

Penerapan Algoritma Decision Tree Untuk Penilaian Agunan Pengajuan Kredit

Budi Setiadi¹⁾, Bambang Lareno²⁾

¹Teknik Informatika, Universitas Islam Kalimantan MAAB, Banjarmasin

²Teknik Informatika, STMIK Indonesia Banjarmasin

e-mail: budibtc75@yahoo.com, blareno@gmail.com

Abstrak

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. Model yang berbeda dapat memberikan keakuratan yang berbeda pula. Karena itu penelitian ini bertujuan menerapkan algoritma decision tree C4.5 untuk penilaian agunan pengajuan kredit. Penelitian ini menggunakan data agunan pengajuan kredit di Kota Banjarmasin. Evaluasi kinerja algoritma menggunakan precision and recall dan AUC kemudian dibandingkan dan dianalisa hasilnya antara metode analisis lain (Naive Bayes, K-NN) dengan hasil prediksi dengan metode klasifikasi algoritma C4.5. Hasilnya, Decision Tree C4.5 dapat diterapkan dalam penilaian agunan kredit dengan akurasi 71% dan Nilai AUC di atas 0,6. Decision Tree C4.5 memprediksi lebih akurat dari pada k-NN, Naive Bayes dan Perhitungan biasa.

Kata kunci: algoritma klasifikasi, decision tree, C4.5, penilaian agunan kredit

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. Apabila prosedur telah berjalan dengan benar dan tidak terdapat rekayasa 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 kreditur. 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.

Proses ini dapat menggunakan *datamining*. Tujuannya agar dapat diketahui pola-pola universal dari data yang ada[1]. Faktor-faktor yang dapat mempengaruhi akan dianalisa dan di klasifikasi untuk kemudian di prediksi sehingga membentuk kelompok-kelompok yang memiliki kesamaan tersembunyi dari data yang telah dianalisa[2]. Klasifikasi data yang dapat digunakan untuk mengetahui tingkat hubungan antar data tersebut adalah dengan metode klasifikasi *decision tree*[3].

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. Salah satu pendekatan untuk menetapkan nilai akurasi pada pohon keputusan tersebut adalah algoritma C.45[4][5]

Josef Zurada dan K. Niki Kunene [6] menegaskan bahwa penting bagi institusi keuangan untuk secara akurat mengevaluasi risiko kredit yang mereka timbulkan tiap kali memutuskan sebuah pemberian pinjaman. 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* (kNN), *classification trees* (CT), *neural networks* (NN), *radial basis function neural networks* (RBFNN), *support vector machines* (SVM), *decision trees* (DT), *ensemble techniques*, dan *genetic programming*. Mostafa Mahmoud and Ahmed Ali Najla Algadi[7] menemukan bahwa sistem pakar berkinerja sangat baik untuk masalah kompleks dan tidak terstruktur jika dibandingkan dengan pendekatan statistik yang lebih tradisional. Dibiidang lain, Ashwani Rao[8] mengekstrak informasi dari 3.316 data bayi dari usia 0-1

bulan yang membentuk pola data untuk di klasifikasi dan membandingkan algoritma pembelajaran dalam hal estimasi yang lebih baik di percaya untuk prediksi data pemberian makanan bayi dari usia 0-1 bulan, pendekatan yang digunakan adalah dengan menggunakan algoritma C.45, *Naive Bayes Multinomial Classifier*, *Logistic Model Trees*, *Regularized Logistic Regression*.

Dari latar di atas, disimpulkan bahwa untuk penilaian agunan kredit dapat menggunakan metode *data mining* berupa algoritma klasifikasi *decision tree*. Sehingga walaupun terkadang data tidak lengkap, keputusan masih dapat diambil akurasi penilaian yang cukup. Dalam penelitian ini akan menerapkan algoritma C4.5 untuk penilaian agunan. Sehingga pertanyaan penelitian adalah: “Bagaimana mengevaluasi algoritma *decision tree* untuk penilaian agunan pengajuan kredit sehingga diketahui mana yang lebih akurat?”

2. Metode Penelitian

Penelitian ini memakai data metode penelitian eksperimen komparasi, yang terdiri: (1)Metode Pengumpulan data dan pengolahan data awal, (2)Metode yang diusulkan, (3)Eksperimen dan pengujian model, (4)Hasil eksperimen dan (5)Evaluasi dan validasi hasil.

2.1. Pengumpulan Data

Penelitian ini memakai data agunan pengajuan kredit yang didapatkan dari lembaga penilai kredit Kota Banjarmasin. Data yang dibutuhkan dalam penelitian ini adalah:

- Data Sekunder
Satu set data model simulasi berdasarkan data nyata yang di dapatkan dari lembaga penilai kredit.
- Data Primer
Data primer yang digunakan adalah data hasil komputasi masing-masing algoritma.

2.2. Pengolahan Data Awal

Data yang didapatkan dari instansi terkait masih berupa data yang terdiri dari berbagai parameter, sehingga harus olah terlebih dahulu. Modifikasi tersebut dilakukan dengan memperhatikan kebutuhan. Dalam hal ini, dilakukan perubahan dan kategori data untuk mendapatkan hasil yang lebih baik. Data diolah dengan membuat kategori-kategori:

1. Kategori wilayah (A, B, C)
Untuk nilai ini, kategori yang diambil adalah persepsi umum dikalangan appraisal kredit.
2. Luas tanah (A, B, C, D, E)
 $A = \text{Luas} > 275 \text{ m}^2$; $B = \text{Luas} > 200 \text{ m}^2$; $C = \text{Luas} > 125 \text{ m}^2$; $D = \text{Luasan sampai dengan } 125 \text{ m}^2$
3. Harga tanah (SM, M, P, T, ST)
 $ST = \text{Harga} > 2.000.000/\text{m}^2$; $T = \text{Harga} > 1.500.000/\text{m}^2$; $P = \text{Harga} > 1.000.000/\text{m}^2$
 $M = \text{Harga} > 500.000/\text{m}^2$; $SM = \text{Harga sampai dengan } 500.000/\text{m}^2$
4. Luas bangunan (A, B, C, D, E)
 $A = \text{Luasan} > 250 \text{ m}^2$; $B = \text{Luasan} > 200 \text{ m}^2$; $C = \text{Luasan} > 150 \text{ m}^2$; $D = \text{Luasan} > 125 \text{ m}^2$
 $E = \text{Luasan} > 75 \text{ m}^2$; $F = \text{Luasan sampai dengan } 75 \text{ m}^2$
5. Harga Bangunan (SM, M, P, T, ST)
 $ST = \text{Harga} > 2.500.000/\text{m}^2$; $T = \text{Harga} > 2.000.000/\text{m}^2$; $P = \text{Harga} > 1.500.000/\text{m}^2$
 $M = \text{Harga} > 1.000.000/\text{m}^2$; $SM = \text{Harga sampai dengan } 1.000.000/\text{m}^2$

2.3. Metode/Model yang diusulkan

Metode yang digunakan adalah perbandingan antara akurasi yang dihasilkan oleh Algoritma C4.5, K-Nearest Neighbor dan Naïve Bayes.

2.3.1. Algoritma C4.5

Sering disebut dengan pohon keputusan (*decision tree*). Mirip sebuah struktur pohon dimana terdapat node internal yang mendeskripsikan atribut-atribut. Setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas[3]. Algoritma C4.5 adalah pengembangan dari algoritma ID3[9]. karena dari pengembangan inilah algoritma C4.5 mempunyai prinsip dasar kerja yang sama dengan algoritma ID3[10].

2.3.2. Nearest Neighbor

Algoritma k-nearest neighbor (k-NN atau KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Teknik ini sangat sederhana dan mudah diimplementasikan. Mirip dengan teknik klustering,

pengelompokkan suatu data baru berdasarkan jarak data baru itu ke beberapa data/tetangga (*neighbor*) terdekat[11].

2.3.3. Naïve Bayes

Algoritma Bayesian adalah sebuah algoritma pengklasifikasi secara statistik. Mereka dapat memprediksi probabilitas keanggotaan kelas, seperti kemungkinan bahwa suatu tupel diberikan milik suatu kelas tertentu[10]

2.4. Eksperimen dan Pengujian Model/Metode

Data yang telah dimodifikasi diproses dengan metode yang telah dipilih, yang dalam hal ini adalah C4.5, kNN dan Naïve Bayes. Algoritma akan diimplementasikan dengan rapidminer, yang kemudian akan menghasilkan nilai akurasi berdasarkan metode *Precision and Recall* dan AUC (*area under curve*) [12]. Pengukuran penelitian dilakukan dengan membandingkan hasil prediksi yang dilakukan pengolahan data dengan metode biasa dan dilakukan simulasi dengan metode algoritma C4.5, kNN dan Naïve Bayes.

2.5. Evaluasi dan Validasi

Evaluasi menggunakan hasil *Precision and Recall* dan AUC, kemudian dibandingkan hasilnya antara penilai yang menggunakan metode analisis lain (Naive Bayes, K-NN) dengan hasil prediksi dengan metode klasifikasi algoritma C4.5. Dari sini akan terlihat tingkat akurasi dari penerapan C4.5

3. Hasil dan Pembahasan

3.1. Hasil Pengujian Model/Metode

Pengujian menggunakan data dengan atribut yang telah di kategori, sebagaimana terlihat pada Tabel 1.

Tabel 1. Data Nilai Agunan

LOKASI	NW	SERTIFIKAT	LANTAI	LT	HARGA TANAH / M	HARGA TANAH	LB	HARGA BANGUNAN / M	HARGA BANGUNAN
kayu tangi	C	SHM	1	D	M	DM	E	P	EP
veteran	C	SHGB	2	B	M	BM	C	M	CM
kayu tangi	C	SHM	1	C	M	CM	E	P	EP
pondok metro	B	SHM	1	B	T	BT	C	P	CP
pondok metro	B	SHM	1	B	T	BT	D	T	DT
gatot subroto	B	SHM	1	C	T	CT	D	M	DM
gatot subroto	B	SHM	1	C	T	CT	D	M	DM
komplek benyamin	B	SHM	2	B	M	BM	B	P	BP
sultan adam	B	SHBG	3	C	ST	CST	A	T	AT
komplek mandiri	C	SHM	1	B	M	BM	D	M	DM
komplek tekwondo	C	SHM	1	C	M	CM	E	M	EM
pasir mas	C	SHM	1	C	M	CM	E	SM	ESM
sutoyo S	C	SHM	1	C	M	CM	D	M	DM
A. Yani	A	SHGB	3	C	ST	CST	A	P	AP
sungai andai	C	SHM	1	C	M	CM	F	SM	FSM
A. Yani	A	SHGB	2	C	ST	CST	C	P	CP
pramuka	C	SHM	1	B	M	BM	C	P	CP
komplek benyamin	B	SHM	1	A	M	AM	B	P	BP
komplek benyamin	B	SHM	2	A	M	AM	A	P	AP
sungai andai	C	SHM	1	C	M	CM	F	SM	FSM
handil bakti	C	SHM	1	C	SM	CSM	F	SM	FSM
handil bakti	C	SHM	1	C	SM	CSM	F	SM	FSM
Komplek DPR	C	SHM	1	A	M	AM	D	M	DM
HKSN	C	SHM	1	C	M	CM	E	M	EM
Komplek AMD	B	SHM	1	C	SM	CSM	F	M	FM
Komplek junjung buih	B	SHM	1	C	M	CM	F	M	FM
ratu jaleha	B	SHM	1	C	M	CM	D	SM	DSM
Cemara Raya	C	SHM	1	B	T	BT	E	P	EP
HKSN	C	SHM	1	B	M	BM	E	P	EP
Citra Garden	A	SHM	2	C	P	CP	D	ST	DST
Citra Garden	A	SHM	2	C	P	CP	D	ST	DST
hasan basri	C	SHGB	2	C	ST	CST	C	P	CP
Beruntung jaya	B	SHM	1	C	P	CP	D	P	DP
Sungai Lutut	C	SHM	1	C	SM	CSM	F	SM	FSM
handil bakti	C	SHM	1	C	SM	CSM	F	SM	FSM
Cempaka	B	SHM	1	A	T	AT	B	P	BP
Teluk Dalam	B	SHM	1	C	M	CM	D	M	DM
Mawar	A	SHM	1	C	P	CP	D	P	DP
pasar lama	C	SHM	1	C	M	CM	F	SM	FSM
komplek wildan	C	SHM	2	B	M	BM	C	M	CM
Cemara Raya	C	SHM	1	C	T	CT	E	P	EP
komplek tekwondo	C	SHM	1	B	M	BM	E	M	EM
komplek mandiri permai	C	SHM	1	C	M	CM	E	P	EP
Beruntung jaya	B	SHM	1	B	T	BT	D	P	DP
Landasan Ulin	C	SHM	1	C	M	CM	E	SM	ESM
Citra Garden	A	SHM	2	C	P	CP	D	ST	DST
pramuka	C	SHM	1	C	M	CM	D	P	DP
komplek benyamin	B	SHM	1	B	M	BM	E	P	EP
veteran	C	SHM	1	B	M	BM	E	M	EM
kayu tangi ujung	C	SHM	1	C	M	CM	E	P	EP
mahoni	C	SHM	2	D	M	DM	D	P	DP
S. Parman	B	SHGB	3	C	ST	CST	A	P	AP

3.1.1. C4.5

Ketika data ini pada tabel 1 diimplementasikan dengan algoritma C4.5, menghasilkan akurasi 71.00%. Pohon yang terbentuk terlihat pada gambar 2.

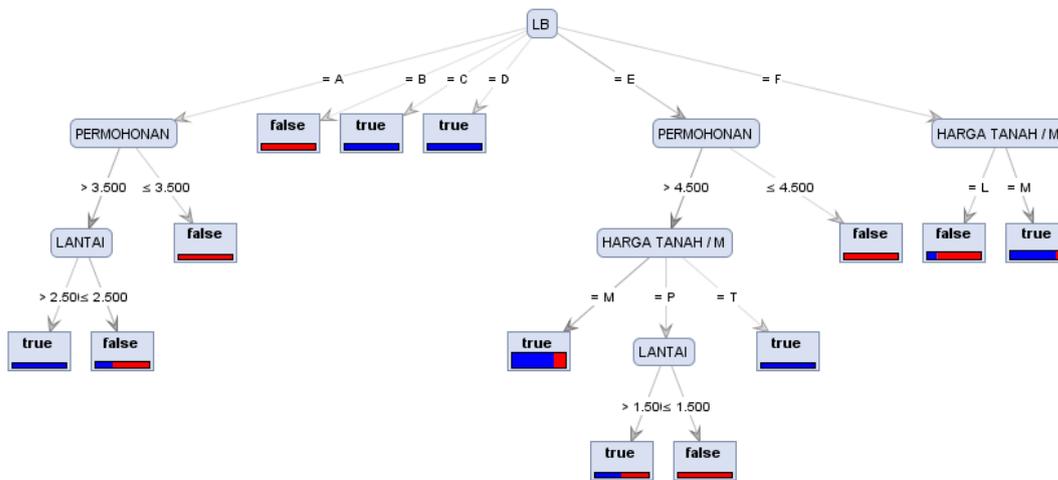
Multiclass Classification Performance
 Annotations

Table View
 Plot View

accuracy: 71.00% +/- 19.21% (mikro: 71.15%)

	true true	true false	class precision
pred. true	25	10	71.43%
pred. false	5	12	70.59%
class recall	83.33%	54.55%	

Gambar 1. Akurasi Hasil C4.5



Gambar 2. Diagram pohon hasil C4.5

3.1.2. k-NN

Ketika data pada tabel 1 diimplementasikan menggunakan algoritma k-NN, akurasi yang dihasilkan 63,67%.

Table View
 Plot View

accuracy: 63.67% +/- 16.50% (mikro: 63.46%)

	true true	true false	class precision
pred. true	23	12	65.71%
pred. false	7	10	58.82%
class recall	76.67%	45.45%	

Gambar 3. Akurasi Hasil k-NN

3.1.3. Naïve Bayes

Ketika data pada tabel 1 diimplementasikan menggunakan algoritma Naïve Bayes, akurasi yang dihasilkan 61,67%.

Table View
 Plot View

accuracy: 61.67% +/- 23.06% (mikro: 61.54%)

	true true	true false	class precision
pred. true	20	10	66.67%
pred. false	10	12	54.55%
class recall	66.67%	54.55%	

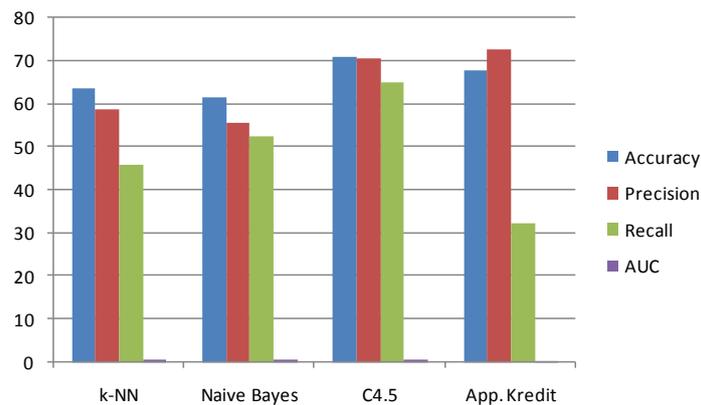
Gambar 4. Akurasi Hasil Naive Bayes

3.2. Evaluasi dan Validasi Hasil

Pada tabel 2, secara umum C4.5 lebih baik dibandingkan dengan k-NN, Naive Bayes dan Perhitungan Biasa. Perhitungan biasa unggul pada tingkat presisi. Hasil perbandingan secara grafik ditampilkan pada gambar 5. Nilai AUC C4.5 berada di atas 0,6, sehingga masuk dalam kategori layak klasifikasi.

Tabel 2. Validasi Hasil Akurasi CART

	k-NN	Naive Bayes	C4.5	App. Kredit
Accuracy	63,67	61,67	71	67,67
Precision	58,82	55,5	70,59	72,73
Recall	45,83	52,5	65	32,5
AUC	0,585	0,575	0,65	0,517



Gambar 5. Perbandingan Nilai Akurasi, Presisi dan Recall, AUC.

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%.

Selain itu, Kota Banjarmasin belum memiliki 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.

4. Simpulan

Dari hasil penelitian yang dilakukan dari tahap awal hingga pengujian, dan hasil perbandingan, disimpulkan bahwa *Decision Tree* C4.5 dapat diterapkan dalam penilaian agunan kredit dengan akurasi 71% dengan nilai AUC di atas 0,6. *Decision Tree* C4.5 memprediksi lebih akurat dari pada k-NN, Naive Bayes dan Perhitungan biasa, 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.

Berapa hal perlu disampaikan untuk pengujian lanjutan dan pencapaian tingkat akurasi yang lebih baik, yaitu bahwa zonasi kota yang lebih jelas sangat diperlukan, khususnya berkait dengan rencana pengembangan wilayah dan tata kota. Selain itu perlu pengembangan algoritma sehingga jumlah kategori untuk setiap atribut dapat ditentukan dengan optimal.

Daftar Pustaka

- [1] Witten I.H, Frank E. Data Mining: Practical Machine Learning Tools and Techniques. Second Edition. San Francisco: Morgan Kauffman; 2005.
- [2] Wu X, Kumar V. The Top Ten Algorithm In Data Mining. London: Taylor & Francis Group; 2009.
- [3] Han J, Kamber M. Data Mining: Concepts and Techniques. Second Edition. San Fransisco: Morgan

-
- Kaufmann; 2007.
- [4] Gamulina J, Gamulinb O, Kermekc D. *Comparing classification models in the final exam performance prediction*. In MIPRO; 2014; Croatia.
 - [5] Johansson U, Boström H, Löfström T. *Conformal Prediction Using Decision Trees*. IEEE 2013. 2013; 1550-4786(13)
 - [6] Zurada J, Kukune N.K. *Comparisons of the Performance of Computational Intelligence Methods for Loan Granting Decisions*. IEEE2011. 2011.
 - [7] Mahmoud M, Najla Algadi AA. *Expert System for Banking Credit Decision*. In International Conference on Computer Science and Information Technology 2008; 2008. p. 813-819.
 - [8] Rao A, Maiden K, Carterette B, Ehrenthal D. *Predicting Baby Feeding Method from Unstructured Electronic Health Record Data*. ACM. 2012; 978-1-4503-1716-0/12/10.
 - [9] Larose D.T. *Discovering Knowledge In Data: An Introduction to Data Mining*. New Jersey: Wiley; 2005.
 - [10] K, Luthfi ET. *Algoritma Data Mining*. Yogyakarta: Andi; 2009.
 - [11] Santosa B. *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu; 2007.
 - [12] Fayyad UM. *Data Mining and Knowledge Discovery: Making Sense out of Data*. IEEE Intelligent Systems. 1996; Vol. 11, pp. 22-23.