

PENERAPAN DATA MINING UNTUK MEMPREDIKSI PERILAKU NASABAH KREDIT: STUDI KASUS BPR MARCORINDO PERDANA CIPUTAT

Syaiful Anwar

Program Studi Manajemen Informatika
Akademik Manajemen Informatika dan Komputer Bina Sarana Informatika (AMIK BSI)
Jl. Banten No.1 Karawang
<http://www.bsi.ac.id>
sfa_bsi@yahoo.com

ABSTRACT

Rural Bank one of the institutions about providing loans to certain conditions and criteria. Credit Analysis takes time and funds are not cheap so we need an appropriate method for analyzing prospective credit customers. Data Mining is one method that can be used to analyze existing data chunks that can be used to summarize the data provide specific information related to the data. Data Mining classification of a decision tree algorithm C4.5 is used in forming the rules of the statement. Decision tree model was able to improve the accuracy in analyzing the credit worthiness of the proposed prospective credit customers. The richer the information or knowledge contained by the training data, the accuracy of the decision tree will increase. And implementation can be done using one of the Visual Basic programming language.

Keywords: credit customer behavior, Data Mining, C4.5 Algorithm

1. PENDAHULUAN

Bank perkreditan rakyat adalah lembaga yang memberikan kredit kepada nasabah yang membutuhkan dana cepat untuk keperluan nasabah yang membutuhkannya. Secara umum peruntukan dana pinjaman dari bank perkreditan rakyat seperti, pembelian kendaraan, pendidikan, renovasi rumah, resepsi pernikahan, modal usaha dan lain-lain.

Tidak dipungkiri, memberikan pinjaman dana kepada nasabah pasti akan muncul permasalahan-permasalahan, seperti nasabah terlambat membayarkan cicilan dana, penyalahgunaan dana untuk keperluan lain, nasabah gagal mengembangkan usahanya sehingga dapat mengakibatkan dana di bank tidak mengalir atau dapat mengakibatkan kredit macet.

Oleh karena itu, penelitian ini dilakukan untuk membantu menyelesaikan permasalahan tersebut dengan merancang sebuah aplikasi *data mining* yang berfungsi untuk memprediksi perilaku nasabah kredit yang berpotensi melakukan peminjaman (kredit) terhadap bank. BPR Marcorindo Perdana yang berlokasi di Ciputat Tangerang Selatan merupakan tempat yang dipilih oleh peneliti sebagai studi kasus.

Jumlah nasabah yang mengajukan pinjaman dana kredit pada tahun 2009 sebanyak 402 orang dengan nasabah bermasalah sebanyak 48 orang dan pada tahun 2010 jumlah nasabahnya

sebanyak 558 orang dengan nasabah bermasalah sebanyak 72 orang. Data yang diizinkan untuk diteliti adalah data pada tahun 2009.

Untuk memprediksi perilaku nasabah kredit ini diperlukan suatu metode atau teknik yang dapat mengolah data-data yang sudah ada di bank tersebut. Salah satu metodenya dapat menggunakan teknik *data mining*

Dari fungsi *data mining* yang sudah disebutkan di atas akan digunakan dalam penelitian ini. Teknik *data mining* yang diterapkan pada aplikasi yang dibangun adalah klasifikasi sedangkan metode klasifikasi yang digunakan adalah *Decision Tree* (pohon keputusan). Algoritma yang dipakai sebagai algoritma pembentuk pohon keputusannya adalah *Algoritma C4.5*. Adapun data yang di proses dalam penelitian ini adalah data angsuran nasabah kredit BPR Marcorindo Perdana pada tahun 2009 dalam format Microsoft Excel.

Hasil penelitian ini adalah sebuah aplikasi yang dapat memudahkan Bagian Dana BPR Marcorindo Perdana dalam memperoleh target pemasaran kredit di masa mendatang. Dengan adanya aplikasi ini Bank Perkreditan Rakyat dapat menyeleksi dengan mudah calon-calon nasabah yang akan mengajukan dana pinjaman.

2. LANDASAN TEORI/ KERANGKA PEMIKIRAN

2.1. Data Mining

Data Mining merupakan disebut sebagai proses ekstraksi pengetahuan dari data yang besar. Sesuai fungsinya, *Data Mining* adalah proses pengambilan pengetahuan dari volume data yang besar yang disimpan dalam basis data, data *warehouse*, atau informasi yang disimpan dalam repositori (Han & Kamber, 2006). Gartner Group dalam (Larose, 2005) menyebutkan bahwa *data mining* adalah proses menelusuri pengetahuan baru, pola dan tren yang dipilih dari jumlah data yang besar yang disimpan dalam repositori atau tempat penyimpanan dengan menggunakan teknik pengenalan pola serta statistik dan tehnik matematika.

Tahapan proses dalam *data mining* dan dapat dijelaskan sebagai berikut:

a. Data Selection

Pemilihan (seleksi) data baru sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD (*Knowledge Discovery in Database*) dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.

b. Preprocessing/Cleaning

Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi).

c. Transformasi

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data

tersebut sesuai untuk proses *data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan cari dalam basis data.

d. Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

e. Interpretasi / Evaluasi

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD

2.2. Algoritma C4.5

Menurut Larose(2005:p116). Algoritma C4.5 merupakan pengembangan dari algoritma ID3. Algoritma C4.5 mengkonstruksi pohon keputusan dari sebuah set *data training*, yang berupa kasus kasus atau *record-record (tuple)* dalam basis data, dimana tiap *record* memiliki atribut *kontinyu* diskrit atau keduanya

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut :

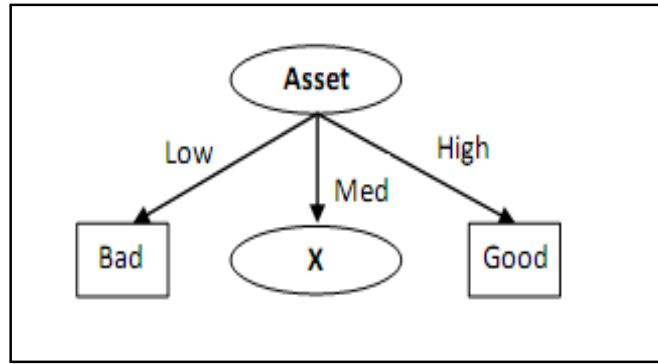
1. Pilih atribut sebagai root
2. Buat cabang untuk masing-masing nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Tabel 1. Tabel Data untuk Klasifikasi Credit Risk

Customer	Savings	Assets	Income
1	Medium	High	75
2	Low	Low	50
3	High	Medium	25
4	Medium	Medium	50
5	Low	Medium	100
6	High	High	25
7	Low	Low	25
8	Medium	Medium	75

Sumber: Larose, 2005

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



Gambar 1. Pohon Keputusan C4.5 dengan Simpul Akar Asset
 Sumber. Larose (2005:p.116)

Data *training* pada Tabel 1 adalah untuk menentukan apakah seorang nasabah bermasalah atau tidak, ditentukan oleh kolom *predictor saving, asset, dan income*. Kolom *credit risk* adalah kelas dari masing-masing *record*. Berikut akan dibahas prediksi apakah nasabah bermasalah atau tidak, menggunakan metode klasifikasi.

Langkah untuk membuat pohon keputusan, yaitu :

1. Tabel 2.1 adalah data *training* beserta kelasnya. Untuk atribut *income* yang bernilai angka, dibuat dalam bentuk kategori, yaitu $income \leq 25$, $income > 25$, $income \leq 50$, $income > 50$, $income \leq 75$, $income > 75$.
2. Hitung nilai *entropy*. Dari data training diketahui jumlah kasus ada 8, yang beresiko kredit *good* 5 *record* dan *Bad* 3 *record* sehingga didapat *entropy*:

$$Entropy(S) = - \frac{5}{8} \log_2 \frac{5}{8} - \frac{3}{8} \log_2 \frac{3}{8} = 0.9544$$

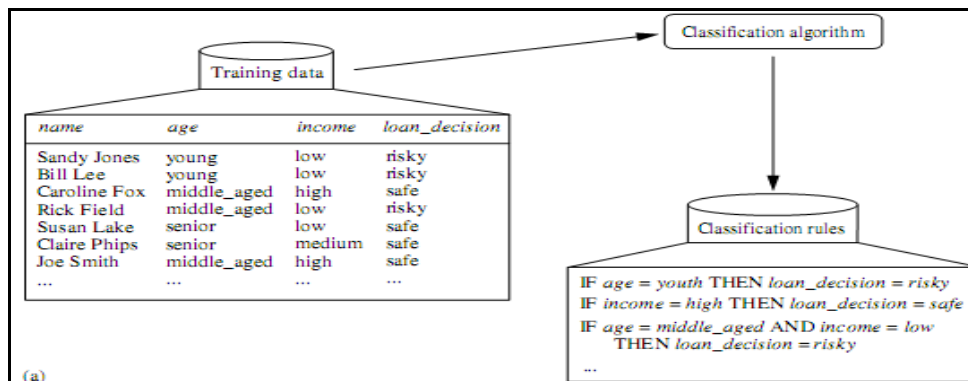
Hitung nilai *gain* untuk tiap atribut, lalu tentukan nilai *gain* tertinggi. Yang

mempunyai nilai *gain* tertinggi itulah yang akan dijadikan akar dari pohon.

2.3. Classification

Data classification memiliki dua tahap proses (Han and Kamber, 2006). Tahap pertama adalah membangun suatu model yang berdasarkan serangkaian *data class*, yang disebut *learned model*. Model tersebut dibangun dengan menganalisa *database tuple*. Setiap *tuple* diasumsikan menjadi *predefined class* yang ditentukan oleh satu atribut yang disebut class label atribut. Akibat terdapat class label maka tahap ini juga dikenal dengan *supervised learning*. Berbeda dengan *unsupervised learning* atau yang dikenal dengan *clustering*.

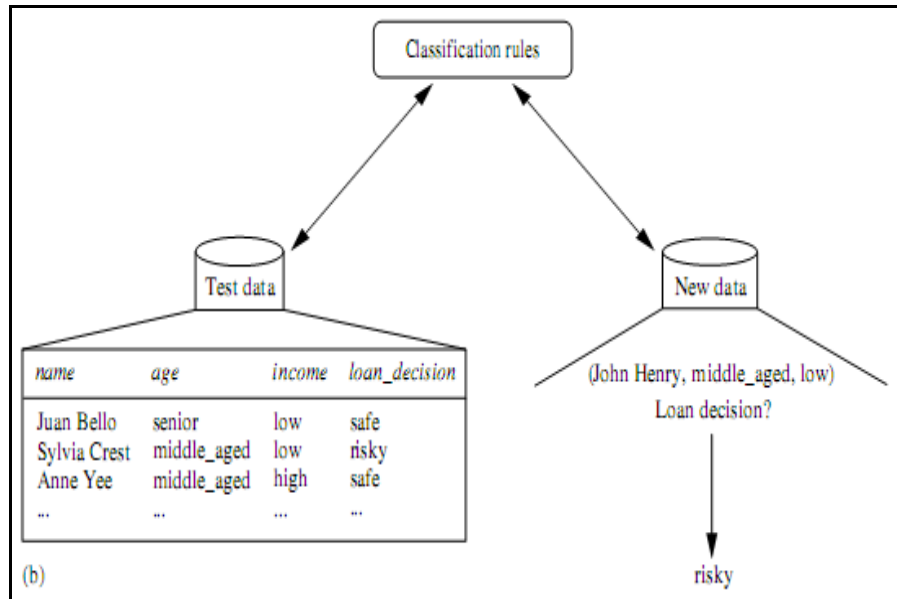
Umumnya, *learned model* digambarkan dalam bentuk *classification rules, decision tree*. Contohnya, diberi database mengenai *customer credit information*. Berdasarkan database tersebut dapat dibentuk *classification rules* yang nantinya dapat digunakan untuk mengidentifikasi *customer* apakah *customer* tersebut memiliki *excellent* atau *fair credit rating*.



Gambar 2. Tahap Pertama Classification: Learning
 Sumber: Han dan Kamber (2006,p287)

Pada tahap kedua, model digunakan untuk mengtes *data testing*. *Clasifikation rules* yang

sudah terbentuk pada tahap pertama digunakan memprediksi data kredit risk dari data baru



Gambar 3. Tahap kedua *Classification*
 Sumber: Han & Kamber (2006,p287)

2.4. Decision Tree

Pohon keputusan merupakan salah satu metode klasifikasi dan prediksi yang sangat kuat dan terkenal dalam penerapan *data mining*. Pada dasarnya *Decision Tree* mengubah data menjadi pohon keputusan (*decision tree*) dan aturan-aturan keputusan (*rule*).

Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target.

Sebuah pohon keputusan mungkin dibangun dengan saksama secara manual atau dapat tumbuh secara otomatis dengan menerapkan salah satu atau beberapa algoritma pohon keputusan untuk memodelkan himpunan data yang belum terklasifikasi. Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain ID3, CART, dan C4.5 (Larose,2005).

Variabel tujuan biasanya dikelompokkan dengan pasti dan model pohon keputusan lebih mengarah pada perhitungan probabilitas dari tiap-tiap *record* terhadap kategori-kategori tersebut atau untuk mengklasifikasi *record* dengan mengelompokkannya dalam satu kelas.

Decision tree membuat *set rule* yang paling efisien dan kemungkinan terkecil yang membuatnya menjadi *predictive model* yang

baik. Jika terdapat *overlap* diantara dua prediktor maka yang terbaik dari keduanya yang akan diambil. Pada sistem *rule induction*, keduanya akan diambil dan pada faktanya, salah satunya akan menjadi lebih lemah atau kurang akurat.

Rule induction beroperasi ‘*bottom up*’ dan mengumpulkan semua pola yang menarik, lalu menggunakannya sebagai *predictive target*. Di pihak lain, *decision tree*, bekerja dari *prediction target* sebagai *root*, lalu terus sampai *leaf* terbawah.

Kelebihan-kelebihan *decision tree* adalah sebagai berikut:

1. Menyediakan *visual result*
2. Dibangun berdasarkan rule-rule yang dapat dimengerti dan dipahami
3. Bersifat *predictive*
4. Memungkinkan untuk melakukan prediksi
5. Menampilkan apa yang penting

Algoritma akan mengidentifikasi atribut yang paling relevan dan akan mendefinisikan suatu *set rule* yang akan memberikan presentase kemungkinan akan terjadi hal demikian di kemudian hari

Decision tree dibuat dengan menggunakan sebuah teknik yang disebut *recursive*

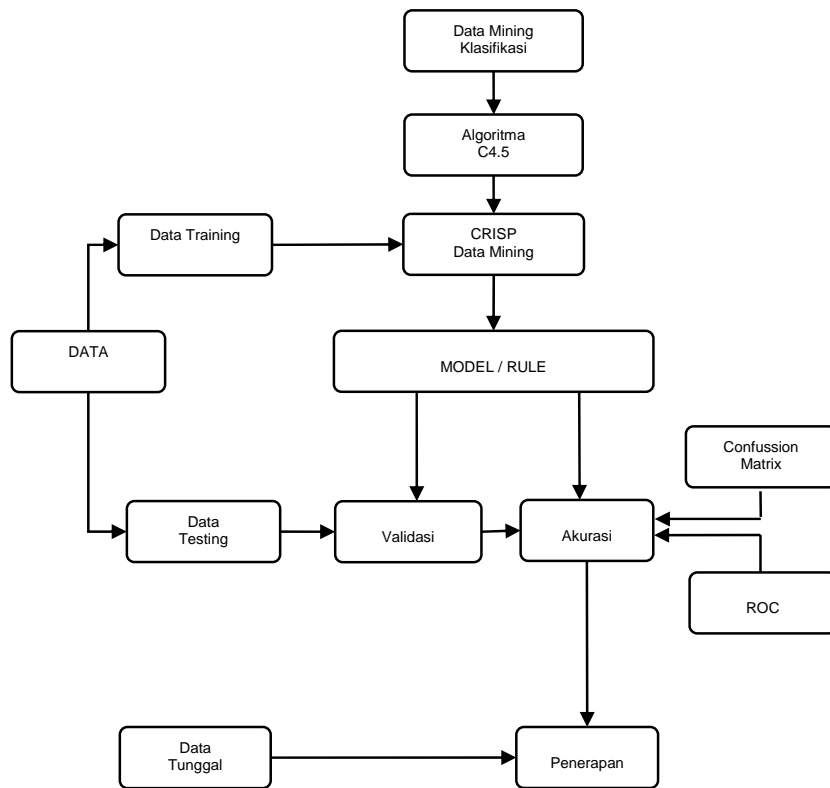
partitioning. Algoritma akan mendefinisikan atribut yang paling relevan dan akan *split* data yang ada berdasarkan atribut tersebut. Setiap *partition* disebut sebagai *rule*. Proses akan diulang terus untuk setiap *subgroup* sampai ditemukan sebuah *good stopping point*. *Information gain measure* digunakan untuk memilih atribut mana yang akan dites pada setiap node di *tree* (Han & Kamber, 2006). Atribut dengan *information gain* yang paling tinggi akan dipilih sebagai *test attribute* untuk *current node*.

2.5 Rapid Miner

Rapid Miner merupakan perangkat lunak yang dibuat oleh Dr. Markus Hofmann dari *Institute*

of Technology Blanchardstown dan Ralf Klinkenberg dari rapid-i.com dengan tampilan GUI (*Graphical User Interface*) sehingga memudahkan pengguna dalam menggunakan perangkat lunak ini. Perangkat lunak ini bersifat *open source* dan dibuat dengan menggunakan bahasa Java di bawah *lisensi GNU Public License* dan *Rapid Miner* dapat dijalankan di sistem operasi manapun. Dengan menggunakan *Rapid Miner*, tidak dibutuhkan kemampuan koding khusus, karena semua fasilitas sudah disediakan. *Rapid Miner* dikhususkan untuk penggunaan *data mining*.

2.6. Kerangka Pemikiran



Gambar 4. Kerangka Pemikiran

Kerangka pemikiran dari penelitian ini, dimulai dari *problem* (permasalahan) analisa kredit yang tidak akurat kemudian dibuat *approach* (model) dalam bentuk *decision tree* dengan algoritma C4.5 untuk memecahkan permasalahan. Untuk mengembangkan aplikasi (*development*) berdasarkan model yang dibuat, digunakan *Rapid Miner*. Tahap terakhir yaitu penerapan hasil model yang sudah terbentuk

menjadi *rule* dalam menganalisa nasabah kredit di BPR Marcorindo Perdana. Kerangka pemikiran dari penelitian ini, dapat digambarkan lebih detail lagi, yaitu: basis data nasabah kredit BPR, lalu data dibagi menjadi dua bagian, sebagian menjadi *data testing* dan sebagian data diambil sebagai data training, *data training* ini diolah dengan menggunakan CRISP-DM, yang mana CRISP-DM ini

digunakan sebagai metode penelitian suatu data mining klasifikasi dengan menggunakan *algoritma C4.5*, dengan CRISP-DM ini terbentuklah sebuah model/ *rule*. Model ini ini akan divalidasi terhadap data selanjutnya. Bersamaan dengan validasi, langkah lainnya yaitu, mengukur akurasi dengan menggunakan *Confusion Matrix* dan ROC. Hasil akurasi yang baik pada *rule* yang terbentuk maka *rule* tersebut dapat direkomendasikan untuk diterapkan dan implementasikan pada sebuah aplikasi yang dapat memprediksi perilaku nasabah kredit bermasalah

3. METODE PENELITIAN

3.1. Jenis Penelitian

Penelitian yang dilakukan adalah penelitian yang berbentuk Eksperimen. Dimana Penelitian Eksperimen ini menekankan pada percobaan dari teori-teori yang sudah ada.

3.2. Eksperimen

Ada beberapa tahap yang dilakukan dalam melakukan eksperimen ini, penulis menggunakan model *Cross-Standard Industry for Data Mining (CRISPDm)* (Larose, 2005) yang terdiri dari 6 tahap, yaitu :

1. *Business Understanding Phase*

Berdasarkan laporan kredit tahun 2009 ditemukan kredit bermasalah sebanyak 48 nasabah, sehingga ini menjadi permasalahan BPR. Marcorindo Persada dan merupakan imbas dari analisa analis yang kurang akurat. Agar akurat, maka dikembangkan model klasifikasi algoritma C4.5 untuk menentukan kelayakan pemberian kredit dengan tujuan analisa yang dilakukan lebih akurat.

2. *Data Understanding Phase*

Data kredit diambil dari BPR Marcorindo Persada pada tahun 2009, dimana dari 402 debitur, 48 diantaranya bermasalah dalam pembayaran atau dengan istilah lain macet dalam pembayaran kreditnya. Ada beberapa atribut yang digunakan di dalam data nasabah yaitu jumlah tanggungan , level penghasilan , level pinjaman , jangka waktu (lama pinjaman dalam tahun), usia, status, pendidikan, jenis kelamin, memiliki rumah, pengalaman mengambil kredit dan mark (label yang diberikan kepada nasabah, label Lancar untuk yang pembayarannya tidak bermasalah, dan bermasalah jika kreditnya macet). Berikut katagori nilai atributnya:

Tabel 1. Katagori Nilai Atribut

Atribut	Nilai Angka
Pendapatan	a. <3 juta b. 3-4 juta c. 5-9 juta d. >=10 juta
Level Pinjaman	a. Kecil b. Sedang c. Besar
Jumlah Tanggungan	a. Besar b. Sedang c. Sedikit d. Kosong
Jangka Waktu	a. 1 Tahun b. 2 Tahun c. 3 Tahun
Pendidikan	a. SD b. SMP c. SMA d. Sarjana
Umur	a. 19-25 Thun b. 26 s/d 55 thn c. >55
Status Perkawinan	a. Menikah b. Belum Menikah c. Duda/Janda(Cerai)

Masakerja	a. >5 tahun b. 2-5 tahun c. <2 tahun
Rumah Sendiri	a. Ada b. Tidak
Pengalaman Kredit	a. 1 b. 2 c. >=3

3. *Data Preparation Phase.*

Tahap persiapan data mencakup semua kegiatan untuk mencari data yang akan dimasukkan ke dalam alat pemodelan, dari data mentah awal. Pada fase ini tugas kemungkinan akan dilakukan beberapa kali, dan itu tidak diperlukan untuk melakukan apapun agar urutan. Tugas termasuk tabel dan catatan, sementara pemilihan atribut meliputi transformasi dan pembersihan data untuk alat pemodelan.

4. *ModellingPhase*

Tahap ini juga dapat disebut tahap *learning* karena pada tahap ini data training diklasifikasikan oleh model dan kemudian menghasilkan sejumlah aturan. Model yang digunakan dalam tahap ini menggunakan algoritma C4.5. Seperti yang telah dijelaskan sebelumnya, ada beberapa tahap yang harus dilalui dalam membentuk pohon keputusan, tentunya algoritma C4.5 digunakan untuk membuat pohon keputusan. Untuk membuat pohon keputusan, dibutuhkan data training yang diambil dari tabel pada lampiran 2.

Langkah-langkah untuk membuat algoritma C.45, yaitu :

- a. Siapkan data *training*
- b. Hitung nilai *entropy*.

Jika diketahui dalam data *training*, jumlah data 168, nasabah yang diterima kredit dalam status Lancar 142 *record* dan yang status Bermasalah 26 *record*. Berikut perhitungan *entropy*-nya :

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

- c. Setelah itu, hitung nilai *gain* untuk setiap atribut dan Split Infonya

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

$$SplitInfo(S,A) = \sum_{i=1}^k \frac{s_i}{s} \log_2 \frac{s_i}{s}$$

Hitung nilai *gain ratio* lalu pilih nilai *gain ratio* yang tertinggi. Nilai *gain ratio* tertinggi itulah yang akan dijadikan akar dari pohon.

$$Gain\ Ratio(S,A) = \frac{Gain(S,A)}{SplitInfo(S,A)}$$

Selanjutnya tentukan simpul berikutnya, yaitu simpul 1.1, dan dilakukan perhitungan *gain ratio* masing-masing atribut dari nilai rumah. Jumlah kasus yang dihitung adalah nilai dari simpul nasabah kredit dan seterusnya sampai semua record dalam simpul, mendapatkan kelas yang sama.

5. *Evaluation Phase*

Tahap ini dapat disebut tahap klasifikasi karena pada tahap ini menguji akurasi dari pohon keputusan yang sudah terbentuk. Evaluasi dilakukan terhadap rule yang dibuat dengan data *training* maupun pengujian dengan data *testing*.

Persamaan yang dapat digunakan untuk menghitung *accuracy*, *precision* dan *recall* (Han & Kamber,2007:p361) yaitu : Model dengan akurasi yang baik selanjutnya akan di test dengan data tunggal pada tahap penerapannya. Tahap ini dijelaskan secara lebih rinci pada bab IV.

6. *Deployment Phase*

Setelah pembentukan model dan dilakukan analisa dan pengukuran pada tahap sebelumnya, selanjutnya pada tahap ini diterapkan model yang paling akurat ke Bank Perkreditan Rakyat berupa *rule* dan aplikasinya untuk penentuan kelayakan pemberian kredit, yaitu menerapkan model algoritma C4.5 ke BPR. Marcorindo Perdana untuk penentuan kelayakan pemberian kredit ke nasabah.

4. HASIL DAN PEMBAHASAN

4.1. Hasil

Data *training* 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 berdasarkan nilai *gain ratio*, yaitu :

1. Siapkan data training
2. Hitung nilai *entropy*. Dari data training diketahui jumlah kasus ada 168, konsumen yang termasuk Lancar 140 *record* dan Bermasalah 28 *record* sehingga didapat *entropy*:

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

$$entropy(S) = (-142/168 \cdot \log_2(142/168)) + (-26/168 \cdot \log_2(26/168)) = 0.6216$$

3. Hitung nilai *gain* untuk tiap atribut (contoh atribut rumah)

$$Gain(S,A) = 0.6216 - ((146/168 \cdot 0.4333) + (22/168 \cdot 0.9760)) = 0.1173$$
4. Hitung nilai dari *Split info* untuk tiap atribut (contoh atribut rumah)

$$SplitInfo(S,A) = \sum_{i=1}^s \frac{s_i}{s} \log_2 \frac{s_i}{s}$$

$$SplitInfo(S,A) = - (146/168 \cdot 0.4333) - (22/168 \cdot 0.9760) = 0.5600$$

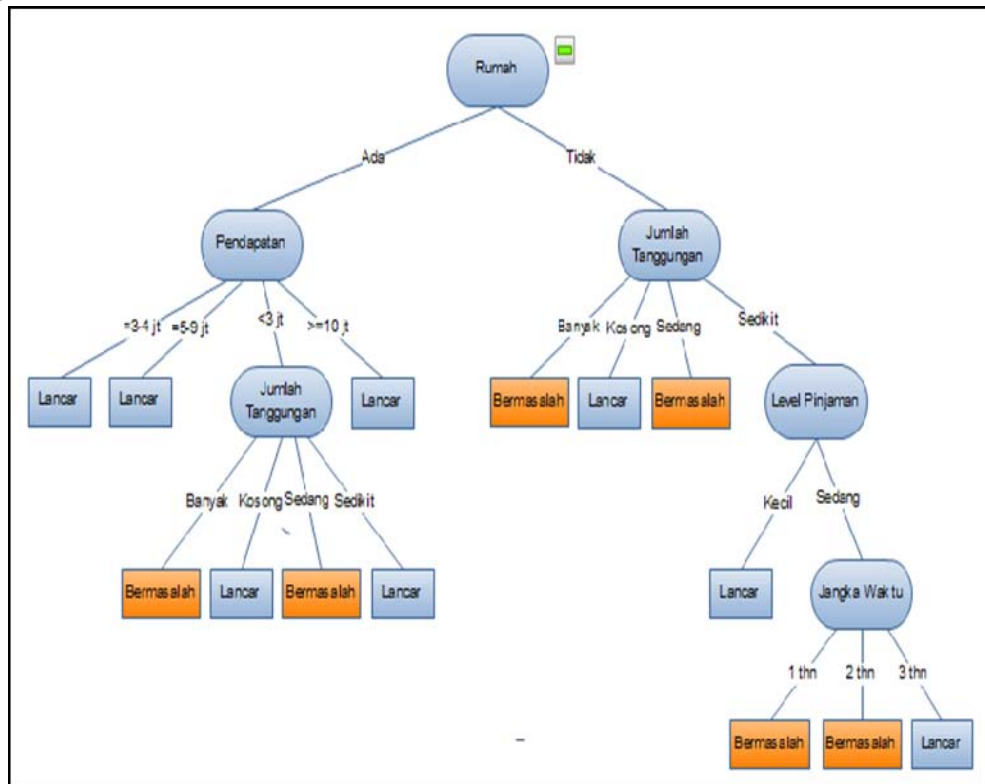
5. Hitung nilai dari *gain ratio* untuk tiap atribut (contoh atribut rumah)

$$Gain\ Ratio(S,A) = \frac{Gain(S,A)}{SplitInfo(S,A)}$$

$$Gain\ Ratio(S,A) = 0.1173 / 0.5600 = 0.2094$$

6. Tentukan nilai *gain ratio* tertinggi. Yang mempunyai nilai *gain ratio* tertinggi itulah yang akan dijadikan akar dari pohon.

Perhitungan *entropy*, *gain*, *splitinfo* dan *gain ratio* untuk semua atribut dilakukan, untuk mendapatkan nilai *gain ratio* tertinggi. Berdasarkan hasil perhitungan *gain ratio* pada atribut dan semua *instance* sudah berada pada partisi yang benar. Sehingga bentuk ke seluruh pohon tersebut bisa dilihat dibawah ini



Gambar 5. Tree

Tree

```

Rumah Sendiri = Ada
| Pendapatan = 3-4 jt: Lancar {Lancar=42, Bermasalah=3}
| Pendapatan = 5-9 jt: Lancar {Lancar=67, Bermasalah=1}
| Pendapatan = T<3 jt
| | Jml. Tanggungan = Banyak: Bermasalah {Lancar=0, Bermasalah=4}
| | Jml. Tanggungan = Kosong: Lancar {Lancar=2, Bermasalah=1}
| | Jml. Tanggungan = Sedang: Bermasalah {Lancar=0, Bermasalah=3}
| | Jml. Tanggungan = Sedikit: Lancar {Lancar=3, Bermasalah=1}
| Pendapatan = T>=10 jt: Lancar {Lancar=19, Bermasalah=0}
Rumah Sendiri = Tidak
| Jml. Tanggungan = Banyak: Bermasalah {Lancar=0, Bermasalah=4}
| Jml. Tanggungan = Kosong: Lancar {Lancar=3, Bermasalah=0}
| Jml. Tanggungan = Sedang: Bermasalah {Lancar=1, Bermasalah=5}
| Jml. Tanggungan = Sedikit
| | Level Pinjaman = Kecil: Lancar {Lancar=2, Bermasalah=0}
| | Level Pinjaman = Sedang
| | | Jangka Waktu = 1 thn: Bermasalah {Lancar=1, Bermasalah=2}
| | | Jangka Waktu = 2 thn: Bermasalah {Lancar=0, Bermasalah=2}
| | | Jangka Waktu = 3 thn: Lancar {Lancar=2, Bermasalah=0}

```

Berdasarkan pohon keputusan diatas, dapat dibentuk *rule classification*, yaitu sebagai berikut:

- a. R8: *IF* Memiliki Rumah = Ada *AND* Pendapatan = 3-4 juta *THEN* status kredit =Lancar
- b. R9: *IF* Memiliki Rumah = Ada *AND* Pendapatan = 5-9 juta *THEN* status kredit =Lancar
- c. R10: *IF* Memiliki Rumah =Ada *AND* Pendapatan <3 juta *AND* Tanggungan = Banyak *THEN* status kredit =Bermasalah
- d. R11: *IF* Memiliki Rumah = Ada *AND* Pendapatan <3 juta *AND* Tanggungan = Kosong *THEN* status kredit =Lancar
- e. R12: *IF* Memiliki Rumah = Ada *AND* Pendapatan <3 juta *AND* Tanggungan = Sedang *THEN* status kredit =Bermasalah
- f. R13: *IF* Memiliki Rumah = Ada *AND* Pendapatan <3 juta *AND* Tanggungan = Sedikit *THEN* status kredit =Lancar
- g. R14: *IF* Memiliki Rumah = Ada *AND* Pendapatan >=10 juta *THEN* status kredit =Lancar
- h. R1: *IF* Memiliki Rumah = Tidak *AND* Tanggungan = Banyak *THEN* status kredit=Bermasalah
- i. R2: *IF* Memiliki Rumah = Tidak *AND* Tanggungan = Kosong *THEN* status

kredit=Lancar

- j. R3: *IF* Memiliki Rumah = Tidak *AND* Tanggungan = Sedang *THEN* status kredit =Bermasalah
- k. R4: *IF* Memiliki Rumah = Tidak *AND* Tanggungan = Sedikit *AND* Pinjaman = Kecil *THEN* status kredit =Lancar
- l. R5: *IF* Memiliki Rumah = Tidak *AND* Tanggungan = Sedikit *AND* Pinjaman = Sedang *AND* Jangka Waktu=1 tahun *THEN* status kredit =Bermasalah
- m. R6: *IF* Memiliki Rumah = Tidak *AND* Tanggungan = Sedikit *AND* Pinjaman = Sedang *AND* Jangka Waktu=2 tahun *THEN* status kredit =Bermasalah
- n. IR7: *IF* Memiliki Rumah = Tidak *AND* Tanggungan = Sedikit *AND* Pinjaman = Sedang *AND* Jangka Waktu=3 tahun *THEN* status kredit =Lancar

4.2 Evaluasi dan Validasi

Seperti yang disebutkan dalam bab 2, untuk membuat model klasifikasi, bisa digunakan banyak metode. Dalam penelitian ini hanya menggunakan metode klasifikasi dengan algoritma C4.5. Metode klasifikasi bisa dievaluasi berdasarkan beberapa kriteria seperti tingkat akurasi, kecepatan, kehandalan.

Nilai *accuracy*, *precision*, dan *recall* dari data training dapat dihitung dengan menggunakan *Rapid Miner*.

Hasil dari pengujian model yang telah dilakukan, dilakukan pengujian tingkat akurasi dengan menggunakan *confussion matrix* dan kurva ROC/AUC (*Area Under Cover*).

Gambar 7 adalah perhitungan akurasi data training menggunakan algoritma C4.5. Diketahui dari 168 data *training* dan 10 atribut

(penghasilan, jumlah tanggungan, level pinjaman, jangka waktu, status, usia, jenis kelamin, tahun kerja, pendidikan, memiliki rumah, pengalaman kredit), dengan menggunakan metode algoritma C4.5 didapat 139 data prediksi Lancar sesuai dengan Lancar, 7 prediksi Lancar ternyata Bermasalah, 1 data prediksi Bermasalah ternyata Lancar, dan 21 data predisi Bermasalah sesuai dengan Bermasalah.

accuracy: 95.24%			
	true Lancar	true Bermasalah	class precision
pred. Lancar	139	7	95.21%
pred. Bermasalah	1	21	95.45%
class recall	99.29%	75.00%	

Gambar 7. *Confussion Matrix(accuracy) data training*

Setelah dilakukan pengujian data *training* maka didapatkan hasil pengukuran terhadap data *training* yaitu yaitu *accuracy* = 95.24%, *precision* = 95.45% dan *recall* = 75.00%.

Sedangkan untuk pengujian data *testing*, model yang telah dikembangkan akan diuji keakuratannya dengan memasukan sejumlah data uji (*test set*) ke dalam model. Untuk mengukur keakuratan model dengan baik, data uji seharusnya bukan data yang

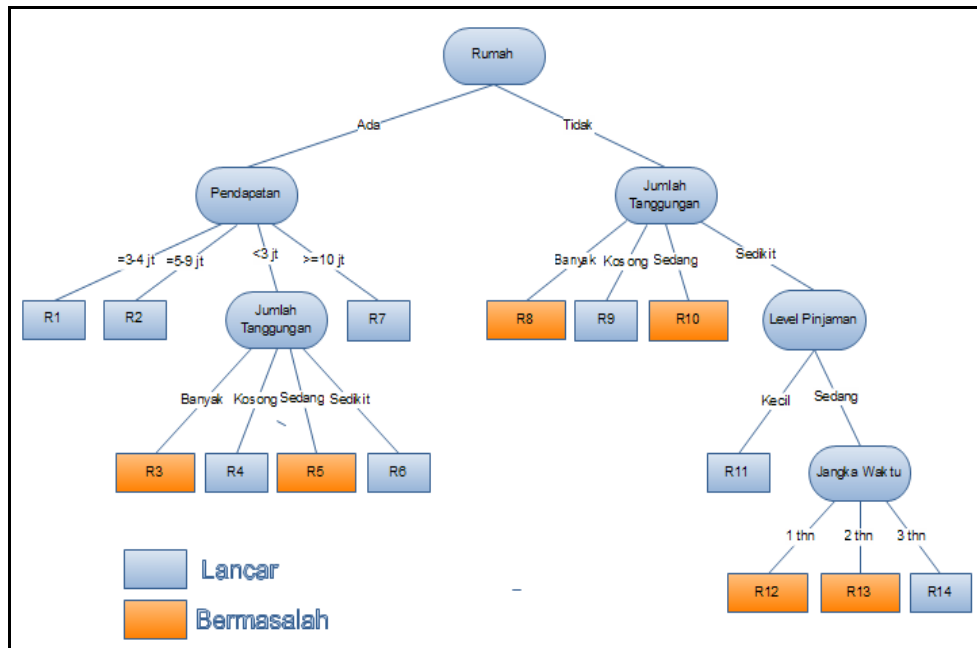
berasal dari data *training* (Han & Kamber, 2006). Data uji diambil dari data murni yang sebelumnya sudah dipecah menjadi dua yaitu 90% untuk data *training* dan 10% untuk data *testing*.

Pengujian *classification rule* yang sudah terbentuk dengan menggunakan data testing. Langkah-langkah pengujiannya sebagai berikut:

Tabel 2. *Data Testing*

Lampiran Data Testing											
Pendapatan	Jumlah Tanggungan	Level Pinjaman	Jangka Waktu	Status	Usia	JK	Lama Kerja	Pendidikan	Rumah	Pengalaman Kredit	Mark (Class)
<3 jt	Kosong	Sedang	3 thn	Cerai	26-55	L	2-5	SD	Ada	<=1	Lancar
<3 jt	Sedang	Sedang	2 thn	Menikah	26-55	L	2-5	SD	Tidak	<=1	Bermasalah
5-9 jt	Banyak	Sedang	3 thn	Menikah	26-55	L	2-5	SMP	Ada	<=1	Lancar
<3 jt	Sedang	Sedang	3 thn	Cerai	26-55	L	2-5	SD	Ada	<=1	Bermasalah
3-4 jt	Sedang	Kecil	2 thn	Menikah	26-55	L	2-5	SD	Ada	<=1	Lancar
<3 jt	Sedikit	Kecil	3 thn	Menikah	19-25	L	2-5	SMP	Ada	<=1	Lancar
<3 jt	Sedikit	Kecil	2 thn	Menikah	26-55	L	2-5	SMP	Ada	<=1	Bermasalah
3-4 jt	Banyak	Sedang	3 thn	Menikah	26-55	P	2-5	SMP	Ada	<=1	Lancar
3-4 jt	Kosong	Sedang	3 thn	Lajang	19-25	L	2-5	SMP	Ada	<=1	Lancar
3-4 jt	Sedang	Sedang	3 thn	Cerai	19-25	L	<=1	SMP	Tidak	2-3	Bermasalah
3-4 jt	Sedang	Sedang	3 thn	Lajang	26-55	L	2-5	SMP	Ada	<=1	Lancar
>=10 jt	Sedikit	Sedang	3 thn	Menikah	T>55	L	>5	SMA	Ada	2-3	Lancar
5-9 jt	Sedang	Sedang	3 thn	Menikah	19-25	L	2-5	SMP	Ada	2-3	Lancar
5-9 jt	Sedang	Sedang	3 thn	Menikah	19-25	L	2-5	SMP	Ada	<=1	Lancar
<3 jt	Banyak	Sedang	3 thn	Menikah	19-25	L	2-5	SMA	Ada	<=1	Bermasalah
<3 jt	Banyak	Sedang	3 thn	Menikah	19-25	L	2-5	SMA	Tidak	2-3	Bermasalah
3-4 jt	Sedang	Kecil	3 thn	Menikah	19-25	L	2-5	SMP	Ada	<=1	Lancar
>=10 jt	Sedikit	Sedang	3 thn	Menikah	26-55	L	>5	SMA	Ada	<=1	Lancar
3-4 jt	Banyak	Sedang	3 thn	Menikah	26-55	L	2-5	SMA	Ada	<=1	Lancar

Lakukan pengujian satu-persatu data testing ke dalam model atau *rule*



Gambar 8. Pohon Keputusan untuk Pengujian Data Testing

Tabel 3. Status Kredit pada Rule Pohon Keputusan

Status Kredit Pada Rule Pohon Keputusan			
Rule	Status Kredit	Rule	Status Kredit
R1	Lancar	R8	Bermasalah
R2	Lancar	R9	Lancar
R3	Bermasalah	R10	Bermasalah
R4	Lancar	R11	Lancar
R5	Bermasalah	R12	Bermasalah
R6	Lancar	R13	Bermasalah
R7	Lancar	R14	Lancar

Tabel 4. Hasil Pengujian Model dengan *data testing*

	True: Lancar	True: Bermasalah	TOTAL
Lancar	13	1	14
Bermasalah	0	5	5
TOTAL	13	6	19

4.2.2 Penerapan Model

Penerapan dari *model decision tree* yang sudah terbentuk di atas dapat diterapkan dengan menggunakan aplikasi yang dibuat

dengan bahasa pemrograman *Visual Basic* seperti contoh dibawah. Aplikasi ini digunakan untuk menguji coba data baru yang akan masuk.

Gambar 9. Penerapan model *decesion tree* dengan data baru

Coding program Visual Basic penerapan dari *rule classifcation* :

Private Sub cHasil_Click()

If Rumah = "Ada" And (Pendapatan >= 3000000 And Pendapatan <= 4000000) Then Hasil = "Lancar"

If Rumah = "Ada" And (Pendapatan >= 5000000 And Pendapatan <= 9000000) Then Hasil = "Lancar"

If Rumah = "Ada" And (Pendapatan < 3000000) And Tanggungan = "Banyak" Then Hasil = "Bermasalah"

If Rumah = "Ada" And (Pendapatan < 3000000) And Tanggungan = "Kosong" Then Hasil = "Lancar"

If Rumah = "Ada" And (Pendapatan < 3000000) And Tanggungan = "Sedang" Then Hasil = "Bermasalah"

If Rumah = "Ada" And (Pendapatan < 3000000) And Tanggungan = "Sedikit" Then Hasil = "Lancar"

If Rumah = "Ada" And (Pendapatan >= 10000000) Then Hasil = "Lancar"

If Rumah = "Tidak" And Tanggungan = "Banyak" Then Hasil = "Bermasalah"

If Rumah = "Tidak" And Tanggungan = "Kosong" Then Hasil = "Lancar"

If Rumah = "Tidak" And Tanggungan = "Sedang" Then Hasil = "Bermasalah"

If Rumah = "Tidak" And Tanggungan = "Sedikit" And Pinjaman = "Kecil" Then Hasil = "Lancar"

If Rumah = "Tidak" And Tanggungan = "Sedikit" And Pinjaman = "Sedang" And JWaktu = 1 Then Hasil = "Bermasalah"

If Rumah = "Tidak" And Tanggungan = "Sedikit" And Pinjaman = "Sedang" And JWaktu = 2 Then Hasil = "Bermasalah"

If Rumah = "Tidak" And Tanggungan = "Sedikit" And Pinjaman = "Sedang" And JWaktu = 3 Then Hasil = "Lancar"

End Sub

5. KESIMPULAN

Penelitian ini dilakukan untuk membantu menyelesaikan permasalahan dengan merancang sebuah aplikasi data mining yang berfungsi untuk memprediksi perilaku nasabah kredit yang berpotensi melakukan peminjaman (kredit) terhadap bank. Setelah dilakukan uji coba dalam penelitian dan perhitungan dengan menggunakan algoritma C4.5 dapat disimpulkan beberapa hal sebagai berikut

1. Algoritma C4.5 menggunakan perhitungan entropy, information gain dan gain ratio untuk membentuk sebuah pola berupa pohon keputusan.
2. Dari hasil penelitian yang dilakukan dapat disimpulkan bahwa algoritma C4.5 dapat digunakan sebagai alat analisa yang dilakukan oleh analis kredit. Hal ini diperkuat dengan hasil evaluasi penelitian bahwa algoritma C4.5 mampu menganalisa kredit yang bermasalah dan yang debitur yang tidak bermasalah sebanyak 90%.
3. Pola dari DecisionTree yang terbentuk dapat diterapkan ke dalam aplikasi untuk mempermudah mendeteksi perilaku nasabah yang bermasalah.

Daftar Pustaka

- 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.
- Hartono, Jogianto. (2005). *Pengenalan Komputer*. Yogyakarta : Andi
- Kasmir. (2011). *Analisis Laporan Keuangan*. Jakarta : PT. Rajagrafindo Persada
- Kusrini, & Luthfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Publishing.
- Larose, D. T. (2005). *Discovering Knowledge in Data*. New Jersey: John Willey & Sons, Inc.
- Liao. (2007). *Recent Advances in Data Mining of Enterprise Data: Algorithms and Application*. Singapore: World Scientific Publishing
- Maimon, Oded & Rokach, Lior. (2005). *Data Mining and Knowledge Discovey Handbook*. New York: Springer
- Michael and Gordon. (2004). *Sistem Informasi Manajemen*, Yogyakarta : Andi
- Myatt, Glenn J. (2007). *Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining*. New Jersey: John Wiley & Sons, Inc.
- Raymond & Goerge. (2008). *Sistem Informasi Manajemen*. Jakarta : Salemba Empat
- Sumathi, & S., Sivanandam, S.N. (2006). *Introduction to Data MiniApplications*. Berlin Heidelberg New York: Springer