JuSIN
Jurnal Sistem Informasi

Website: ojs.itb-ad.ac.id/index.php/JUSIN
E-ISSN: 2797-8516

METODE DECISION TREE DALAM KLASIFIKASI KREDIT PADA NASABAH PT BANK PERKREDITAN RAKYAT (STUDI KASUS : PT BPR LUBUK RAYA MANDIRI)

Diana Yusuf [1], Ellya Sestri [2 (*)]

[1]*ITB Ahmad Dahlan, Jakarta*
[2]*ITB Ahmad Dahlan, Jakarta*

Abstract

Data mining is a new technology that has been successfully applied in many fields. Many problems are solved by data mining algorithms C4.5 as a supporter. Classification is one of the methods contained in data mining and not a few researchers who use the classification methods in solving problems. Credit analysis will be done with the digging data against existing data customer credits based on atribut-atributnya with tekanik data mining algorithm C4.5. The algorithm C4.5 is itself a group of decision tree algorithm. This algorithm has input in the form of training and samples. Data mining technique used to classify loans with algorithm C4.5. Analysis and processing of data use applied tools of RapidMiner v 7.3

Kata Kunci: *Data mining, Classification, Algorithm C4.5, RapidMiner v.7.3, Decision Tree, Kredit*

Juli – Desember 2020, Vol 1 (1) : hlm 21-28
©2020 Institut Teknologi dan Bisnis Ahmad Dahlan.
All rights reserved.

(*) Korespondensi: dianayusuf01@gmail.com (Diana Yusuf), ellyasestri.24@gmail.com (Ellya Sestri)

PENDAHULUAN

Perbankan ialah lembaga keuangan yang kegiatan utamanya menghimpun dana masyarakat dan menyalurkannya kepada masyarakat dalam bentuk kredit serta memberikan jasa-jasa keuangan lainnya. Kredit merupakan sumber utama penghasilan sebuah bank, sekaligus sumber resiko operasi bisnis terbesar. Kredit macet atau kredit bermasalah merupakan permasalahan yang mendarah daging dalam perbankan, karena kredit macet dalam jumlah besar dapat mengganggu sendi kehidupan ekonomi dan akan menggerogoti dana operasional bank itu sendiri, serta membahayakan likuiditas keuangan bank dan menurunkan kepercayaan masyarakat dan luar negeri terhadap profesionalisme pengelolaan bisnis perbankan nasional.

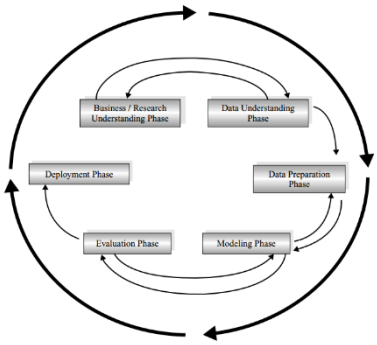
Algoritma C 4.5 telah digunakan untuk dalam berbagai penelitian, di antaranya (Iriadi dan Nuraeni, 2016) menggunakan algoritma C 4.5 untuk memprediksi kelayakan kredit pada Bank Mayapada Jakarta. Selain itu Gu dan Zhou (2012) dalam penelitiannya membahas mengenai bagaimana melakukan analisis dari setiap ujian para siswa dimulai dari semester pertama serta parameter-parameter lainnya, guna memprediksi skor yang didapatkan untuk masuk perguruan tinggi. Menurut mereka hal ini dapat membantu dalam perbaikan proses mengajar pada sekolah menengah dan kualitas mengajar menjadi lebih baik. Adi Sucipto (2015) menggunakan algoritma C 4.5 dalam memprediksi kredit macet melalui perilaku nasabah pada koperasi simpan pinjam. Kemudian (David Hartanto Kamagi *et al*, 2014), melakukan penelitian tentang bagaimana memprediksi tingkat kelulusan mahasiswa.

Menganalisis data mentah secara manual dan menemukan informasi yang benar dalam sebuah perusahaan adalah proses yang sulit. Namun teknik secara otomatis mendeteksi pola yang relevan dari data mentah tersebut dengan menggunakan algoritma yang ada dalam . Metode pohon keputusan adalah metode terbaik dan umumnya sering digunakan (Sathyadevan dan Nair, 2015).

METODE

Metode penelitian yang digunakan *Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang dikembangkan tahun 1996 oleh analis dari beberapa industri seperti Daimler Chrysler, SPSS dan NCR. CRISP DM menyediakan standar proses *Data Mining* sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian.

Dalam CRISP DM, sebuah proyek *Data Mining* memiliki siklus hidup yang terbagi dalam enam tahapan (Gambar 2.3). Keseluruhan tahap yang berurutan tersebut bersifat adaptif. Tahap berikutnya dalam urutan bergantung kepada keluaran dari tahap sebelumnya. Hubungan penting antar tahapan digambarkan dengan panah. Sebagai contoh, jika proses berada pada tahap *modeling*. Berdasar pada perilaku dan karakteristik model, proses mungkin harus kembali pada tahap *data preparation* untuk perbaikan lebih lanjut terhadap data atau berpindah maju kepada tahap *evaluation*.



1. Fase Pemahaman Bisnis (*Bussiness / Research Understanding Phase*)
   a. Penentuan tujuan proyek dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan.
   b. Menerjemahkan tujuan dan batasan menjadi formula dari permasalahan *data mininig*.
   c. Menyiapkan strategi awal untuk mencapai tujuan.
2. Fase Pemahaman Data (*Data Understanding Phase*)
   a. Mengumpulkan data.
   b. Menggunakan analisis penyelidikan data untuk mengenali lebih lanjut data dan pencarian pengetahuan awal.
   c. Mengevaluasi kualitas data.
   d. Jika diinginkan, pilih sebagian kecil group data yang mungkin mengandung pola dari permasalahan.
3. Fase Pengolahan Data (*Data Preparation Phase*)
   a. Siapkan data dari awal, kumpulan data yang akan digunakan untuk keseluruhan tahapan berikutnya. Tahap ini merupakan pekerjaan berat yang perlu dilaksanakan secara intensif.
   b. Pilih kasus dan variabel yang ingin dianalisis dan yang sesuai analisis yang akan dilakukan.
   c. Lakukan perubahan pada beberapa variabel jika dibutuhkan.
   d. Siapkan data awal sehingga siap untuk perangkat pemodelan.
4. Fase Pemodelan (*Modeling Phase*)
   a. Pilih dan aplikasikan teknik pemodelan yang sesuai.
   b. Kalibrasi aturan model untuk mengoptimalkan hasil.
   c. Perlu diperhatikan bahwa beberapa teknik mungkin untuk digunakan pada permasalahan *Data Mining* yang sama.
   d. Jika diperlukan, proses dapat kembali ke tahap pengolahan data untuk menjadikan data ke dalam bentuk yang sesuai dengan spesifikasi kebutuhan teknik *Data Mining* tertentu.

Klasifikasi

Klasifikasi merupakan proses pembelajaran / pelatihan suatu fungsi tujuan (target) *f* yang memetakan tiap himpunan atribut x ke satu dari label kelas y yang didefenisikan sebelumnya. Pekerjaan pelatihan tersebut akan menghasilkan suatu model yang kemudian disimpan sebagai memori.

Klasifikasi juga merupakan suatu pekerjaan menilai objek data untuk memasukkan ke dalam kelas dari sejumlah kelas yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan yaitu, pembangunan model sebagai prototype untuk disimpan sebagai memori dan penggunaan model tersebut untuk melakukan pengenalan/klasifikasi/prediksi/ pada suatu objek data lain agar diketahui dikelas mana objek data tersebut dalam model yang sudah disimpannya (Susanti, 2014).

Decision Tree

Konsep dasar dari Decision Tree adalah mengubah data menjadi pohon keputusan dengan aturan-aturannya. Atribut yang dipilih akan menghasilkan partisi dengan data yang lebih seragam dan dapat menghasilkan pohon keputusan yang sederhana dengan perulangan yang sedikit. Sebuah pohon keputusan terdiri dari sekumpulan aturan yang bertujuan untuk membagi sejumlah populasi yang heterogen menjadi lebih kecil dan lebih homogen dengan memperhatikan variabel tujuan.

Metode ini adalah sebuah stuktur pohon dimana setiap node pohon mempresentasikan atribut yang telah diuji, setiap cabang merupakan suatu pembagian hasil uji, dan node daun

(leaf) mempresentasikan kelompok kelas tertentu. Level node teratas dari sebuah Decision Tree adalah akar (root) yang biasanya berupa atribut yang paling memiliki pengaruh terbesar pada suatu kelas tertentu. Pada umumnya *Decision Tree* melakukan strategi pencarian secara top down untuk solusinya. Pada proses mengklasifikasi data yang tidak diketahui, nilai atribut akan diuji dengan cara melacak jalur dari node akar *(root)* sampai node akhir (daun) dan kemudian akan diprediksi kelas yang dimiliki oleh suatu data baru tertentu. Secara singkat *Decision Tree* merupakan salah satu metode klasifikasi pada *Text mining* (Rusito et al, 2016).

Algoritma C.45

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami dan juga dapat diekspresikan dalam bentuk bahasa basis data seperti Structured Query Language untuk mencari record pada kategori tertentu (Rusito, et all, 2016). Secara umum, algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

1. Pilih atribut sebagai akar.
2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang.
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Berikut pengembangan algoritma C4.5 dalam pembuatan pohon keputusan sebagai berikut (Rusito, et all, 2016) :

1. Menyiapkan *data training*. *Data training* biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan kedalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung dahulu nilai *entropy*. Untuk menghitung nilai *entropy* digunakan rumus :
   $Entropy\ (S) = \sum_{i=1}^{n} - pi.log2pi$ .....(2.2)
   Di mana
   S = himpunan kasus
   n = jumlah partisi S
   pi = proporsi S_i terhadap S
3. Kemudian hitung nilai *gain* menggunakan rumus :
   $$Gain\ (S,A) = Entropy\ (S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy\ (Si) \quad (2.3)$$
   Di mana
   S = himpunan kasus
   A = fitur
   n = jumlah partisi atribut A
   $|Si|$ = proporsi Si terhadap S
   $|S|$ = jumlah kasus dalam S

4. Ulangi langkah ke-2 hingga semua *record* terpartisi
5. Proses partisi pohon keputusan akan berhenti pada saat :
   a. Semua *record* dalam simpul N mendapat kelas yang sama
   b. Tidak ada atribut di dalam *record* yang dipartisi lagi
   c. Tidak ada *record* di dalam cabang yang kosong

HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini adalah data debitur kredit pada PT BPR Lubuk Raya Mandiri. Setelah memperoleh data mentah kemudian dilakukan proses *Cleaning Data* yang menghasilkan seperti berikut ini.

Tabel 1 Hasil *Cleaning Data*

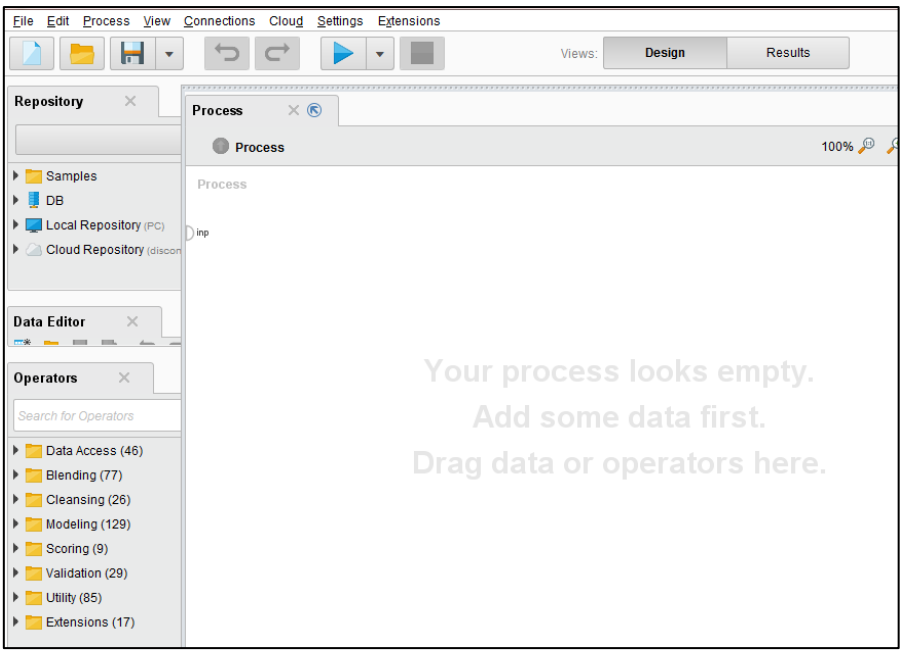
No	No Rek	NAMA	Jen	Tgl.Akad	Jt.Tempo	Plafond Awa	Kol	Gol	Sektor	Guna	Jaminan No	Dati
1	140.00xxxx		9	29/12/2018	29/12/2022	10.000.000	4	875	1020	39	10.400.000	3492
2	140.00xxxx		9	27/12/2018	27/06/2021	9.000.000	4	875	1007	10	98.000.000	3492
3	140.00xxxx		6	23/12/2018	23/12/2021	10.000.000	1	875	1009	20	10.200.000	3492
4	140.00xxxx		12	22/12/2018	22/12/2021	58.000.000	1	875	1009	10	85.000.000	3492
5	140.00xxxx		12	22/12/2018	22/12/2020	140.000.000	1	875	1020	39	148.000.000	3492
6	140.00xxxx		12	08/12/2018	08/12/2019	250.000.000	1	875	1007	10	295.000.000	3492
7	140.00xxxx		6	30/11/2018	30/05/2019	500.000.000	1	875	1011	20	632.175.000	3492
8	140.00xxxx		9	22/11/2018	22/05/2022	74.000.000	4	875	1015	10	87.000.000	3492
9	140.00xxxx		9	21/11/2018	21/11/2020	352.000.000	1	875	1009	20	378.000.000	3497
10	140.00xxxx		12	21/11/2018	21/11/2020	50.000.000	1	875	1020	39	63.000.000	3492
11	140.00xxxx		9	18/11/2018	18/11/2019	40.000.000	1	875	1009	20	57.000.000	3403
12	140.00xxxx		6	17/11/2018	17/11/2022	18.000.000	1	875	1009	20	18.500.000	3492
13	140.00xxxx		6	17/11/2018	17/11/2019	8.000.000	1	875	1020	39	8.450.000	3492
14	140.00xxxx		12	15/11/2018	15/11/2023	26.000.000	4	875	1020	39	26.240.000	3492

Selanjutnya dilakukan proses seleksi data yang akan digunakan untuk membuat sebuah pohon keputusan. Data yang sudah mempunyai informasi yang lengkap dalam setiap atribut tersebut diseleksi. Seleksi dilakukan untuk mengelompokkan atribut sesuai dengan informasi yang dibutuhkan. Dari atribut yang sudah diseleksi, maka atribut yang akan digunakan dalam pembuatan pohon keputusan adalah atribut jenis kredit, atribut waktu, atribut plafond, atribut jaminan, dan atribut *collectability*.

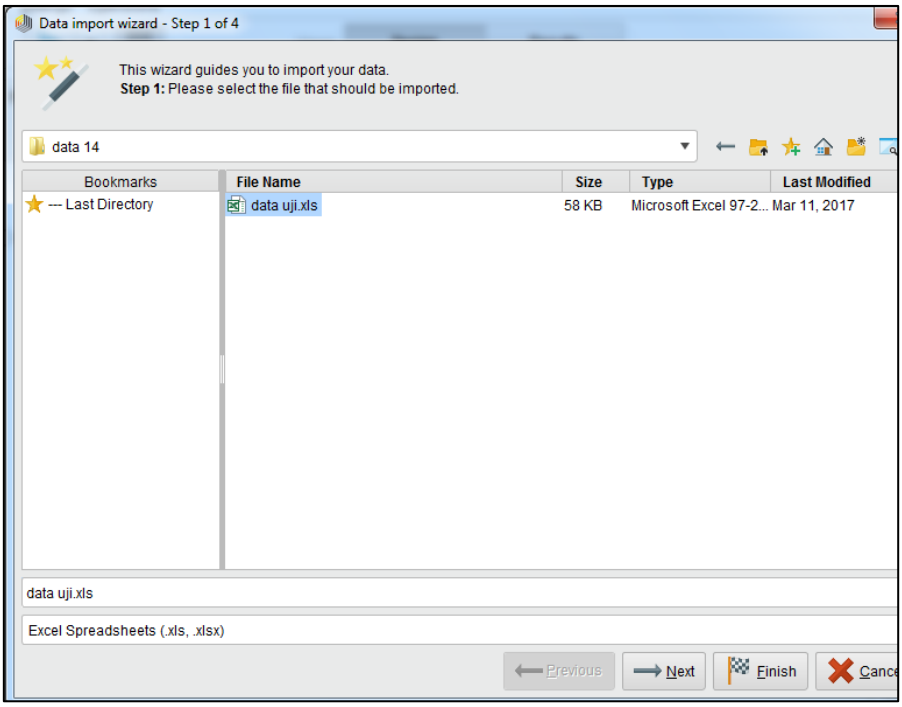
Format data setelah dilakukan pra-proses (*cleaning data* dan *selection data*) dapat dilihat pada tabel berikut ini :

Tabel 2. Format data hasil *cleaning data* dan *selection data*

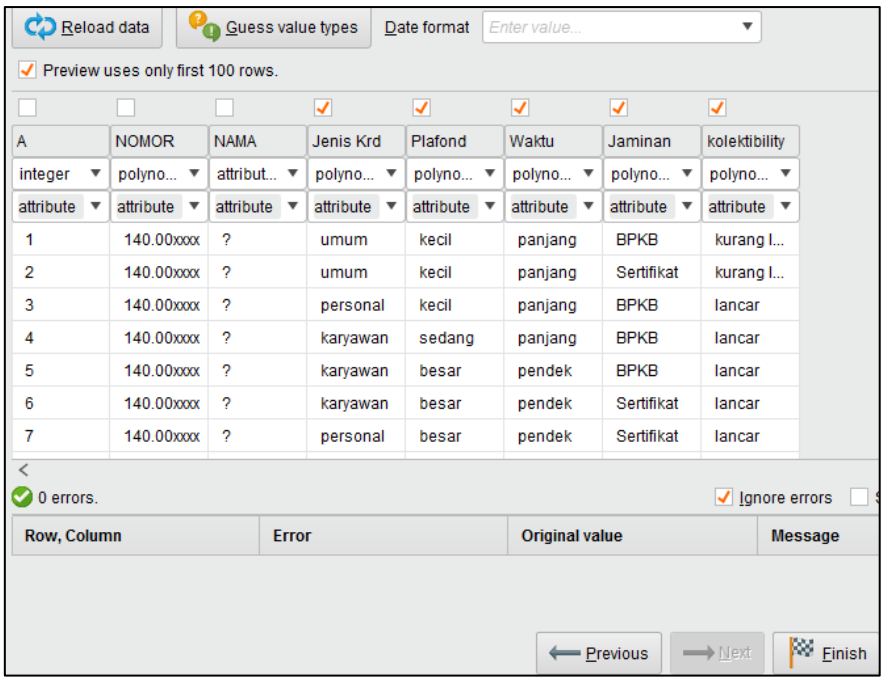
No	No Rek	NAMA	Jenis Krd	Plafond	Waktu	Jaminan	kolektibility
1	140.00xxxx		umum	kecil	panjang	BPKB	kurang lancar
2	140.00xxxx		umum	kecil	panjang	Sertifikat	kurang lancar
3	140.00xxxx		personal	kecil	panjang	BPKB	lancar
4	140.00xxxx		karyawan	sedang	panjang	BPKB	lancar
5	140.00xxxx		karyawan	besar	pendek	BPKB	lancar
6	140.00xxxx		karyawan	besar	pendek	Sertifikat	lancar
7	140.00xxxx		personal	besar	pendek	Sertifikat	lancar
8	140.00xxxx		umum	sedang	panjang	BPKB	kurang lancar
9	140.00xxxx		umum	besar	pendek	BPKB	lancar
10	140.00xxxx		karyawan	sedang	pendek	BPKB	lancar
11	140.00xxxx		umum	sedang	pendek	Sertifikat	lancar
12	140.00xxxx		personal	sedang	panjang	Sertifikat	lancar
13	140.00xxxx		personal	kecil	pendek	BPKB	lancar
14	140.00xxxx		karyawan	sedang	panjang	Sertifikat	kurang lancar



Gambar 1. Tampilan Halaman *Process Rapidminer 7.3*



Gambar 2. Import Data Excel



Reload data | Guess value types | Date format | Enter value...

Preview uses only first 100 rows.

A	NOMOR	NAMA	Jenis Krd	Plafond	Waktu	Jaminan	kolektibility
integer	polyno...	attribut...	polyno...	polyno...	polyno...	polyno...	polyno...
attribute	attribute	attribute	attribute	attribute	attribute	attribute	attribute
1	140.00xxxx	?	umum	kecil	panjang	BPKB	kurang l...
2	140.00xxxx	?	umum	kecil	panjang	Sertifikat	kurang l...
3	140.00xxxx	?	personal	kecil	panjang	BPKB	lancar
4	140.00xxxx	?	karyawan	sedang	panjang	BPKB	lancar
5	140.00xxxx	?	karyawan	besar	pendek	BPKB	lancar
6	140.00xxxx	?	karyawan	besar	pendek	Sertifikat	lancar
7	140.00xxxx	?	personal	besar	pendek	Sertifikat	lancar

0 errors. | Ignore errors

Row, Column	Error	Original value	Message

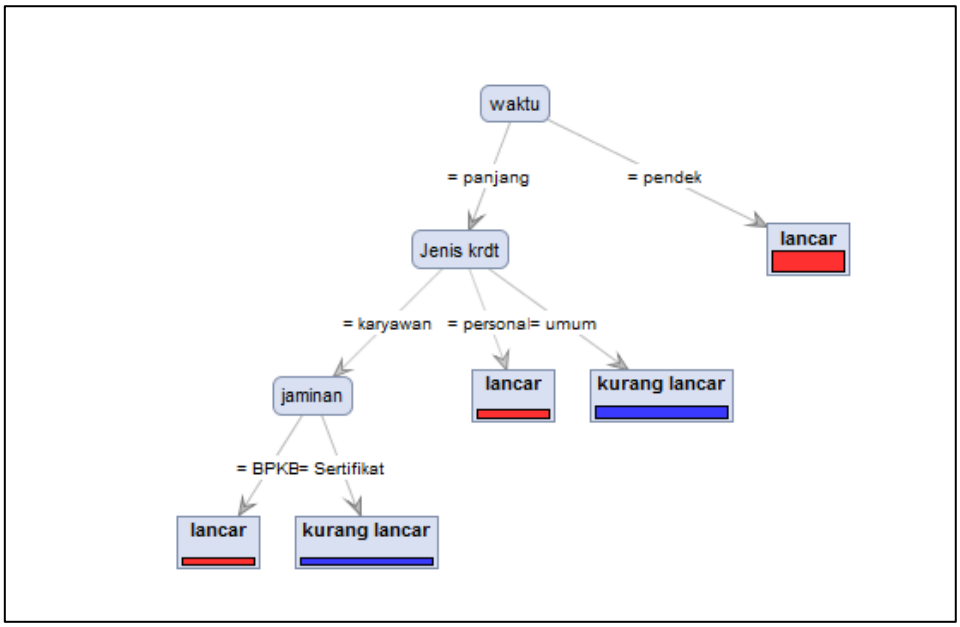
Previous | Next | Finish

Gambar 3. Tampilan Data Import

accuracy: 100.00%

	true kurang lancar	true lancar	class precision
pred. kurang lancar	11	0	100.00%
pred. lancar	0	20	100.00%
class recall	100.00%	100.00%	

Gambar 4. *PerformanceVector*



Gambar 5. *Decision Tree*

KESIMPULAN

1. Algoritma C4.5 dapat digunakan untuk mengklasifikasikan kredit pada nasabah PT. Bank Perkreditan Rakyat (Studi Kasus : PT BPR Lubuk Raya Mandiri)
2. Keputusan *decision tree* tidak hanya untuk keputusan yang bersifat "Lancar" atau "Kurang Lancar" tapi juga bisa digunakan untuk banyak keputusan seperti mengetahui klasifikasi apa yang baik dalam menentukan sebuah kredit.
3. Pohon keputusan dengan menggunakan algoritma C4.5 sangat mudah untuk digunakan, dianalisa dan diterjemahkan.
4. *RapidMiner* merupakan *software* yang sangat baik untuk digunakan dalam membuat *decision tree*. Selain tidak berbayar, *RapidMiner* juga mudah digunakan dan *output* yang dihasilkan juga mudah untuk direpresentasikan.

DAFTAR PUSTAKA

Han, Jiawei., Micheline Kamber dan Jian Pei. (2012). "Data Mining Concepts and Techniques". 3th. USA : Morgan Kaufmann Publisher.1-8

Wei-Li, Jiang. (2011). "Research and Application of Credit Score Based on Decison Tree Model". ICAIC 2011, Part I, CCIS 224, pp. 493-501.

Iriandi, Nandang dan Nia Nuraeni. (2016). "Kajian Penerapan Metode Klasifikasi Data Mining Algoritma C4.5 Untuk Prediksi Kelayakan Kredit Pada Bank Mayapada Jakarta". Vol. II No. 1.

Rusito, *et all*. (2016). "Implementasi Metode Decison Tree dan Algoritma C4.5 untuk Klasifikasi Data Nasabah Bank". Infokam Nomor I Th. XII.

Pudjo, Prabowo., dkk. (2013). "Penerapan Data Mining dengan Matlab".

Susanti. (2014). "Klasifikasi Kredit Menggunakan Metode Decison Tree Pada Nasabah PD BPR BKK Gabus".

Sathyadevan, Shiju., and Remya R. Nair. "Comparative Analysis of Decision Tree Algorithms: ID3, C4.5 and Random Forest". Volume I.

Tampubolon, Kennedi., et all. (2013). "Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan". Volume I. ISSN :2339-210X.

Gu, Ping., and Qi Zhou. (2012). "Student Performance Prediction Based on Improved C4.5 Decision Tree Algorithm". ECICE 2012, AISC 146, pp. 1-8

Sucipto, Adi. (2015). "Prediksi Kredit Macet Melalui Perilaku Nasabah Pada Koperasi Simpan Pinjam dengan Menggunakan Metode Algoritma Klasifikasi C4.5". Volume 6 No 1.

Kim, DaEun. (2006). "Minimizing Structural Risk on Decision Tree Classification". SCI 16, 241-260.