas, dan belum ada resume data pada pihak yang berwenang mengenai kisaran harga tanah pada wilayah kota, sehingga kategori yang diambil adalah persepsi umum dikalangan appraisal kredit. Guna menaikkan akurasi klasifikasi, perlu data zona wilayah ini selain standarisasi kategori.
- b. Keterbatasan lain dari penelitian ini adalah hanya uji coba untuk menilai tingkat akurasi penerapan decision tree C4.5. Agar penelitian ini dapat memberikan kontribusi yang lebih besar, perlu untuk pengembangan dengan cara menerapkan decision tree dalam suatu alat bantu (berupa software), sehingga dapat langsung diterapkan untuk penyelesaian masalah penilaian agunan dengan cepat dan mudah.
- c. Perlu pengembangan algoritma hybrid, yaitu dengan memanfaatkan algoritma lain selain keluarga decision tree, seperti memanfaatkan algoritma neural network.

Daftar Pustaka

- [1] Jozef Zurada and K. Niki Kunene, "Comparisons of the Performance of Computational Intelligence Methods for Loan Granting Decisions," in *Proceedings of the 44th Hawaii International Conference on System Sciences 2011*, Hawaii, 2011, pp. 1-10.
- [2] Mostafa Mahmoud and Ahmed Ali Najla Algadi, "Expert System for Banking Credit Decision," in *International Conference on Computer Science and Information Technology 2008*, 2008, pp. 813-819.
- [3] W Yogi Yusuf, "Perbandingan Performance Algoritma Decision Tree C5.0, CART, CHAID dan C4.5 : Kasus Prediksi Status Resiko Kredit Di Bank X," in *Seminar Nasional Aplikasi Teknologi Informasi (SNATI 2007)*, Yogyakarta, 2007.
- [4] Larose Daniel T, *Discovering knowledge in data: An Introduction to Data Mining.*: Wiley Interscience, 2005.
- [5] Susanto Sani and Suryadi Dedy, *Pengantar Data Mining: Menggali Pengetahuan Dari Bongkahan Data*. Yogyakarta: Penerbit Andi, 2010.

Halaman sengaja dikosongkan