

Klasifikasi Nilai Kelayakan Calon Debitur Baru Menggunakan *Decision Tree* C4.5

Bambang Hermanto*¹, Azhari SN²

¹Prodi S2 Ilmu Komputer FMIPA UGM, Yogyakarta

²Jurusan Ilmu Komputer dan Elektronika, FMIPA UGM, Yogyakarta

e-mail: *bbg.hermanto@gmail.com, azharisn@gmail.com

Abstrak

Dalam upaya meningkatkan kualitas layanan konsumen khususnya dalam hal penilaian kelayakan calon debitur seiring dengan bertambahnya jumlah calon debitur baru pembiayaan kredit pembelian kendaraan bermotor, maka perusahaan membutuhkan alat bantu pengambilan keputusan sehingga dapat dengan mudah dan cepat memperkirakan manakah debitur yang mampu melunasi kredit.

Penelitian ini membahas proses generate decision tree menggunakan algoritma C4.5 dan memanfaatkan sekelompok learning dataset debitur pembiayaan kendaraan sepeda motor. Pohon keputusan tersebut kemudian diinterpretasikan kedalam bentuk aturan-aturan keputusan yang dapat dipahami dan digunakan sebagai acuan dalam mengolah data-data calon debitur dalam menentukan kelayakan calon debitur baru. Nilai kelayakan mengacu pada nilai parameter tujuan yakni status kredit. Bila nilai status kredit adalah lunas berarti diperkirakan calon debitur yang bersangkutan mampu melunasi kredit (berpotensi kredit lunas), akan tetapi jika parameter status kredit bernilai tarikan berarti diperkirakan calon debitur yang bersangkutan tidak mampu melunasi kredit (berpotensi kredit macet).

Pengujian sistem dilakukan dengan cara membandingkan hasil antara testing data dengan learning data dalam 3 skenario dengan hasil keputusan bahwa data valid di atas 70% untuk semua skenario kasus. Selain itu, dalam pembuatan pohon keputusan dan aturan keputusan (generate rules) dibutuhkan waktu yang cukup cepat, yakni tidak lebih dari 15 menit untuk setiap skenario pengujian.

Kata kunci—kredit, C4.5, decision tree

Abstract

In an effort to improve the quality of customer service, especially in terms of feasibility assessment of borrowers due to the increasing number of new prospective borrowers loans financing the purchase of a motor vehicle, then the company needs a decision making tool allowing you to easily and quickly estimate Where the debtor is able to pay off the loans.

This study discusses the process generates C4.5 decision tree algorithm and utilizing the learning group of debtor financing dataset motorcycle. The decision tree is then interpreted into the form of decision rules that can be understood and used as a reference in processing the data of borrowers in determining the feasibility of prospective new borrowers. Feasibility value refers to the value of the destination parameter credit status. If the value of the credit is paid off status mean estimated prospective borrower is able to repay the loan in question, but if the credit status parameters estimated worth pull means candidates concerned debtor is unable to pay loans..

System testing is done by comparing the results of the testing data by learning data in three scenarios with the decision that the data is valid at over 70% for all case scenarios. Moreover, in generated tree and generate rules takes fairly quickly, which is no more than 15 minutes for each test scenario

Keywords—credit, C4.5, Decision tree

1. PENDAHULUAN

1.1. Latar Belakang

Istilah debitur menurut undang-undang no.10 tahun 1998 tentang perubahan atas undang-undang nomor 7 tahun 1992 tentang perbankan adalah nasabah yang memperoleh fasilitas kredit atau pembiayaan berdasarkan prinsip syariah atau yang dipersamakan dengan itu berdasarkan perjanjian bank dengan nasabah yang bersangkutan. Definisi kredit menurut kamus besar Bahasa Indonesia adalah pinjaman uang dengan pembayaran pengembalian secara mengangsur.

Penelitian ini membahas proses *generate decision tree* menggunakan algoritma C4.5 dan memanfaatkan sekelompok *learning dataset* debitur pembiayaan kendaraan sepeda motor. Pohon keputusan tersebut kemudian diinterpretasikan kedalam bentuk aturan-aturan keputusan (*rules*) yang dapat dipahami dan digunakan sebagai acuan dalam mengolah data-data calon debitur (*test dataset*) dalam menentukan kelayakan calon debitur baru. Nilai kelayakan mengacu pada nilai parameter tujuan (*target variable*) yakni status kredit. Bila nilai status kredit adalah lunas berarti diperkirakan calon debitur yang bersangkutan mampu melunasi kredit (berpotensi kredit lunas), akan tetapi jika parameter status kredit bernilai tarikan berarti diperkirakan calon debitur yang bersangkutan tidak mampu melunasi kredit (berpotensi kredit macet).

Menurut [1] pohon keputusan (*decision tree*) dapat membagi kumpulan data yang besar (*learning dataset*) menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan (*rules*), sehingga anggota himpunan hasil menjadi mirip satu dengan lainnya dengan memperhatikan pada variabel tujuannya (*target variable*). *Decision tree* merupakan salah satu metoda klasifikasi *data mining* yang populer digunakan karena mudah diinterpretasikan oleh manusia dengan konsep dasar mengubah data menjadi pohon keputusan dan aturan-aturan keputusan [10], mudah mengintegrasikan dengan sistem basis data dan memiliki tingkat ketelitian (akurasi) yang baik diatas 90% [2]. Peranan pohon keputusan sebagai alat bantu pengambilan keputusan (*decision support tool*) telah dikembangkan oleh manusia untuk membantu mencari dan membuat keputusan masalah dengan memperhitungkan berbagai macam faktor yang ada di dalam lingkup masalah tersebut. Dengan pohon keputusan, manusia dapat dengan mudah mengidentifikasi dan melihat hubungan antara faktor-faktor yang mempengaruhi suatu masalah dan dapat mencari penyelesaian terbaik dengan memperhitungkan faktor-faktor tersebut.

Beberapa penelitian yang sudah pernah dilakukan dan berkaitan dengan metoda klasifikasi *data mining* yang digunakan dalam penelitian ini. Menurut [3] menggabungkan algoritma GATS (*Genetic Algorithm-Tabu Search*) sebagai strategi pencarian dan algoritma C4.5 sebagai fungsi evaluasi dengan percobaan kategori dataset *www*, *mail*, *p2p*. Dengan algoritma ini dapat meningkatkan komputasi kinerja tanpa dampak negatif pada akurasi klasifikasi. Penelitian lainnya dilakukan [4] menggunakan algoritma C4.5 untuk peramalan resiko keuangan pada informasi *real* dari akuntansi pada perusahaan keuangan pada tahun 2005 dan 2006. Penelitian ini melibatkan indeks finansial dan non-finansial sebagai objek penelitian, dan hasilnya kemampuan peramalan pada penelitian ini lebih baik daripada hanya melibatkan indeks finansial saja. Penelitian selanjutnya menggabungkan algoritma C4.5 dan teori himpunan kasar (*rough set theory*). Teori himpunan ini mengurangi atribut yang digunakan menggunakan data tadi sebagai input pada algoritma C4.5 untuk membuat *decision tree*. Hasil penelitian menunjukkan optimasi algoritma C4.5 pada penelitian ini mempunyai efisiensi dan akurasi yang lebih tinggi dibandingkan hanya menggunakan algoritma C4.5 [5].

Beberapa penelitian yang telah dilakukan tentang *data mining* menggunakan pendekatan optimasi algoritma *decision tree C4.5* antara lain penelitian yang dilakukan [6], teknologi deteksi gangguan pada keamanan jaringan menggunakan metode *neural network* dan C4.5. Ide utama dari penelitian tersebut adalah memanfaatkan kemampuan klasifikasi dari kedua metode tersebut untuk menanggulangi beberapa serangan yang berbeda-beda. Penelitian kedua, [7] dengan menggabungkan *decision tree* dengan algoritma genetika dan diuji pada 24 basis data yang berbeda. Dari hasil penelitian, didapatkan bahwa tingkat kesalahan pada penggabungan

kedua metode ini lebih kecil daripada hanya menggunakan *decision tree*. Penelitian yang lain menggunakan algoritma C4.5 dan ditingkatkan efisiensinya pada pemilihan atribusi dan metode partisi (R-C4.5). Model ini menghindari munculnya fragmentasi dengan menyatukan cabang yang memiliki efek yang kecil jika diklasifikasikan. Penelitian ini berfokus pada penerapan R-C4.5 untuk penelitian tentang perawatan kesehatan yang memprediksi lama tinggal para pasien rawat inap. Hasil penelitian ini dapat membantu institusi kesehatan untuk mengatur dan memanfaatkan dengan maksimal fasilitas rumah sakit yang ada [5]. [8] dengan menggabungkan algoritma GATS (*Genetic Algorithm-Tabu Search*) sebagai strategi pencarian dan algoritma C4.5 sebagai fungsi evaluasi. Algoritma menguji kategori dataset *www*, *www*, *mail*, *p2p*. Dengan algoritma ini dapat meningkatkan komputasi kinerja tanpa dampak negatif pada akurasi klasifikasi. Penelitian yang lain dilakukan [6] menggunakan algoritma C4.5 yang digunakan untuk peramalan resiko keuangan pada informasi *real* dari akuntansi pada perusahaan keuangan pada tahun 2005 dan 2006. Penelitian ini melibatkan indeks finansial dan non-finansial sebagai objek penelitian, dan hasilnya kemampuan peramalan pada penelitian ini lebih baik daripada hanya melibatkan indeks finansial saja.

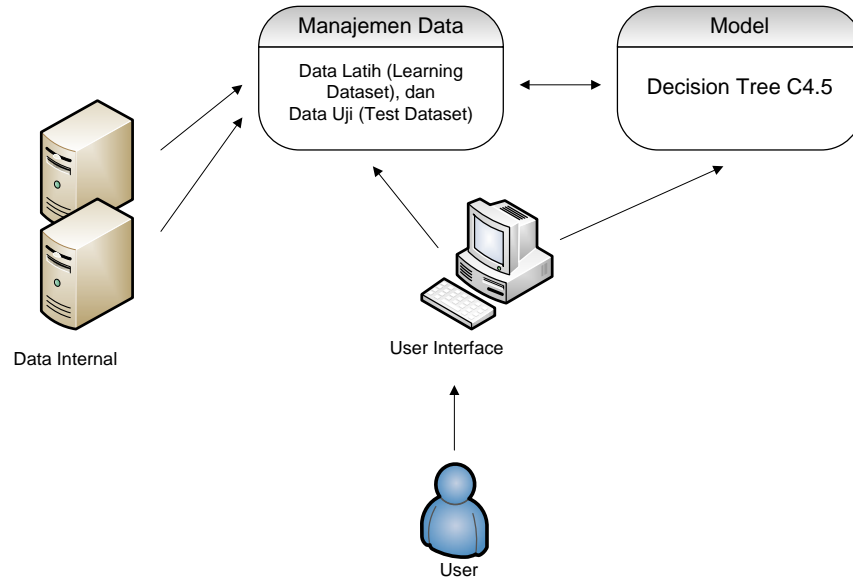
1.2. Tujuan Penelitian

Menghasilkan sebuah *software* sebagai pendukung keputusan dalam penilaian kelayakan calon debitur baru menggunakan metoda algoritma C4.5 sehingga bermanfaat bagi manajer perusahaan dalam pembiayaan kendaraan bermotor.

2. METODE PENELITIAN

2.1. Arsitektur Sistem

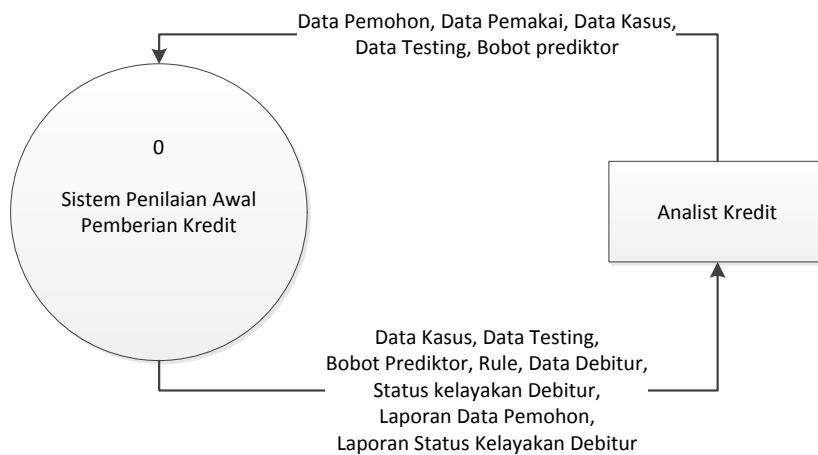
Sistem yang dibangun menggunakan arsitektur sistem *database* tunggal, artinya *database* dan aplikasi diletakkan pada komputer yang sama atau tidak berada dalam lingkup jaringan komputer [9]. Data internal yang diproses dalam manajemen data berupa data debitur lama yang sudah diketahui nilai kelayakannya (data latih) dan juga data calon debitur baru (data uji). Data latih tersebut berisi tentang data jenis pekerjaan, besar penghasilan (*salary*), tenor, nilai angsuran (*installment*) dan status kredit debitur. Sedangkan data uji berisi data nama nasabah, jenis pekerjaan, besar penghasilan (*salary*), tenor, dan nilai angsuran calon debitur baru. Dari beberapa atribut yang tersedia, sebagai variabel *predictor* adalah jenis pekerjaan, besar penghasilan, tenor dan nilai angsuran. Sedangkan status kredit dijadikan sebagai variabel target. Sebagai contoh, pada kriteria penghasilan (*salary*) akan diklasifikasi menjadi 3 kelompok yakni penghasilan tinggi, penghasilan sedang dan penghasilan rendah. Setelah data internal tersebut diproses pada manajemen data, maka data tersebut akan digunakan pada manajemen model untuk proses perhitungan pembuatan pohon keputusan menggunakan algoritma C4.5. Hasilnya digunakan sistem untuk memperkirakan calon debitur baru berpotensi kredit lunas. Arsitektur system yang dibuat terlihat pada Gambar 1.



Gambar 1 Arsitektur Sistem

2.2. Diagram Konteks

Sistem Kredit ini akan diakses oleh pengguna Analist kredit. Analist kredit bisa menyimpan dan merubah data pemohon (debitur), input data pemohon kredit dan input data kriteria. Kemudian Analist Kredit juga akan mendapatkan konfirmasi dari pengguna yang telah ditambahkan dan data keputusan yang diproses oleh sistem. Analist Kredit juga mendapatkan laporan dari proses yang dilakukan oleh system. Proses klasifikasi kelayakan yang terjadi pada sistem dimulai dengan perbandingan data kasus yang diberikan bobot sesuai dengan kriteria. Hasil dari perbandingan data kasus dan bobot menghasilkan sebuah tree yang bisa dijadikan rule untuk menghasilkan status kelayakan dari seorang debitur yang mengajukan kredit untuk pembiayaan kendaraan bermotor. Diagram konteks sistem yang dibuat terlihat pada Gambar 2.

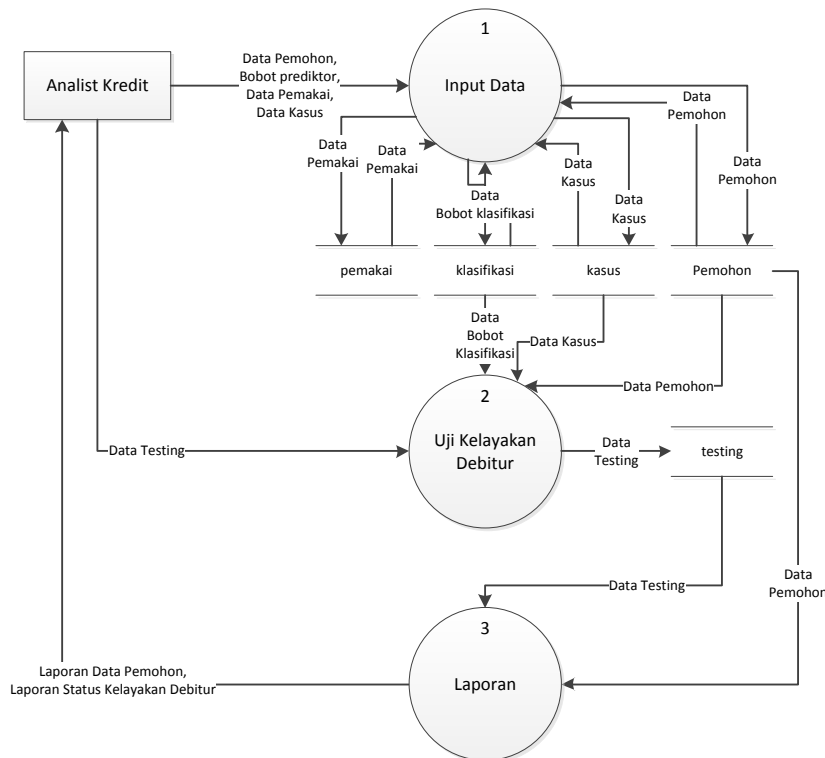


Gambar 2 Diagram Konteks

2.3. Diagram Level 1

Data flow diagram level 1 memiliki 3 proses, yaitu proses input data, proses uji kelayakan, dan proses laporan. Proses input data terjadi proses pengaturan hak akses sistem, input data

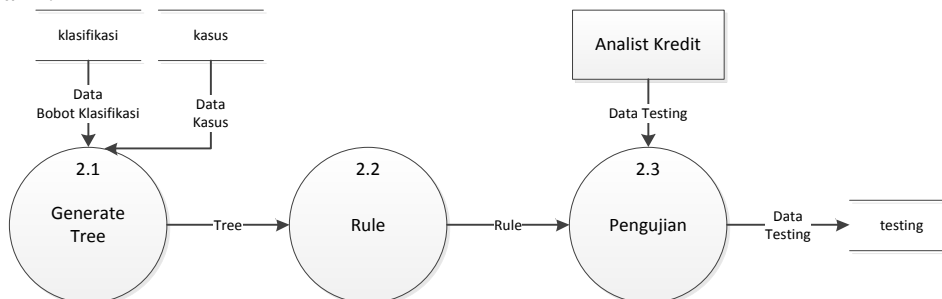
bobot klasifikasi, input kasus dan input data pemohon (debitur). Proses uji kelayakan merupakan proses utama yaitu melakukan proses yang menghasilkan seorang pemohon (debitur) dinyatakan layak atau tidaknya untuk mendapatkan kredit. Hasil proses uji kelayakan didapat dari pemrosesan bobot klasifikasi dan data kasus sehingga mendapatkan hasil uji kelayakan yang akurat. Proses terakhir adalah proses laporan yang mengambil data dari data pemohon (debitur) dan hasil uji kelayakan. DFD level 1 dapat dilihat pada Gambar 3.



Gambar 3 Data Flow Diagram Level 1

2.4. Diagram Level 2 Proses 2

DFD Level 2 Proses 2 menggambarkan subproses yang terjadi pada proses uji kelayakan. Subproses uji kelayakan terdiri dari *Generate Tree*, *Rule* dan *Pengujian*. *Generate tree* memproses data bobot klasifikasi dan data kasus. Hasil dari *generate tree* adalah berupa pohon keputusan dimana pohon keputusan tersebut digunakan untuk membuat rule klasifikasi. Setelah *rule* terbentuk, dilakukan pengujian data. Data flow diagram level 2 proses 2 terlihat pada Gambar 4.



Gambar 4 Data Flow Diagram Level 2 Proses 2

3. HASIL DAN PEMBAHASAN

3.1. Pembentukan pohon keputusan

Berikut uraian langkah-langkah dalam algoritma C4.5 dalam membangun sebuah pohon keputusan untuk menyelesaikan kasus debitur yang berpotensi kredit lunas atau tidak lunas (tarikan) berdasarkan kriteria jenis pekerjaan (*occupation*), besar penghasilan (*salary*), tenor dan nilai angsuran (*installment*).

Tabel 1 Data latih nasabah debitur pembiayaan kendaraan

NO	OCCUPATION	SALARY	TENOR	INSTALLMENT	STATUS KREDIT
1	PTPN	2,950,000	31	905,000	LUNAS
2	PTPN	2,400,000	17	1,425,000	LUNAS
3	HONORER	2,800,000	36	895,000	LUNAS
4	PNS	3,000,000	21	1,110,000	TARIKAN
5	HONORER	1,500,000	31	589,000	TARIKAN

Langkah awal yaitu mempersiapkan data latih (*learning dataset*). Tabel 1 berisi data-data nasabah debitur berjumlah 5 *record* yang akan digunakan untuk membentuk pohon keputusan. Atribut-atribut yang digunakan sebagai *predictor* adalah jenis pekerjaan (*occupation*), besar penghasilan (*salary*), tenor dan nilai angsuran (*installment*) dimana semuanya sudah memiliki nilai masing-masing. Dan sebagai kelasnya adalah nilai atribut status kredit yaitu kelas “Lunas” dan kelas “Tarikan”. Lalu nilai-nilai atribut *predictor* akan diklasifikasikan berdasarkan tabel aturan klasifikasi nilai atribut *predictor* nasabah debitur yang ditetapkan oleh pihak PT. Mega Finance, dengan langkah sebagai berikut.

Pada Tabel 1 baris pertama diketahui informasi debitur dengan pekerjaan PTPN, penghasilan per bulan = Rp. 2.950.000, tenor = 31 bulan, dan nilai angsuran per bulan = Rp. 905.000,-. Dengan melihat Tabel 1, maka diperoleh klasifikasi pekerjaan PTPN adalah “O1”, penghasilan Rp. 2.950.000 adalah “S1”, tenor 31 bulan adalah “T3”, dan *installment* Rp. 905.000,- adalah “I2”. Setelah mengklasifikasikan semua nilai atribut *predictor* dengan cara yang sama, maka hasilnya terlihat pada Tabel 2.

Tabel 2 Hasil klasifikasi nilai atribut *predictor* Tabel 1

NO	OCCUPATION	SALARY	TENOR	INSTALLMENT	STATUS KREDIT
1	O1	S1	T3	I2	LUNAS
2	O1	S1	T2	I2	LUNAS
3	O3	S1	T3	I2	LUNAS
4	O1	S2	T2	I2	TARIKAN
5	O3	S1	T3	I1	TARIKAN

Keterangan:

- O1 = Pekerjaan Pegawai Negeri/BUMN (PNS, POLRI, TNI, PLN, PTPN, ASKES)
- O2 = Pekerjaan Swasta/Wirusaha (Pedagang, Pengusaha, Instansi Swasta, pabrik, toko)
- O3 = Pekerjaan lainnya (Honorar, Sukarelawan, Belum bekerja)
- S1 = Penghasilan antara Rp. 1.500.000 s.d. Rp. 3.000.000
- S2 = Penghasilan antara Rp. 3.000.001 s.d. Rp. 6.000.000
- S3 = Penghasilan antara Rp. 6.000.001 ke atas
- T1 = Tenor antara 11 bulan s.d. 16 bulan
- T2 = Tenor antara 17 bulan s.d. 30 bulan

- T3 = Tenor antara 31 bulan ke atas
- I1 = Nilai Angsuran antara Rp.500.000 s.d. Rp.700.000
- I2 = Nilai Angsuran antara Rp.700.001 s.d. Rp.1.500.000
- I3 = Nilai Angsuran antara Rp.1.500.001 ke atas

Langkah berikutnya yaitu perhitungan *Node* 1. Dimulai dengan menghitung nilai *entropy* dan *gain*, dengan uraian sebagai berikut.

Hitung jumlah kasus untuk keputusan “Lunas” dan “Tarikan” dari semua kasus, dan juga kasus yang dibagi berdasarkan atribut pekerjaan (*occupation*), penghasilan (*salary*), tenor dan nilai angsuran (*installment*). Tabel 3 menampilkan hasil perhitungan untuk semua kasus data latih, dimana diperoleh total 5 kasus dengan 3 data bernilai status kredit lunas dan 2 data bernilai status kredit tarikan. Nilai pada kolom *entropy* diperoleh dari persamaan (1).

$$\text{Entropy}(S) = -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus} \quad (1)$$

Keterangan

- S : ruang (data) sampel yang digunakan
- p_{\oplus} : jumlah yang bersolusi positif atau mendukung pada data sampel untuk kriteria tertentu
- p_{\ominus} : jumlah yang bersolusi negatif atau tidak mendukung pada data sampel untuk kriteria tertentu

$$\text{Entropy}(\text{Total}) = (-3/5 \times \log_2 (3/5)) - (2/5 \times \log_2 (2/5)) = 0,97095 \quad (1)$$

Tabel 3 Hasil perhitungan total kasus dan entropi data latih

SUM (KASUS)	SUM (LUNAS)	SUM (TARIKAN)	ENTROPY
5	3	2	0.97095

Dengan menggunakan persamaan (1), akan dihitung semua nilai *entropy* untuk kasus yang dibagi berdasarkan atribut pekerjaan (*occupation*), penghasilan (*salary*), tenor dan nilai angsuran (*installment*). Setelah itu, hitung nilai *gain* menggunakan persamaan (2).

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^m \left(\frac{|S_i|}{|S|} * \text{Entropy}(S_i) \right) \quad (2)$$

Keterangan

- A : Atribut
- S : Sampel
- n : Jumlah partisis himpunan atribut A
- $|S_i|$: Jumlah sampel pada pertisi ke- i
- $|S|$: Jumlah sampel dalam S

Sebagai contoh, akan dihitung *entropy* dan *gain* pada salah satu atribut yakni *occupation*. Entropy (O2) tidak perlu dihitung karena tidak ada *record* atribut pekerjaan dengan klasifikasi O2.

$$\text{Entropy} (O1) = (-2/3 \times \log_2 (2/3)) - (1/3 \times \log_2 (1/3)) = 0.9183 \quad (1)$$

$$\text{Entropy} (O3) = (-1/2 \times \log_2 (1/2)) - (1/2 \times \log_2 (1/2)) = 1 \quad (1)$$

$$\text{Gain}(\text{Occupation}) = 0.97095 - ((3/5 * 0.9183)+(2/5 * 1)) = 0.01997 \quad (2)$$

Setelah menghitung semua nilai *entropy* dan *gain* dari masing-masing nilai atribut dengan cara yang sama, akan terlihat hasil akhir perhitungan *node* 1 seperti ditunjukkan pada Tabel 4.

Tabel 4 Perhitungan *node* 1

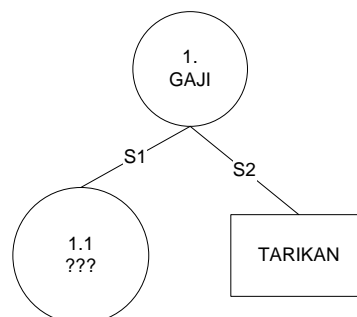
NODE	ATRIBUT	NILAI	KASUS	LUNAS	TARIKAN	ENTROPY	GAIN
1	TOTAL		5	3	2	0.97095	
	OCCUPATION						0.01997
		O1	3	2	1	0.9183	
		O3	2	1	1	1	
	SALARY						0.321926
		S1	4	3	1	0.81128	
		S2	1	0	1	0	
	TENOR						0.01997
		T2	2	1	1	1	
		T3	3	2	1	0.9183	
	INSTALLMENT						0.321926
		I1	1	0	1	0	
		I2	4	3	1	0.81128	

Keterangan:

- O1 = Pekerjaan Pegawai Negeri/BUMN (PNS, POLRI, TNI, PLN, PTPN, ASKES)
- O3 = Pekerjaan lainnya (Honorar, Sukarelawan, Belum bekerja)
- S1 = Penghasilan antara Rp. 1.500.000 s.d. Rp. 3.000.000
- S2 = Penghasilan antara Rp. 3.000.001 s.d. Rp. 6.000.000
- T2 = Tenor antara 17 bulan s.d. 30 bulan
- T3 = Tenor antara 31 bulan ke atas
- I1 = Nilai Angsuran antara Rp.500.000 s.d. Rp.700.000
- I2 = Nilai Angsuran antara Rp.700.001 s.d. Rp.1.500.000

Dari Tabel 4 dapat diketahui bahwa atribut dengan nilai *gain* tertinggi adalah *salary*, yaitu sebesar **0.321926**. Dengan demikian, atribut *salary* menjadi simpul akar (*root node*). *Node* 1 diisi oleh *salary*. Karena *salary* S2 memiliki *entropy* = 0 dengan kelas Tarikan sebanyak 1 *record*, sehingga tidak perlu dilakukan perhitungan lebih lanjut. Dengan demikian *salary* S2 menjadi simpul daun (*leaf node*). Sedangkan untuk *salary* S1 masih perlu dilakukan perhitungan.

Dari hasil analisis tersebut dapat digambarkan sebuah pohon keputusan sementara seperti pada Gambar 5.

Gambar 5 Pohon keputusan *node* 1

Berdasarkan pembentukan pohon keputusan *node* 1, *node* 1.1 akan dianalisis lebih lanjut untuk mengetahui *node* cabang dari *salary* S1 dengan uraian sebagai berikut.

Hitung jumlah kasus untuk keputusan “Lunas” dan “Tarikan” dengan syarat *salary* S1, kemudian hitung nilai *entropy* menggunakan persamaan (1). Begitu pula untuk kasus yang dibagi berdasarkan atribut pekerjaan (*occupation*), tenor dan nilai angsuran (*installment*) dengan syarat atribut *salary* bernilai S1. Setelah dilakukan perhitungan nilai *entropy*, lakukan perhitungan *gain* untuk setiap atribut menggunakan persamaan (2). Setelah menghitung semua nilai *entropy* dan *gain* dari masing-masing nilai atribut, akan terlihat hasil akhir perhitungan *node* 1.1 seperti ditunjukkan pada Tabel 5.

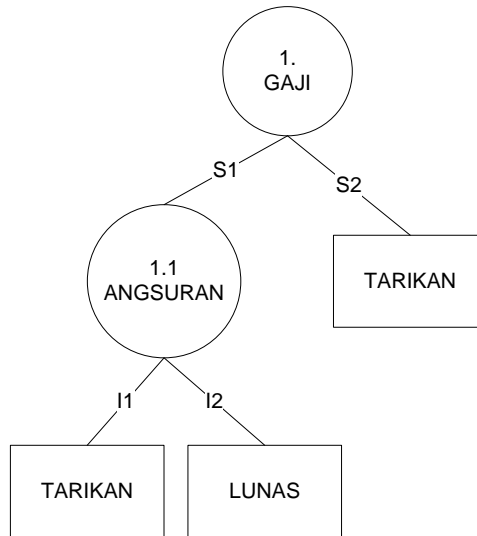
Tabel 5 Perhitungan *node* 1.1

NODE	ATRIBUT	NILAI	KASUS	LUNAS	TARIKAN	ENTROPY	GAIN
1.1	SALARY S1		4	3	1	0.81128	
	OCCUPATION						0.31128
		O1	2	2	0	0	
		O3	2	1	1	1	
	TENOR						0.122555
		T2	1	1	0	0	
		T3	3	2	1	0.9183	
	INSTALLMENT						0.81128
		I1	1	0	1	0	
		I2	3	3	0	0	

Keterangan

- O1 = Pekerjaan Pegawai Negeri/BUMN (PNS, POLRI, TNI, PLN, PTPN, ASKES)
- O3 = Pekerjaan lainnya (Honorar, Sukarelawan, Belum bekerja)
- S1 = Penghasilan antara Rp. 1.500.000 s.d. Rp. 3.000.000
- T2 = Tenor antara 17 bulan s.d. 30 bulan
- T3 = Tenor antara 31 bulan ke atas
- I1 = Nilai Angsuran antara Rp.500.000 s.d. Rp.700.000
- I2 = Nilai Angsuran antara Rp.700.001 s.d. Rp.1.500.000

Dari hasil pada Tabel 5 diketahui bahwa atribut dengan nilai *gain* tertinggi adalah *installment* yaitu sebesar **0.81128**. Dengan demikian *installment* dapat menjadi *node* cabang dari atribut *Salary* S1. Karena nilai *entropy* dari *installment* I1 dan *installment* I2 adalah 0 sehingga perhitungan dianggap selesai, dimana *installment* I1 sebanyak 1 *record* pada kelas Tarikan, dan *installment* I2 sebanyak 3 *record* pada kelas Lunas. Dengan demikian *installment* I1 dan *installment* I2 menjadi simpul daun (*leaf node*). Dari hasil analisis tersebut dapat digambarkan sebuah pohon keputusan seperti pada Gambar 6.



Gambar 6 Pohon keputusan node 1.1

Dengan memperhatikan pohon keputusan pada Gambar 6 diketahui bahwa semua kasus sudah masuk dalam kelas. Dengan demikian, pohon keputusan pada Gambar 6 merupakan pohon keputusan yang terbentuk.

Pohon keputusan merupakan sekumpulan himpunan aturan "Jika...Maka". Setiap *path* dalam *tree* dihubungkan dengan sebuah aturan, dimana premis terdiri atas sekumpulan *node-node* yang ditemui, dan kesimpulan dari aturannya terdiri atas kelas yang terhubung dengan *leaf* dari *path*. Gambar 7 merupakan aturan yang terbentuk hasil dari proses pembuatan pohon keputusan pada Gambar 6.

ATURAN (RULE)

Jika GAJI = S1, Maka :

Jika ANGSURAN = I1, Maka "TARIKAN"

Jika ANGSURAN = I2, Maka "LUNAS"

Jika GAJI = S2, Maka : "TARIKAN"

KETERANGAN

- S1 = Penghasilan/Gaji antara Rp. 1,5 Jt s.d. Rp. 3 Jt
- S2 = Penghasilan/Gaji antara Rp. 3 Jt s.d. Rp. 6 Jt
- I1 = Nilai Angsuran antara Rp.500rb s.d. Rp.700rb
- I2 = Nilai Angsuran antara Rp.700rb s.d. Rp.1,5 Jt

Gambar 7 Aturan keputusan hasil ilustrasi pembuatan pohon keputusan

Dari Gambar 7 diperoleh aturan bahwa nasabah calon debitur baru yang diperkirakan mampu melunasi kredit (potensial kredit lunas) adalah dengan kriteria penghasilan antara Rp. 1.500.000 hingga Rp. 3.000.000, dan nilai angsuran bulanan antara Rp. 700.000 hingga Rp. 1.500.000.

4. KESIMPULAN

Berdasarkan hasil pembahasan diatas, maka didapatkan kesimpulan :

- 1) Telah dibangun sebuah *prototype* aplikasi yang mampu menghasilkan pohon keputusan dengan algoritma C4.5 untuk memudahkan dalam sistem penilaian kelayakan calon debitur baru. Nilai kelayakan dilihat dari nilai atribut status kredit debitur. Bila nilai status kredit adalah lunas berarti diperkirakan calon debitur baru mampu melunasi kredit (layak), akan tetapi jika nilai status kredit adalah tarikan berarti diperkirakan calon debitur baru tidak mampu melunasi kredit (tidak layak).
- 2) Teknik klasifikasi data mining menggunakan algoritma C4.5 untuk kasus data latih nasabah debitur pembiayaan kendaraan sepeda motor memberikan hasil akurasi dan performa yang cukup baik. Terbukti sistem telah mampu menghasilkan keputusan bahwa data valid di atas 70% untuk semua skenario kasus. Selain itu, dalam pembuatan pohon keputusan (*generate tree*) dan aturan keputusan (*generate rules*) dibutuhkan waktu yang cukup cepat, yakni tidak lebih dari 15 menit untuk setiap skenario pengujian.

5. SARAN

Penelitian yang dilakukan masih terdapat kekurangan yang memerlukan penelitian lebih lanjut guna menghasilkan hasil penelitian yang lebih baik lagi. Oleh karena itu untuk penelitian selanjutnya dapat menggunakan atribut *predictor* dan data yang lebih banyak dan variatif. Penelitian ini dapat pula dilanjutkan dengan menggunakan kombinasi metode rekomendasi yang lain atau dapat menggunakan bahasa pemrograman yang lain..

DAFTAR PUSTAKA

- [1] Berry, Michael J.A dan Gordon S. Linoff., 2004, *Data Mining Techniques for Marketing, Sales, Customer Relationship Management*, Second Edition, Willey Publishing, Inc.
- [2] Chen, Y., Dai, L., dan Chen, X,Q., 2008, GATS-C4.5: An Algorithm for Optimizing Features in Flow Classification, *EEE Communications Society subject matter experts for publication in the IEEE CCNC 2008 proceedings.*, 466-470.
- [3] Gorunescu, F., 2011, *Data Mining Concept Model and Techniques*, Berlin: Springer. ISBN 978-3-642-19720-8
- [4] Han, J., and Kamber, M. 2006, *Data Mining Concept and Tehniques*. San Fransisco: Morgan Kauffman. ISBN 13: 978-1-55860-901-3
- [5] Jaber, K.M, Abdullah, R., Rashid N., 2012, HDT-HS: A Hybrid Decision Tree/Harmony Search Algorithm for Biological Datasets, 2012 International Conference on Computer dan Information Science (ICCIS), 341-345.
- [6] Kornienko, Y., Borisov, A, 2003, Investigation of a Hybrid Algorithm for Decision Tree Generation, *IEEE Inrernational Workshop 00 Infelligem Dah Acquisition and Advanced Coquting System: Techlogy and Applications, Lviu. Ukraine*, Institute of Information Technology, Riga Technical University, Latvia, 8-10 September 2003, 63-68
- [7] Larose, D.T., 2005, *Discovering knowledge in data : An Introduction to Data Mining*, JohnWiley and Sons, Inc., New Jersey.
- [8] Liu, C., dan Jiang, Q., 2009, Mixed Financial Forecasting Index System Construct and Financial Forecasting Study on the C4.5 Decision Tree, *School of Information*

*Management and Engineering Shanghai University of Finance and Economics Shanghai ,
China*

- [9] Pressman, R.S., 1997, *Software Engineering: A Practitioner's Approach*, The McGraw-Hill Companies, Inc.
- [10] Witten, I. H., Frank, E., Hall, M. A., 2011, *Data Mining Practical Machine Learning Tools and Techniques* (3rd ed). USA: Elsevier