

KOMPARASI PENERAPAN ALGORITMA C45, KNN DAN NEURAL NETWORK DALAM PROSES KELAYAKAN PENERIMAAN KREDIT KENDARAAN BERMOTOR

PUJI ASTUTI

Poetie12@gmail.com

Program Studi Teknik Informatika Fakultas Teknik, Matematika dan Ilmu Pengetahuan Alam
Universitas Indraprasta PGRI

Abstrak. Dalam perkembangan bisnis, masalah perkreditan tetap menarik untuk dikaji dan diungkap. Kebanyakan masalah terjadi tidak pada system yang diterapkan oleh bank tetapi masalah terjadi justru pada SDM yang mengelola kredit ,baik pada hubungan nya dengan konsumen atau pada kesalahan di pihak leasing yang salah prediksi dalam menilai konsumen yang mengajukan kredit. Beberapa penelitian bidang komputer telah banyak dilakukan untuk mengurangi resiko kredit yang menyebabkan kerugian pada perusahaan. Dalam penelitian ini dilakukan komparasi algoritma *C4.5*, *KNN* dan *neural network* yang diaplikasikan terhadap data konsumen yang mendapat kelayakan penerimaan kredit bermotor baik yang bermasalah dalam pembayaran angsurannya maupun tidak. Metode saat ini belum dapat menentukan data mining yang tepat. Proses penghitungan ke tiga algoritma dan di tambah dengan program rapid miner dapat menghasilkan data yang akurat dan berguna bagi semua khususnya pihak *bess finance* untuk lebih mempermudah system dalam hal penentuan penerimaan kredit kendaraan bermotor.dan hasil yang di peroleh ternyata algoritma *C45* lebih akurat di bandingkan dengan kedua algoritma lainnya.

Kata kunci: *C4.5*, *KNN*, *neural network*, *Rapid Miner*, *Data Mining*

Abstract. In the development of business,credit problems remain to be studied to reveal interesting. Most problems the system imposed by the bank but the problem occurs precisely the human resources to manage credit, either on its relationship with the consumer or the mistake in leasing the wrong predictions in assessing consumers who apply for credit. Some computers have a lot of official research conducted to reduce the credit risk of causing harm to the company. In this study a comparison algorithm C4.5, KNN and the neural network which is applied to the data consumer who gets the credit worthiness of motor good reception is problematic in the installment payment or not. The current method has not been able to determine the appropriate data mining. The process of counting to three algorithms and programs added with rapid miner can produce data that is accurate and useful for all parties especially bess finance to further simplify the system in terms of determining the credit acceptance of vehicles. Results obtained C45 turns algorithm is more accurate in comparison with other algorithms.

Keywords: *C4.5*, *KNN*, *neural network*, *Rapid Miner*, *Data Mining*

PENDAHULUAN

Evaluasi rasio kredit merupakan masalah yang menarik dalam analisa keuangan .mengenai analisis kelayakan pemberian kredit untuk konsumen khususnya penerimaan kredit kendaraan bermotor dengan metode klasifikasi data mining telah banyak dilakukan,Dalam ini akan di

bahas mengenai penerapan algoritma c45. Metode yang dipakai saat ini belum dapat menentukan data mining yang tepat dalam proses kelayakan penerimaan kredit bermotor. Dalam pengolahan datanya digunakan metode Data mining dimana data mining itu sendiri adalah proses menelusuri pengetahuan yang baru, pola, dan tren yang dipilah dari jumlah data yang besar yang disimpan dalam repositori atau tempat penyimpanan dengan menggunakan tehnik pengenalan pola serta statistik dan teknik matematika menurut Gartner group dalam (Vercellis, 2009) Jadi sangat bermanfaat dalam proses pengolahan datanya. Ditambah dengan program rapid miner untuk memasukkan hasil data yang akan diproses sehingga dapat lebih memudahkan pihak lesing.

Dalam ini dibahas tentang komparasi tiga algoritma yaitu K-nearest neighbor, C 45 dan neural network. K-nearest neighbor adalah klasifikasi yang menyimpan semua data *training* dan melakukan klasifikasi dengan cara membandingkan antara atribut data baru yang paling cocok dengan atribut *record* yang terdapat pada data *training*, selain itu dibandingkan juga dengan algoritma C4.5, dimana algoritma ini adalah pohon keputusan, yang berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variable input dengan sebuah variable target. Dan selanjutnya dengan neural network yang Sejak tahun 1950-an, neural network telah digunakan untuk tujuan prediksi, bukan hanya klasifikasi tapi juga regresi dengan atribut target continue (Hariani, 2010).

Secara rinci identifikasi masalah yang akan dikemukakan dalam ini adalah:

- a. Memilih satu dari tiga buah metode algoritma yaitu algoritma k-nearest neighbor ,C4.5, dan neural network,yang paling baik dalam memilih calon penerima kelayakan penerimaan kredit bermotor.
- b. Dengan kajian komparasi algoritma ini diharapkan dapat membantu pihak lesing dalam menentukan penerima kredit bermotor tersebut.

TINJAUAN PUSTAKA

Kredit

Istilah kredit berasal dari perkataan latin *credo*, yang berarti *I believe, I trust*, saya percaya atau saya menaruh kepercayaan (Rivai, 2006) Kredit adalah penyerahan barang, jasa, atau uang dari satu pihak (kreditor/pemberi pinjaman) atas dasar kepercayaan kepada pihak lain (nasabah atau penguatang/*borrower*) dengan janji membayar dari penerima kredit kepada pemberi kredit pada tanggal yang telah disepakati kedua belah pihak.

Memperoleh kredit berarti memperoleh kepercayaan. Atas dasar kepercayaan kepada seseorang yang memerlukannya maka di berikan uang, barang atau jasa dengan syarat membayar kembali atau memberikan penggantinya dalam satu jangka waktu yang telah diperjanjikan (Linof ,2011).

Data Mining

Ada beberapa sumber yang mengartikan data mining dengan berbagai macam pengertian diantaranya adalah:

1. Menurut Linof dan Berry (David olson ,2008) Data mining adalah proses bisnis untuk menjelajahi sejumlah besar data untuk menemukan pola yang bermakna dan beraturan.
2. Olson dan Delen (Triantaphyllou ,20010) menyatakan bahwa Data mining disebut eksplorasi analisis data. Data yang dihasilkan dari cash registers, scanning, dari data base topik tertentu di perusahaan, dieksplorasi, dianalisis, dikurangi, dan digunakan kembali.

3. Triantaphyllou (Larose ,2005) menyatakan bahwa Data mining dan *Knowledge discovery* adalah bagian dari metode komputasi yang bertujuan untuk mengumpulkan dan menganalisis data yang terkait dengan fungsi dari sebuah sistem kepentingan untuk tujuan memperoleh pemahaman yang lebih baik.

Algoritma C 4.5

Algoritma C 4.5 adalah suatu deretan algoritma untuk permasalahan klasifikasi didalam sebuah mesin dan himpunan data. Dengan nilai data yang bervariasi , dimana kejadian diuraikan oleh koleksi atribut dan mempunyai salah satu dari satu set kelas yang eksklusif, algoritma C4.5 yaitu suatu pemetaan dari atribut menilai ke kelas yang dapat diberlakukan bagi menggolongkan kejadian tak terlihat baru. (Wu ,2009).

Neural network

Neural network adalah satu set unit input/output yang terhubung dimana tiap relasinya memiliki bobot. Neural network dimaksudkan untuk mensimulasikan perilaku system biologi susunan syaraf manusia, yang terdiri dari sejumlah besar unit pemroses yang disebut neuron, yang beroperasi secara parallel (Alpayadin ,2010) Neuron mempunyai relasi dengan synapse yang mengelilingi neuron-neuron lainnya. Susunan syaraf tersebut dipresentasikan dalam neural network berupa graf yang terdiri dari simpul (neuron) yang dihubungkan dengan busur, yang berkorespondensi dengan synapse. Sejak tahun 1950-an, neural network telah digunakan untuk tujuan prediksi, bukan hanya klasifikasi tapi juga regresi dengan atribut target continue (Hariani ,2010).

Backpropagation (Kusrini, 2009) bekerja melalui proses secara iterative menggunakan data *training*, membandingkan nilai prediksi dari jaringan dengan setiap data yang terdapat pada data *training*. Dalam setiap proses, bobot relasi dalam jaringan dimodifikasi untuk meminimalkan nilai *Mean Squared Error* (MSE) antara nilai prediksi dari *network* dengan nilai sesungguhnya. Modifikasi relasi *neural network* tersebut dilakukan dengan arah mundur, dari *output layer* pertama dari *hidden layer* sehingga algoritma ini disebut *backpropagation*.

METODE

Jenis Penelitian

- a. Penelitian Eksperimental
Penelitian eksperimental merupakan penelitian yang bersifat uji coba, memanipulasi dan mempengaruhi hal-hal yang terkait dengan seluruh variabel atau atribut.
- b. Penelitian Perbandingan atau studi komparasi yakni dengan membandingkan antara tiga macam algoritma yaitu algoritma C4.5, KNN dan Neural Network.

Populasi

Populasi dalam penelitian ini merupakan konsumen pada *bess finance* tahun 2012/2013. Dengan data konsumen sebanyak 486 record data konsumen , yang terdiri dari 14 atribut. Dimana 13 atribut *predictor* dan 1 atribut hasil. Responden atau sample dalam penelitian ini di ambil dari pihak-pihak yang terkait dalam kelayakan penerimaan kredit kendaraan bermotor

Analisis Data

Analisis data dalam penelitian ini menggunakan Data Kuantitatif berupa kaidah-kaidah matematika terhadap data numerik. Yang dilakukan pengujian pada masing-masing algoritma yaitu algoritma C4.5, algoritma KNN dan Neural Network.

HASIL DAN PEMBAHASAN

Algoritma C4.5

Data *training* pada tabel 4 adalah untuk menentukan apakah seorang nasabah bermasalah atau tidak. Berikut akan dibahas prediksi apakah nasabah bermasalah atau tidak, menggunakan metode klasifikasi.

Langkah untuk membuat pohon keputusan, yaitu :

1. Tabel 3.3 adalah data *training* beserta kelasnya
2. Hitung nilai *entropy*. Dari data training diketahui jumlah kasus ada 486, konsumen yang termasuk kelas *good* 309 *record* dan *Bad* 177 *record* sehingga didapat *entropy*:

$$Entropy(S) = \sum_{i=1}^n -p_i \cdot \log_2 p_i$$

$$= (-309/486 \cdot \log_2 (309/486)) + (-177/486 \cdot \log_2 (177/486))$$
$$= 0.9461$$

3. Hitung nilai *entropy* per atribut terlebih dahulu dengan rumus sama dengan di atas

1. Status perkawinan

a. Menikah

$$Entropy(S) = \sum_{i=1}^n -p_i \cdot \log_2 p_i$$

$$= (-93/334 \cdot \log_2 (93/334)) + (-241/334 \cdot \log_2 (241/334))$$
$$= 0.8533$$

2. Jumlah Tanggungan

a. Jumlah tanggungan 1

$$Entropy(S) = \sum_{i=1}^n -p_i \cdot \log_2 p_i$$

$$= (-19/45 \cdot \log_2 (19/45)) + (-26/45 \cdot \log_2 (26/45))$$
$$= 0.9825$$

dst

3. Hitung nilai *gain* untuk tiap atribut, lalu tentukan nilai *gain* tertinggi. Yang mempunyai nilai *gain* tertinggi itulah yang akan dijadikan akar dari pohon. Misalkan untuk atribut status perkawinan = menikah, didapat nilai *gain*:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

a. Perhitungan Gain status pernikahan

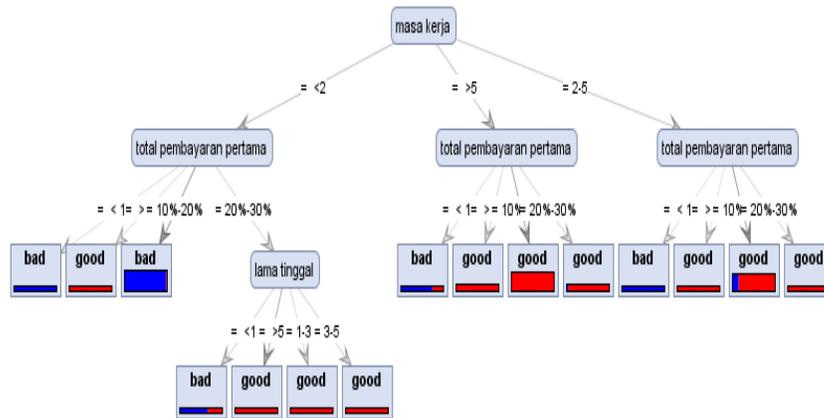
$$Gain(S, A) = 0.9461 - (334/486(0.8533)) + 128/486(0.9700) + 24/486(0.8709)$$
$$= 0.0612$$

Perhitungan *entropy* dan *gain* untuk semua atribut dilakukan, untuk mendapatkan nilai *gain* tertinggi. Hasil perhitungan seluruh atribut terlihat pada Tabel 1.

Tabel 1. Nilai entropy dan gain untuk menentukan simpul akar

Atribut	kasus	good	bad	entropy	gain
status perkawinan	486	309	177	0.9461	0.0612
Menikah	334	241	93	0.8533	
belum menikah	128	51	77	0.9700	
janda/duda	24	17	7	0.8709	
jumlah tanggungan	486	309	177		0.0813
1 orang	45	26	19	0.9825	
2-3 orang	224	168	56	0.8113	
> 3 orang	69	55	14	0.7277	
tidak ada	148	60	88	0.9740	
pendidikan terakhir	486	309	177		0.0346
>S1	4	4	0	0	
S1	23	18	5	0.7554	
Diploma	21	16	5	0.7919	
SLTA	217	123	94	0.9871	
SLTP	103	58	45	0.9885	
SD	105	81	24	0.7755	
tidak sekolah	13	9	4	0.8905	
Usia	486	309	177		0.1680
< 21 tahun atau > 60 tahun	64	23	41	0	
21 - 55 tahun	414	278	136	0.9134	
55 - 60 tahun	8	8	0	0	
kepemilikan rumah	486	309	177		0.0828
milik sendiri (PBB/sertifikat/AJB/rek listrik)	141	106	35	0.8085	
milik sendiri (PBB a/n orang lain)	43	33	10	0.7824	
KPR	41	34	7	0.6594	
orang tua	187	85	102	0.9940	
Keluarga	26	13	13	1.0000	
Dinas	3	3	0	0	
sewa/kontrak >= tenor	22	19	3	0.5746	
sewa/kontrak < tenor	23	16	7	0.8865	
Kost	0	0	0	0	
lama tinggal	486	309	177		0.0171
> 5 tahun	346	213	133	0.9611	
3 - 5 tahun	68	43	25	0.9488	
1 - 3 tahun	49	41	8	0.6421	
< 1 tahun	23	12	11	0.9986	
kondisi rumah	486	309	177		0.0138

Dari hasil perhitungan *entropy* dan *gain* yang didapat pada Tabel 1, terlihat bahwa atribut masa kerja mempunyai nilai gain tertinggi yaitu 0.4994. Oleh karena itu maka masa kerja merupakan simpul akar pada pohon keputusan. Untuk menentukan simpul berikutnya, yaitu simpul 1.1 dilakukan lagi perhitungan *entropy* dan *gain* berdasarkan atribut masa kerja. Jumlah kasus yang dihitung adalah sejumlah kasus dengan nilai dari simpul akar (masa kerja). Penghitungan tidak dilakukan secara keseluruhan, namun hasil dari pohon keputusan akan terlihat setelah data konsumen di aplikasikan pada rapid miner dan akan terlihat pada gambar 2



Gambar 1. Pohon Keputusan hasil perhitungan dengan metode C4.5

Algoritma K-NN

Dalam algoritma KNN perhitungan dilakukan dengan mendefinisikan bobot kedekatan untuk semua jumlah data, yaitu bobot antara satu atribut dengan atribut lain dan bobot antara nilai nilai dalam atribut. jika terdapat kasus baru dan dari kasus baru tersebut termasuk dalam hal bermasalah atau tidak, maka dilakukan perhitungan kedekatan kasus baru dengan kasus sebelumnya. Untuk mengukur jarak antara atribut akan kita beri nilai antara 0 sampai dengan 1. nilai 0 artinya jika atribut tidak berpengaruh dan sebaliknya nilai 1 jika

Table 2. Pembobotan atribut

NO	Atribut	Bobot
1	Status Perkawinan	0.5
2	Jumlah tanggungan	1
3	Pendidikan terakhir	0.5
4	Usia	0.5
5	Keperilikan rumah	0.8
6	lama tinggal	1
7	Kondisi rumah	0.6
8	Jenis Pekerjaan	0.8
9	Status Perusahaan	0.5
10	Status Kepegawaian	0.8
11	Masa kerja	1
12	Penghasilan perbulan	1
13	Pembayaran pertama	1

Penentuan kedekatan antar nilai atribut akan kita hitung di mulai pada atribut status perkawinan, yang terdiri dari tiga nilai kategori,yaitu menikah, belum menikah, dan janda/duda.

Table 3. Kedekatan nilai atribut status perkawinan

Atribut	Nilai atribut 1	Nilai Atribut 2	Bobot
Status Perkawinan	Menikah	Menikah	0
	Menikah	Belum Menikah	1
	Menikah	janda/duda	0.5
	belum menikah	Belum Menikah	0
	belum menikah	janda/duda	0.5
	janda/duda	janda/duda	0

Pembobotan nilai atribut dilakukan untuk ke semua atribut,langkah berikutnya adalah menghitung kemiripan.

Table 4 berisi sampel data training yang merupakan kasus lama dan akan di ukur kedekatannya dengan kasus yang baru

Table 4. Sampel data training

No	status perkawinan	jumlah tanggungan	pendidikan terakhir	usia	kepemilikan rumah
1	Menikah	tdk ada	SLTA	<21/>60	Ortu
2	belum menikah	tdk ada	SLTA	21-55	Ortu

Table 5. Lanjutan

lama tinggal	kondisi rumah	jenis pekerjaan	status perusahaan	status kepegawaian	masa kerja
3_5	Permanen	karyawan	swasta menengah	Kontrak	<2
3_5	Permanen	karyawan	swasta menengah	Tetap	>5

Table 6. Lanjutan

penghasilan perbulan	total pembayaran pertama	remark
> 2x ang slip	10-20%	bad
> 3x ang slip	10-20%	good

Table 7. Sampel data testing

No	status perkawinan	jumlah tanggungan	pendidikan terakhir	usia	kepemilikan rumah
1	belum menikah	tdk ada	SLTP	<21/>60	Ortu

Table 8. Lanjutan

lama tinggal	kondisi rumah	jenis pekerjaan	status perusahaan	status kepegawaian	masa kerja
>5	permanen	Karyawan	Swasta Kecil	Kontrak	<2

Tabel 9. Lanjutan

penghasilan perbulan	total pembayaran pertama	Remark
> 2x ang slip	10-20%	Bad

Penghitungan kedekatan kasus baru pada data testing dengan dua kasus lama pada data training.

Kedekatan kasus baru dengan kasus nomor 1

- A. Kedekatan bobot atribut status perkawinan (menikah dengan belum menikah) = 1
- B. Bobot atribut status perkawinan = 0.5
- C. Kedekatan jumlah tanggungan (tidak ada dengan tidak ada) = 0
- D. Bobot atribut jumlah tanggungan = 1
- E. Kedekatan bobot pendidikan terakhir (SLTA dengan SLTA) = 0
- F. Bobot atribut pendidikan terakhir = 0.5
- G. Kedekatan bobot usia (<21 thn />60 thn dengan < 21 thn .60 thn) = 0
- H. Bobot atribut usia = 0.5
- I. Kedekatan bobot kepemilikan rumah (orang tua dengan orang tua) = 1
- J. Bobot atribut kepemilikan rumah = 0.8
- K. Kedekatan bobot lama tinggal (3-5 thn dengan > 5 thn) =0.5
- L. Bobot atribut lama tinggal = 1
- M. Kedekatan bobot kondisi rumah (permanent dengan permanen) = 0
- N. Bobot atribut kondisi rumah = 0.6
- O. kedekatan bobot jenis pekerjaan (karyawan dengan karyawan)= 0
- P. bobot atribut jenis pekerjaan 0.8
- Q. kedekatan bobot status perusahaan(swasta menengah dengan swsata kecil) = 0.5
- R. bobot atribut status perusahaan = 0.5
- S. kedekatan bobot status kepegawain (kontrak dengan kontrak) ;0
- T. bobot atribut status kepegawain = 0.8
- U. kedekatan bobot masa kerja(<2 thn dengan <2thn) = 0
- V. bobobt atribut masa kerja = 1
- W. kedekatan bobot penghasilan perbulan (>2x slip dengan > 2x ang slip) = 1
- X. bobot atribut penghasilan perbulan =1
- Y. kedekatan bobot pembayaran pertama (10 – 20 % dengan 10-20%) = 0
- Z. bobot atribut pembayaran pertama

Hasil hitung

$$\begin{aligned}
 \text{similarity} &= [(A*B) + (C*D) + (E*F) + (G*H) + (I*J) + (K*1) + (M*N) + (O*P) + \\
 & (Q*R) + (S*T) + (U*V) + (W*X) + (Y*Z)] / B + D + F + H + J + L + N + P + R + T \\
 & + V + X + Z) \\
 &= [(1 * 0.5) + (0*1) + (0*0.5) + (0*0.5) + (1*0.8) + (0.5*1) + (0*0.6) + (0*0.8) + (0.5*0.5) \\
 & + (0*0.8) + 0*1) + (0*1) + (0*1)] / (0.5 + 0.5 + 0.5 + 0.8 + 1 +0.6 + 0.8 + 0.5 +0.8 \\
 & + 1 +1 +1) \\
 &= (0.5 + 0 + 0 + 0 + 0.8 + 0.5 + 0 + 0.25 + 0 + 0 + 0 + 0 + 0) / 10 \\
 &= 2.05 / 10 \\
 &= 0.205
 \end{aligned}$$

Kedekatan kasus baru dengan kasus nomor 2

- A. Kedekatan bobot atribut status perkawinan (menikah dengan belum menikah) = 1
- B. Bobot atribut status perkawinan = 0.5
- C. Kedekatan jumlah tanggungan (>3 dengan tidak ada) = 1
- D. Bobot atribut jumlah tanggungan = 1
- E. Kedekatan bobot pendidikan terakhir (SLTA dengan SLTA) = 0
- F. Bobot atribut pendidikan terakhir = 0.5
- G. Kedekatan bobot usia (21-55 thn dengan < 21 thn .60 thn) = 0.5
- H. Bobot atribut usia = 0.5
- I. Kedekatan bobot kepemilikan rumah (KPR dengan orang tua) = 1
- J. Bobot atribut kepemilikan rumah = 0.8
- K. Kedekatan bobot lama tinggal (3-5 thn dengan > 5 thn) = 0.5
- L. Bobot atribut lama tinggal = 1
- M. Kedekatan bobot kondisi rumah (permanen dengan permanen) = 0
- N. Bobot atribut kondisi rumah = 0.6
- O. kedekatan bobot jenis pekerjaan (karyawan dengan karyawan) = 0
- P. bobot atribut jenis pekerjaan 0.8
- Q. kedekatan bobot status perusahaan (swasta menengah dengan swasta kecil) = 0.5
- R. bobot atribut status perusahaan = 0.5
- S. kedekatan bobot status kepegawain (tetap dengan kontrak) = 1
- T. bobot atribut status kepegawain = 0.8
- U. kedekatan bobot masa kerja (<5 thn dengan <2thn) = 1
- V. bobot atribut masa kerja = 1
- W. kedekatan bobot penghasilan perbulan (>3x slip dengan > 2x anggs slip) = 0.5
- X. bobot atribut penghasilan perbulan = 1
- Y. kedekatan bobot pembayaran pertama (10 – 20 % dengan 10-20%) = 0
- Z. bobot atribut pembayaran pertama = 1

Hasil hitung

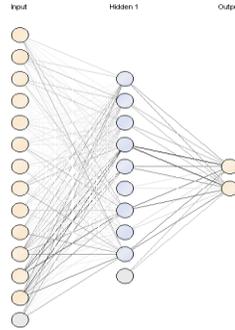
$$\begin{aligned} \text{similarity} &= [(A*B) + (C*D) + (E*F) + (G*H) + (I*J) + (K*1) + (M*N) + (O*P) + \\ & (Q*R) + (S*T) + (U*V) + (W*X) + (Y*Z)] / B + D + F + H + J + L + N + P + R + T \\ & + V + X + Z \\ &= [(1 * 0.5) + (1*1) + (0*0.5) + (0.5*0.5) + (1*0.8) + (0.5*1) + (0*0.6) + (0*0.8) + \\ & (0.5*0.5) + (1*0.8) + 1*1) + (0.5*1) + (0*1)] / (0.5 + 1 + 0.5 + 0.8 + 1 + 0.6 + 0.8 \\ & + 0.5 + 0.8 + 1 + 1 + 1) \\ &= (0.5 + 1 + 0 + 0.25 + 0.8 + 0.5 + 0 + 0.25 + 0.8 + 1 + 0.5 + 0) / 10 \\ &= 5.6 / 10 \\ &= 0.56 \end{aligned}$$

Setelah di hitung ternyata kedekatan yang terendah adalah kasus yang terdekat dengan kasus baru adalah kasus nomor 1.

Neural Network

Gambar 3 adalah *neural net* yang dihasilkan dari pengolahan data *training* dengan metode *neural network* adalah *multilayer perceptron* yang dihasilkan dari data *training* pada Tabel 4. Terdiri dari tiga *layer*, yaitu *Input layer* terdiri dari tiga belas simpul, sama dengan jumlah atribut prediktor ditambah satu simpul bias. Pada pembahasan ini digunakan satu *hidden layer*

yang terdiri dari sembilan simpul ditambah satu simpul bias. Di bagian *output layer* terdapat dua simpul yang mewakili atribut kelas yaitu *good* dan *bad*



Gambar 3. *Neural net* yang dihasilkan dengan metode *neural network*

Untuk setiap data pada data *training* dari Tabel 4, dihitung input untuk simpul berdasarkan nilai input dan jaringan saat itu. Bobot awal untuk *input layer*, *hidden layer*, dan bias diinisialisasi secara acak. Simpul bias terdiri dari dua, yaitu pada *input layer* yang terhubung dengan simpul-simpul pada *hidden layer*, dan pada *hidden layer* yang terhubung pada *output layer*. Setelah semua nilai awal diinisialisasi, kemudian dihitung masukan, keluaran, dan *error*. Selanjutnya membangkitkan output untuk simpul menggunakan fungsi aktivasi sigmoid. Setelah didapat nilai dari fungsi aktivasi, hitung nilai *error* antara nilai yang diprediksi dengan nilai yang sesungguhnya. Setelah nilai *error* dihitung, selanjutnya dibalik ke *layer* sebelumnya (*backpropagated*). Nilai *Error* yang dihasilkan dari langkah sebelumnya digunakan untuk memperbarui bobot relasi. Hasil perhitungan akhir *backpropagation*.

Tabel 10. Nilai bobot akhir untuk *hidden layer*

Hidden Layer (Sigmoid)								
1	2	3	4	5	6	7	8	9
1.167	0.964	-1.913	0.682	2.675	-0.087	-0.489	1.862	3.481
0.155	0.064	5.294	1.243	1.747	-1.747	-0.456	2.567	-0.856
1.172	-0.903	8.467	-1.102	-1.093	-1.717	4.001	0.641	3.193
0.456	4.538	-0.954	-0.842	0.928	-0.455	1.498	0.72	-1.691
5.92	0.752	3.148	3.667	-1.701	0.396	2.811	0.449	3.086
-2.526	-5.657	-1.026	-3.32	0.284	1.007	5.028	-3.928	-0.976
-1.057	-0.871	-2.946	1.582	-2.227	-1.493	2.518	-1.111	-0.684
3.51	1.523	-2.256	-0.098	0.036	1.033	-1.428	1.983	1.993
-3.834	-0.623	2.911	2.672	-3.18	-0.687	-0.285	-2.061	2.422
-3.036	-0.569	-4.338	-1.359	-3.845	-0.253	-3.268	-2.299	-1.872
-1.651	-0.688	-0.68	4.977	3.776	9.358	-1.908	2.957	3.568
7.276	13.293	-5.149	0.37	-0.962	0.839	0.603	1.738	-1.252
2.64	-6.89	-1.254	7.353	5.557	7.04	-1.089	4.936	-0.839
-0.049	0.571	2.617	1.008	-0.932	4.391	-0.206	-1.698	0.777

Tabel 10 adalah nilai akhir fungsi aktivasi pada *output layer*. Kolom pertama pada Tabel 11 menyatakan *class*, yaitu atribut kelas yang dinyatakan dengan simpul pada *output layer* seperti pada gambar 3. Nilai yang terdapat pada kolom berlabel angka satu sampai Sembilan adalah nilai bias terbaru yang terdapat pada relasi antara simpul pada *hidden layer* dan simpul pada *output layer*.

Tabel 11 Nilai Bobot Akhir untuk *Output Layer*

Class	output (sigmoid)						threshold			
	1	2	3	4	5	6				
Good	6.26	-12.451	-6.973	5.235	3.53	8.909	-6.74	2.092	4.401	-3.676
Bad	-6.259	12.45	6.972	-5.236	-3.531	-8.908	6.741	-2.09	4.401	3.677

Evaluasi dan Validasi

Penelitian ini bertujuan untuk melihat akurasi analisis kredit pada *leasing* dalam menilai kelayakan konsumen dibandingkan dengan menggunakan algoritma C4.5, *K-NN*, dan *neural network*, kemudian menganalisa akurasi dengan membandingkan ketiga metode tersebut.

Pengujian

Pengujian yang telah dihitung tingkat akurasi dengan memasukkan data uji yang berasal dari data *training*. Data dalam penelitian ini 486 data maka digunakan metode *cross validation* untuk menguji tingkat akurasi. Untuk nilai akurasi model untuk metode C4.5 sebesar 92.8%, metode *K-NN* sebesar 77.78%, dan metode *neural network* sebesar 91.1%

1. Confusion Matrix

Tabel 4.1 adalah perhitungan berdasarkan data *training* pada Tabel 4.9, diketahui dari 486 data, 152 diklasifikasikan *bad* sesuai dengan prediksi yang dilakukan dengan metode C4.5, lalu 8 data diprediksi *bad* tetapi ternyata *good*, 299 data *class good* diprediksi sesuai, dan 27 data diprediksi *good* ternyata *bad*.

Tabel 12 Model confusion Matrix untuk C45

accuracy: 92.79% +/- 2.13% (mikro: 92.80%)			
	true bad	true good	class precision
pred. bad	152	8	95.00%
pred. good	27	299	91.72%
class recall	84.92%	97.39%	

Tabel 4.10 adalah *confusion matrix* untuk metode *K-NN*. Diketahui dari 486 data, 157 diklasifikasikan *bad* sesuai dengan prediksi yang dilakukan dengan metode C4.5, lalu 86 data diprediksi *bad* tetapi ternyata *good*, 221 data *class good* diprediksi sesuai, dan 22 data diprediksi *good* ternyata *bad*.

Tabel 13 Model confusion matrik untuk metode K-NN

accuracy: 77.78% +/- 5.17% (mikro: 77.78%)			
	true bad	true good	class precision
pred. bad	157	86	64.61%
pred. good	22	221	90.95%
class recall	87.71%	71.99%	

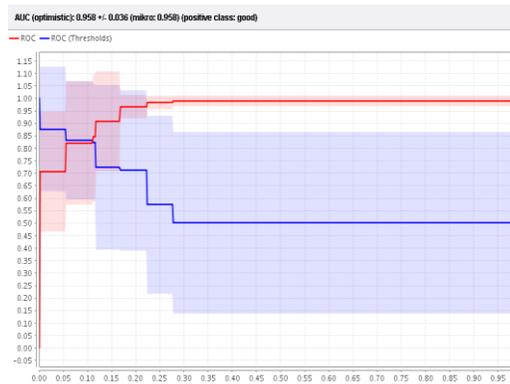
Dengan metode *neural network*, menghasilkan kondisi seperti pada Tabel 4.11 Diketahui dari 486 data, 157 diklasifikasikan *bad* sesuai dengan prediksi yang dilakukan dengan metode C4.5, lalu 21 data diprediksi *bad* tetapi ternyata *good*, 283 data *class good* diprediksi sesuai, dan 22 data diprediksi *good* ternyata *bad*.

Tabel 14 Model Confusion Matrix untuk metode neural network

accuracy: 91.16% +/- 3.58% (mikroc: 91.15%)			
	true bad	true good	class precision
pred. bad	157	21	88.20%
pred. good	22	286	92.86%
class recall	87.71%	93.16%	

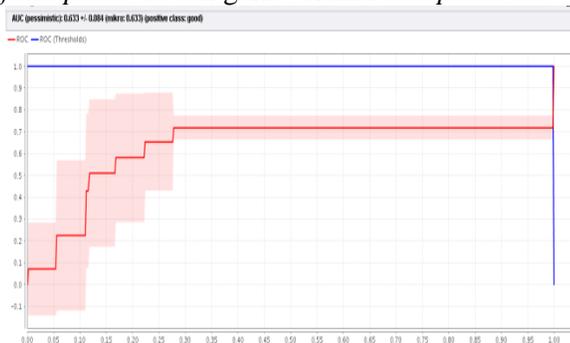
Kurva ROC

Hasil perhitungan divisualisasikan dengan kurva ROC. Perbandingan ketiga metode komparasi bisa dilihat pada Gambar 4.5 yang merupakan kurva ROC untuk algoritma C4.5.



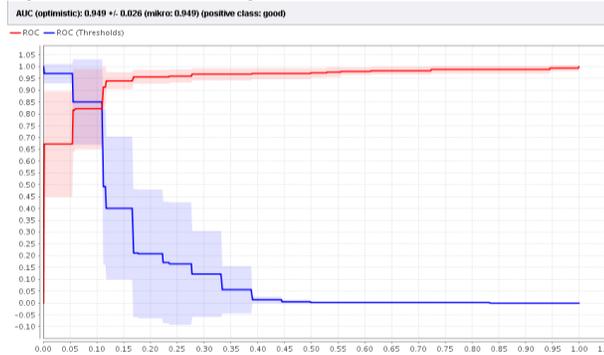
Gambar 4 Kurva ROC dengan algoritma C4.5

Kurva ROC pada gambar 4.5 mengekspresikan *confusion matrix* dari Tabel 4.9. Garis horizontal adalah *false positives* dan garis vertikal *true positives*.



Gambar 5 Kurva ROC dengan Metode K-NN

Seperti terlihat pada Gambar 4, Gambar 5, dan Gambar 6, area di bawah kurva pada Gambar 4.3 paling luas diantara ketiga metode



Gambar 6 Kurva ROC dengan Metode *Neural Network*

Pebandingan hasil perhitungan nilai AUC untuk metode C4.5, *K-NN*, dan *neural network* dapat dilihat pada Tabel 4.13.

Tabel 15 Komparasi Nilai AUC

	C4.5	<i>K-NN</i>	<i>Neural network</i>
AUC	0.958	0.633	0.949

Analisis Hasil Komparasi

Analisis yang dihasilkan dengan metode C4.5, *K-NN*, dan *neural network* diuji menggunakan metode *Cross Validation*, terlihat perbandingan nilai *accuracy*, pada Tabel 4.13, untuk metode C4.5 memiliki nilai *accuracy* yang paling tinggi, diikuti dengan metode *neural network*, dan yang terendah adalah *K-NN*.

Tabel 16 Komparasi Nilai *Accuracy* dan AUC

	C4.5	<i>K-NN</i>	<i>Neural network</i>
<i>Accuracy</i>	92.89%	77.78%	91.10%
AUC	0.958	0.905	0.949

Tabel 4.14 membandingkan *accuracy* dan AUC dari tiap metode. Terlihat bahwa nilai *accuracy* C4.5 paling tinggi begitu pula dengan nilai AUC-nya. Untuk metode *neural network* dan *K-nn* juga menunjukkan nilai yang sesuai. Untuk klasifikasi *data mining*, nilai AUC dapat dibagi menjadi beberapa kelompok (Gorunescu,2011).

- 0.90-1.00 = klasifikasi sangat baik
- 0.80-0.90 = klasifikasi baik
- 0.70-0.80 = klasifikasi cukup
- 0.60-0.70 = klasifikasi buruk
- 0.50-0.60 = klasifikasi salah

Berdasarkan pengelompokan di atas dan Tabel 4.14 maka dapat disimpulkan bahwa metode C4.5, *K-NN*, dan *neural network* termasuk klasifikasi sangat baik karena memiliki nilai AUC antara 0.90-1.00.

Perancangan Sistem

Data baru diterapkan pada algoritma yang memiliki akurasi paling tinggi, dalam hal ini adalah algoritma C4.5. Data baru yang digunakan sebanyak sepuluh record diuji dengan menggunakan confusion matrix dan diperoleh akurasi dan persisi sebesar 90 %.

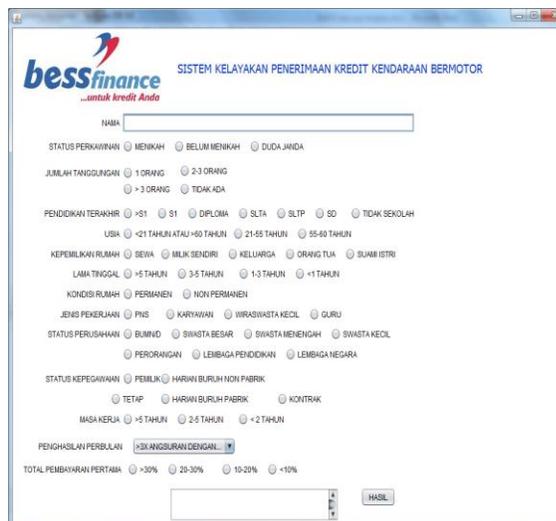
Tabel 17 Tabel Nilai Akurasi dan Persisi Data Baru Algoritma C4.5

Accuracy : 90%
Precision : 100%

	Good	Bad	
Prediction Good	1	7	87.50%
PredictionBad	2	0	100%
Classrecall	66.67%	100%	

Dari tabel 4.15 diketahui bahwa data baru yang digunakan sebanyak 10 record. Sebanyak 1 record diprediksi secara Good dan 7 record diprediksi secara benar *Bad*. Rule pada algoritma C4.5 digunakan sebagai rule untuk pembuatan *interface* untuk memudahkan pihak leasing dalam penentuan kelayakan penerimaan kredit. *Interface* sebagai implementasi hasil penelitian ini dirancang menggunakan bahasa pemrograman Java.

Berdasarkan aturan-aturan di atas dirancang model penentu kelayakan kredit sebagai berikut:



Gambar 4.8 Tampilan implementasi model penentu kelayakan kredit

PENUTUP

Simpulan

1. Dalam penelitian ini dilakukan pengujian menggunakan algoritma C4.5, *K-NN* dan *neural network* menggunakan data konsumen yang mendapat penerimaan kredit motor. Kemudian hasil, dikomparasi untuk mengetahui algoritma yang paling baik dalam penentuan resiko kredit kendaraan bermotor. Untuk mengukur kinerja ketiga algoritma

tersebut digunakan metode pengujian *Cross Validation*, *Confusion Matrix* dan Kurva ROC, diketahui bahwa algoritma C4.5 memiliki nilai *accuracy* dan AUC paling tinggi, diikuti oleh metode *neural network*, dan yang paling rendah metode *K-NN*.

2. Dengan demikian, metode *decision tree* merupakan metode yang cukup baik dalam pengklasifikasian data, dengan demikian algoritma C4.5 dapat memberikan pemecahan untuk permasalahan penentuan kelayakan konsumen yang mendapatkan kredit kendaraan bermotor pada *leasing*.

DAFTAR PUSTAKA

- Alpayadin, Ethem. 2010. **Introduction to Machine Learning**. The MIT Press. London.
- Bungin, B. 2005. **Metodologi Penelitian Kuantitatif**. Kencana. Jakarta.
- Gorunescu, Florin. 2011. **Data Mining: Concepts, Models, and Techniques**. Verlag Berlin Heidelberg. Springer
- Han, J., & Kamber, M.. 2006. **Data Mining Concept and Tehniques**. San Fransisco. Morgan Kauffman.
- Hariani, Iswi. 2010. **Restrukturisasi dan Penghapusan Kredit Macet**. Jakarta: PT Elexmedia Komputindo.
- Jiang, Yi. et al. 2007. **A Bank Customer Credit Evaluation Based on the Decision Tree and the Simulated Annealing Algorithm**. *Journal of Department of Computer Science Xiamen University* . IEEE International Co 8-11 July 2008.
- Kusrini, dan Luthfi, Emha Taufik. 2009. **Algoritma Data Mining**, Edisi I, Yogyakarta: Andi Publishing.
- Larose, Daniel T. 2005. **Discovering Knowledge in Data: An Introduction to Data mining**. Jhon Willey & Son Inc., New Jersey.
- Liao. 2007. **Recent Advances in Data Mining of Enterprise Data: Algorithms and Application**. Singapore. World Scientific Publishing
- Linof, Gordon S & Berry, Michael J. 2011. **Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management**. Indiana. Wiley Publishing.
- Maimon, Oded & Rokach, Lior. 2010. **Data Mining and Knowledge Discovey Handbook**. New York: Springer
- Mania, Hasan & Patel. 2011. **Comparative study of Naïve Bayes Classifier and KNN for Tuberculosis**. *International Journal of Computer Applications (IJCA)*.
- Mujiasih, Subekti. 2011. **Pemanfaatan Data Mining Untuk Prakiraan Cuaca**. *Jurnal Meteorologi dan Geofisika*. 12 (2):189-195.
- Rivai, Veithzal., & Veithzal, Andria Permata. 2006. **Credit Management Handbook**. Jakarta: Raja Grafindo Persada.
- Satchidananda, S S & Jay B. Simha. 2006. **Comparing Decision Trees With Logistic Regression For Credit Risk Analysis (SAS APAUGC)**
- Sugiyono 2001. **Metode Penelitian Bisnis**. CV. Alfabeta. Bandung.
- Vercellis, Carlo. 2009. **Business Inteligent: Data Mining and Optimization for Decision Making**. Southern Gate. Chichester. west Sussex. John Willey & Sons, Ltd.
- Yadav, Kumar , Surjeet & Pal, Saurabh. 2012. **Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification**. *World of Computer Science and Information Technology Journal (WCSIT)* . 2 (2) 51-56.