

Analisis Kinerja *Decision Tree* C4.5 dalam Prediksi Potensi Pelunasan Kredit Calon Debitur

Bambang Hermanto¹, Azhari SN², Fajri Profesio Putra³

¹ Jurusan Ilmu Komputer FMIPA Universitas Lampung

² Jurusan Ilmu Komputer dan Elektronika, FMIPA UGM Yogyakarta

³ Jurusan Teknik Informatika, Politeknik Negeri Bengkalis

e-mail: ¹bbg.hermanto@gmail.com, ²arisn.softcomp@gmail.com, ³fajri@polbeng.ac.id

Abstrack - In an effort to anticipate the occurrence of errors in the selection of prospective borrowers while improving the quality of customer service, finance companies need decision making tools that simplify and speed up the process of predicting prospective borrowers who are able to pay off credit. The study discusses the application design process in constructing decision tree using C4.5 algorithm and utilizing a group of training data of motorcycle financing debtor, then interpreted in the form of decision rule as a reference in estimating potential loan repayment of debtor. The test results through 5 categories of tests performed in the process of the resulting tree required an average time of 112 seconds with the fastest time obtained in the first test category with the amount of data 3000 records worth 9 seconds. While in the process of generate rule it takes an average time of 1.78 seconds with the fastest time obtained in the first test category with the amount of data 3000 records worth 1.23 seconds. Comparison of the amount of data in each test category affects the value of execution time, the more data make longer the process of generating trees and rules. In the data accuracy test obtained the average percentage of data accuracy value of 51.2% with the highest gain in the first test category with total data 3000 records worth 54%.

Keywords: debtor, credit, c4.5, decision tree

Intisari -. Dalam upaya mengantisipasi terjadinya kesalahan dalam pemilihan calon debitur sekaligus meningkatkan kualitas layanan konsumen, perusahaan pembiayaan membutuhkan alat bantu pengambilan keputusan sehingga mempermudah dan mempercepat proses prediksi calon debitur yang mampu melunasi kredit. Penelitian membahas proses rancangan bangun aplikasi dalam membangun pohon keputusan menggunakan algoritma C4.5 dan memanfaatkan sekelompok data latih debitur pembiayaan kendaraan sepeda motor, kemudian diinterpretasikan dalam bentuk aturan keputusan sebagai acuan dalam memperkirakan potensi pelunasan kredit calon debitur. Hasil pengujian melalui 5 kategori uji yang dilakukan dalam proses *generate tree* dibutuhkan rata-rata waktu 112 detik dengan perolehan waktu tercepat pada kategori uji pertama dengan jumlah data 3000 *record* senilai 9 detik. Sedangkan dalam proses *generate rules* dibutuhkan rata-rata waktu 1,78 detik dengan perolehan waktu tercepat pada kategori uji pertama dengan dengan jumlah data 3000 *record* senilai 1,23 detik. Perbandingan jumlah data disetiap kategori uji mempengaruhi nilai *execution time*, makin banyak datanya maka semakin lama untuk proses *generate tree* dan *rules*. Pada pengujian akurasi data diperoleh prosentase rata-rata nilai akurasi data 51,2% dengan perolehan tertinggi pada kategori uji pertama dengan total data 3000 *record* senilai 54%.

Kata kunci: debitur, kredit, C4.5, pohon keputusan

I. PENDAHULUAN

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 training data debitur pembiayaan kendaraan sepeda motor. Pohon keputusan tersebut kemudian diinterpretasikan kedalam bentuk aturan-aturan keputusan (rules) yang dapat dipahami dan digunakan sebagai acuan dalam memperkirakan potensi pelunasan calon debitur sebagai salah satu kriteria kelayakan calon debitur.

Menurut [1] pohon keputusan (decision tree) dapat membagi kumpulan data yang besar (training data) 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 [1], 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*, *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.

Pada penelitian ini dilakukan analisis untuk kemampuan algoritma C45 dan pohon keputusan dalam memprediksi potensi pelunasan kredit sebagai salah satu kriteria kelayakan calon debitur, sehingga bermanfaat bagi manajer perusahaan sebagai pendukung keputusan penilaian kelayakan calon debitur baru.

II. SIGNIFIKASI STUDI

A. Studi Literatur

Klasifikasi adalah proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui [4]. Algoritma klasifikasi yang banyak digunakan secara luas, yaitu *Decision/classification trees*, *Bayesian classifiers/ Naïve Bayes classifiers*, *Neural networks*, *Analisa Statistik*, *Algoritma Genetika*, *Rough sets*,

k-nearest neighbor, *Metode Rule Based*, *Memory based reasoning*, dan *Support vector machines (SVM)*.

Klasifikasi data terdiri dari 2 langkah proses. Pertama adalah *learning* (fase training), dimana algoritma klasifikasi dibuat untuk menganalisa data training lalu direpresentasikan dalam bentuk rule klasifikasi. Proses kedua adalah klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari rule klasifikasi [4].

Proses klasifikasi didasarkan pada empat komponen [3] :

1. Kelas. Merupakan variabel dependen yang berupa kategorikal yang merepresentasikan 'label' yang terdapat pada objek. Contohnya: resiko penyakit jantung, resiko kredit, customer loyalty, jenis gempa.
2. *Predictor*. Merupakan variabel independen yang direpresentasikan oleh karakteristik (atribut) data. Contohnya: merokok, minum alkohol, tekanan darah, tabungan, aset, gaji.
3. Training dataset. Merupakan satu set data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan *predictor*.
4. Testing dataset. Berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi.

Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami, juga dapat diekspresikan dalam bentuk bahasa basis data seperti *Structured Query Language* untuk mencari *record* pada kategori tertentu. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Karena pohon keputusan

memadukan antara eksplorasi data dan pemodelan, pohon keputusan sangat bagus sebagai langkah awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain [7].

Secara umum Algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut [7]:

- a. Pilih atribut sebagai akar
- b. Buat cabang untuk masing-masing nilai
- c. Bagi kasus dalam cabang
- d. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus seperti yang tertera berikut:

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

Keterangan:

S : Himpunan kasus

A : Variabel penentu

n : Jumlah partisi atribut A

|Si| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

Sebelum mendapatkan nilai Gain adalah dengan mencari nilai Entropi. Entropi digunakan untuk menentukan seberapa informatif sebuah masukan atribut untuk menghasilkan sebuah atribut. Rumus dasar dari Entropi adalah sebagai berikut:

$$Entropy(X) = \sum_{i=0}^m -p_i * \log_2 p_i \tag{2}$$

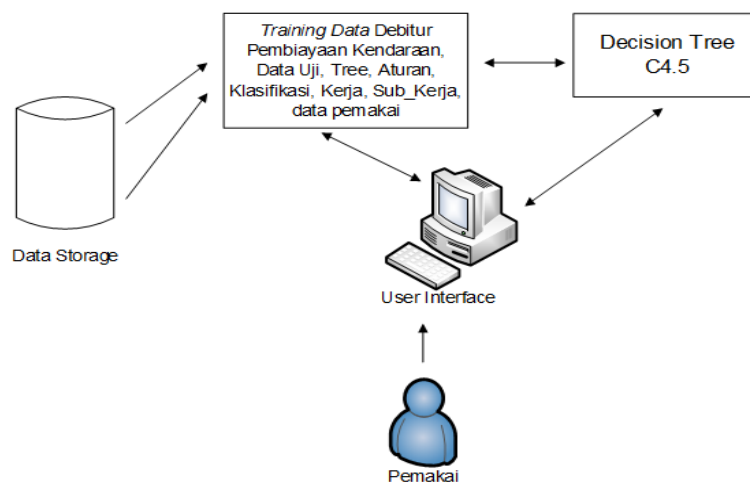
Keterangan

- X : himpunan kasus
- m : jumlah partisi variabel tujuan dari S
- pi : probabilitas kasus dalam partisi ke-i

III. HASIL DAN PEMBAHASAN

A. Arsitektur System

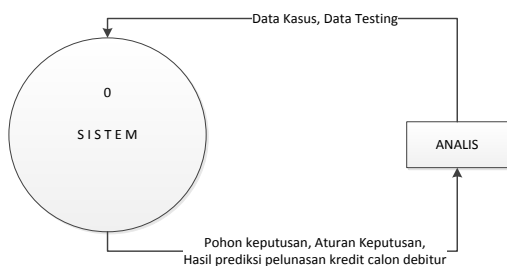
Sistem yang dibangun menggunakan arsitektur sistem database tunggal, artinya database dan aplikasi diletakkan pada computer yang sama atau tidak berada dalam lingkup jaringan computer [9]. Data yang diproses diantaranya adalah data latih (training) data) dan data calon debitur baru sebagai data uji (testing data) dengan parameter jenis pekerjaan, besar penghasilan, tenor, nilai angsuran dan status kredit debitur. Data latih (training data) akan dimanfaatkan sebagai data masukan untuk proses pembuatan pohon keputusan menggunakan algoritma C4.5. Hasilnya berupa pohon keputusan yang kemudian diinterpretasikan dalam aturan keputusan yang dapat digunakan untuk memperkirakan apakah calon debitur baru mampu melunasi kredit. Antarmuka (user interface) disediakan untuk menghubungkan user dengan aplikasi, dalam mengelola data dan mengetahui output sistem. Arsitektur sistem dibuat terlihat pada Gambar 1.



Gambar 1. Arsitektur Sistem

B. Diagram Konteks

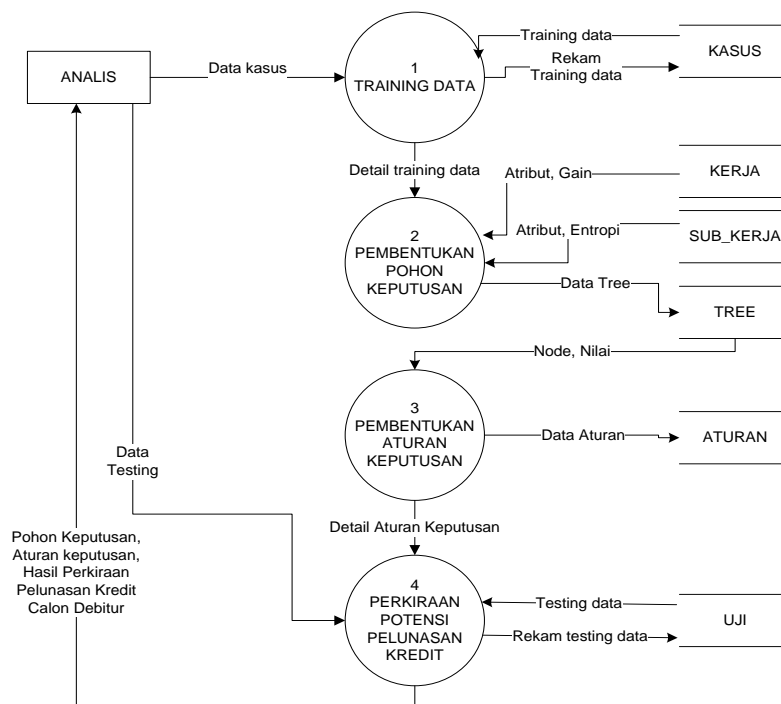
Diagram konteks sistem prediksi potensial pelunasan kredit calon debitur baru terlihat pada Gambar 2. Sistem melibatkan seseorang *user* yang memiliki wewenang dalam melakukan analisis kredit calon debitur termasuk wewenang untuk input *training data* dan *testing data*. Dari sejumlah *training data*, pohon keputusan akan dibentuk (*generate*) menggunakan metoda algoritma C4.5 oleh sistem yang kemudian diinterpretasikan kedalam bentuk aturan-aturan (*rules*) keputusan. Dengan memanfaatkan aturan-aturan keputusan, maka dapat diperkirakan potensi pelunasan kredit sebagai salah satu syarat kelayakan calon debitur.



Gambar 2. Diagram Konteks

C. Diagram Level 1

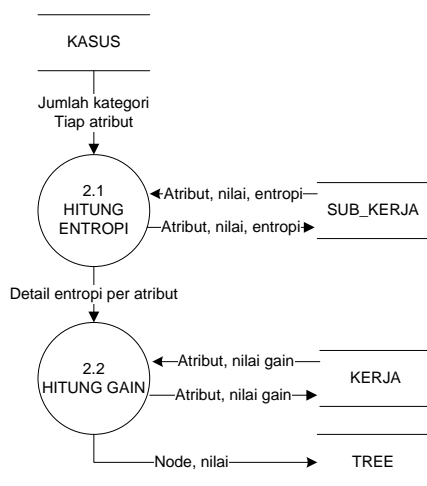
Data flow diagram (DFD) level 1 memiliki beberapa proses diantaranya proses pembentukan pohon keputusan, proses pembentukan aturan keputusan dan hasil prediksi potensi pelunasan calon debitur. *Training data* adalah data input yang akan digunakan untuk proses pembuatan pohon keputusan. Kemudian diinterpretasikan kedalam bentuk aturan-aturan keputusan. Proses prediksi potensial pelunasan kredit merupakan proses pengujian untuk memperkirakan apakah calon debitur baru (*testing data*) yang ditentukan mampu melunasi kredit pembelian kendaraan, dan untuk mengetahuinya dilihat dari nilai parameter status kreditnya. Bila status kredit bernilai lunas berarti diperkirakan calon debitur tersebut mampu melunasi kredit, akan tetapi jika nilai status kredit adalah tidak lunas berarti diperkirakan calon debitur tersebut tidak mampu melunasi kredit. Proses perkiraan kredit lunas adalah proses yang berfungsi memberikan keputusan kepada setiap *testing data* yang masuk. DFD level 1 dapat dilihat pada Gambar 3.



Gambar 3. Data Flow Diagram Level 1

D. Diagram Level 2 Proses 2

DFD Level 2 Proses 2 menggambarkan subproses yang terjadi pada proses pembentukan pohon keputusan yang terdiri dari proses hitung entropi dan hitung gain. Proses hitung entropi berfungsi untuk menghitung nilai entropi total dan masing-masing atribut. Proses hitung gain berfungsi untuk menghitung nilai *gain* untuk masing-masing atribut. Pembentukan pohon keputusan (*Generate tree*) dimulai dengan menghitung jumlah total data dan menghitung jumlah data kredit lunas dan tidak lunas untuk tiap atribut di masing-masing kategori. Proses dilanjutkan dengan menghitung nilai entropi total, entropi tiap atribut, dan nilai *gain* untuk tiap kategori. *Data flow diagram* level 2 proses 2 terlihat pada Gambar 4.



Gambar 4. Data Flow Diagram Level 2 Proses 2

E. Hasil

Proses pengujian akan memanfaatkan sejumlah data latih (*training data*) dan data uji (*testing dataset*) yang diproses melalui *software* aplikasi yang sudah dibangun. Data dalam proses pengujian akan dibagi dalam 5 kategori uji berdasarkan pemetaan data pengujian yang ditetapkan. Hasil pengujian yang dilakukan berupa informasi lama waktu eksekusi (*execution time*) dalam membangun pohon keputusan (*generate tree*) dan aturan keputusan, serta nilai prosentase akurasi data.

TABEL I
PEMETAAN DATA PENGUJIAN

Uji	Total Data	Data Latih	Data Uji
1	3000	2400	600
2	10000	8000	2000
3	25000	20000	5000
4	50000	40000	10000
5	65000	52000	13000

Proses pemilihan data

Data yang akan digunakan ditentukan secara acak *by user* yang terdiri dari 5 kategori uji, yaitu:

- 3000 *record* dengan pembagian 2400 *record* data latih terdiri dari 1200 *record* status lunas dan 1200 *record* status kredit tidak lunas, 600 *record* data uji terdiri dari 300 *record* status kredit lunas dan 300 *record* status kredit tidak lunas.
- 10000 *record* dengan pembagian 8000 *record* data latih terdiri dari 4000 *record* status kredit lunas dan 4000 *record* status kredit tidak lunas, 2000 *record* data uji terdiri dari 1000 *record* status kredit lunas dan 1000 *record* status kredit tidak lunas.
- 25000 *record* dengan pembagian 20000 *record* data latih terdiri dari 10000 *record* status kredit lunas dan 10000 *record* status kredit tidak lunas, 5000 *record* data uji terdiri dari 2500 *record* status kredit lunas dan 2500 *record* status kredit tidak lunas.
- 50000 *record* dengan pembagian 40000 *record* data latih terdiri dari 20000 *record* status kredit lunas dan 20000 *record* status kredit tidak lunas, 10000 *record* data uji terdiri dari 5000 *record* status kredit lunas dan 5000 *record* status kredit tidak lunas.
- 65000 *record* dengan pembagian 52000 *record* data latih terdiri dari 26000 *record* status kredit lunas dan 26000 *record* status kredit tidak lunas, 13000 *record* data uji terdiri dari 6500 *record* status kredit lunas dan 6500 *record* status kredit tidak lunas.

F. Proses pengujian dan hasil

1. Menghitung lama waktu proses pembentukan pohon keputusan.

TABEL II
HASIL PEROLEHAN WAKTU EKSEKUSI PEMBUATAN POHON KEPUTUSAN

Uji	Total Data	Waktu Eksekusi Generate Tree
1	3000	9 Detik
2	10000	43 Detik
3	25000	100 Detik
4	50000	175 Detik
5	65000	235 Detik



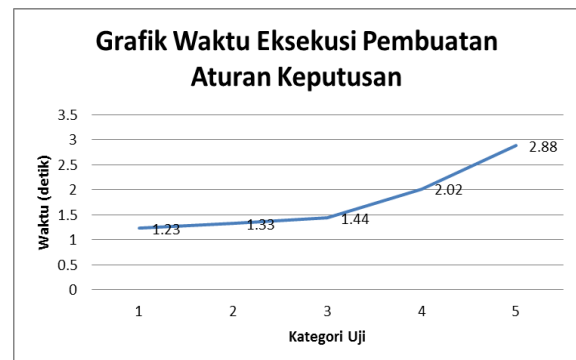
Gambar 5. Grafik waktu eksekusi pembuatan pohon keputusan

Dari Tabel II dan Gambar 5 dapat disimpulkan bahwa makin besar jumlah data percobaan maka waktu eksekusi yang dibutuhkan relatif lebih lama untuk membuat suatu pohon keputusan. Waktu rata-rata yang dibutuhkan untuk membangun pohon keputusan dari 5 kategori uji adalah 112 detik. Waktu tercepat terjadi pada uji pertama dengan total data 3000 record. Hasil dipengaruhi oleh makin banyak data maka makin banyaknya pula proses iterasi data dalam proses pembuatan semua simpul pohon menggunakan algoritma C4.5 dimana didalamnya juga terdapat proses perhitungan nilai entropi dan gain masing-masing simpul.

2. Menghitung lama waktu proses pembentukan aturan keputusan

TABEL III
HASIL PEROLEHAN WAKTU EKSEKUSI PEMBUATAN ATURAN KEPUTUSAN

Uji	Total Data	Waktu Eksekusi Generate Rules
1	3000	1.23 Detik
2	10000	1.33 Detik
3	25000	1.44 Detik
4	50000	2.02 Detik
5	65000	2.88 Detik



Gambar 6. Grafik waktu eksekusi pembuatan aturan keputusan

Dari Tabel III dan Gambar 6 dapat disimpulkan bahwa waktu yang dibutuhkan untuk membuat aturan cukup singkat yakni rata-rata 1,78 detik. Waktu tercepat diperoleh pada kategori uji pertama dengan jumlah data 3000 record. Berbeda dengan proses pembentukan pohon keputusan karena pada proses generate rules sistem tidak perlu melakukan proses perhitungan nilai gain dan entropi melainkan hanya membaca record data pada tabel database aturan yang menyimpan informasi guna menterjemahkan dari tree menjadi rules (aturan keputusan).

3. Menghitung prosentase akurasi data uji yang sesuai (benar)

Untuk mengetahui akurat tidaknya dengan cara membandingkan antara hasil nilai status kredit melalui *software* aplikasi yang dibangun dengan hasil nilai status kredit sesuai data riil. Bila hasilnya sama, maka data tersebut dianggap benar (akurat). Dari Tabel IV dapat disimpulkan bahwa dari kelima skenario uji akurasi data

menghasilkan rata-rata nilai akurasi data yang bernilai benar (akurat) adalah 51,2%. Paling tinggi diperoleh dari kategori uji pertama dengan jumlah data 3000 record.

TABEL IV
DATA PROSENTASE UJI AKURASI DATA BERNILAI SESUAI (BENAR)

UJI	TOTAL DATA	PROSENTASE HASIL BENAR
1	3000	54 %
2	10000	50 %
3	25000	50 %
4	50000	51 %
5	65000	

4. Menghitung Prosentase nilai kelayakan data uji yang tidak sesuai (salah)

TABEL V
DATA PROSENTASE UJI AKURASI DATA BERNILAI TIDAK SESUAI (SALAH)

Uji	Total Data	Prosentase Hasil Salah
1	3000	46 %
2	10000	50 %
3	25000	50 %
4	50000	49 %
5	65000	49 %

Dari Tabel V kelima skenario uji akurasi data menghasilkan rata-rata nilai akurasi data yang bernilai tidak akurat adalah 48,8%. Paling rendah diperoleh dari kategori uji pertama dengan jumlah data 3000 record.

IV. KESIMPULAN

Berdasarkan hasil pembahasan diatas, maka didapatkan kesimpulan:

1. Telah dibangun rancang bangun aplikasi dengan menerapkan algoritma C4.5 pada pohon keputusan untuk memprediksi potensi pelunasan kredit sebagai salah satu kriteria kelayakan calon debitur baru.
2. Dari hasil pengujian 5 kategori uji yang dilakukan dalam proses *generate tree* dibutuhkan rata-rata waktu 112 detik

dengan perolehan waktu tercepat pada kategori uji pertama dengan dengan jumlah data 3000 record yang terdiri dari 2400 *record* data latih dan 600 *record* data uji senilai 9 detik. Sedangkan dalam proses *generate rules* dibutuhkan rata-rata waktu 1,78 detik dengan perolehan waktu tercepat pada kategori uji pertama dengan dengan jumlah data 3000 *record* yang terdiri dari 2400 *record* data latih dan 600 *record* data uji senilai 1,23 detik. Perbandingan jumlah data disetiap kategori uji mempengaruhi nilai *execution time*, makin banyak datanya maka semakin lama untuk proses *generate tree* dan *rules*. Pada pengujian akurasi data diperoleh prosentase rata-rata nilai akurasi data 51,2% dengan perolehan tertinggi pada kategori uji pertama dengan total data 3000 *record* terdiri dari 2400 *record* data latih dan 600 *record* data uji senilai 54%.

Penelitian yang dilakukan masih terdapat beberapa kekurangan diantaranya keterbatasan jumlah *training data* dan atribut prediktor sehingga berpengaruh pada kualitas pohon keputusan yang dihasilkan sebagai dasar aturan untuk memprediksi potensi pelunasan kredit sebagai salah satu kriteria kelayakan calon debitur baru. Diharapkan pada penelitian selanjutnya dapat menggunakan *training data* yang lebih variatif dan jumlah atribut *predictor* lebih, serta dilengkapi dengan proses pemangkasan pohon (*prunning*) untuk lebih meningkatkan tingkat akurasi.

REFERENSI

[1] Berry Michael, J.A., dan Linoff, G.S., 2004, *Data Mining Techniques for Marketing, Sales, Customer Relationship Management*, Second Edition, Willey Publishing, Inc.

[2] Chen, Y., Dai, L., dan Cheng, 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., dan Rashid N., 2012, HDT-HS: A Hybrid Decision Tree/Harmony Search Algorithm for Biological Datasets, 2012 International Conference on Computer dan Information Science (ICIS), 341-345.
- [6] Kornienko, Y., dan 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, *International Conference on Management and Service Science, MASS '09*, pp.1-4, 20-22 Sept. 2009
- [9] Pressman, R.S., 1997, *Software Engineering: A Practitioner's Approach*, The McGraw-Hill Companies, Inc.
- [10] Witten, I. H., Frank, E., dan Hall, M. A., 2011, *Data Mining Practical Machine Learning Tools and Techniques* (3rd ed). USA: Elsevier