

KLASIFIKASI KELAYAKAN KREDIT CALON DEBITUR BANK MENGUNAKAN ALGORITMA DECISION TREE C4.5

CLASSIFICATION OF DEBITOR'S BANK CREDITWORTHINESS USING DECISION TREE C4.5 ALGORITHM

Heri Hidayanto¹, Achmad Wahid Kurniawan²

Mahasiswa Program Studi Teknik Informatika-S1, Fakultas Ilmu Komputer, Universitas Dian
Nuswantoro Semarang

Jl. Imam Bonjol No. 207, Jl. Nakula No.5-11, Semarang, Kode Pos.50131, Telp.(024) 3517261

Email : Herihidayanto28@gmail.com , 111201106000@mhs.dinus.ac.id

ABSTRAK

Kredit adalah salah satu pelayanan yang diberikan oleh bank, resiko kredit yang terjadi dalam pemberian pinjaman kredit, dalam halnya nasabah tidak mampu membayar pinjaman yang diterima selalu diperhatikan oleh bank, dan melakukan pengawasan terhadap nasabah guna mengurangi resiko. Resiko utama untuk bank dan lembaga keuangan adalah membedakan kreditur yang memiliki potensi terjadi kredit macet, krisis ini menjadi perhatian lembaga keuangan tentang resiko kredit. Dalam pengambilan Keputusan untuk memberikan kredit ke pemohon digunakan prinsip *The Five C's of Credit Analysis*.

Pada penelitian ini digunakan teknik data mining dengan algoritma C4.5, dengan sumber data yang diperoleh berasal data historis debitur PD.BPR "BKK" Purwodadi-Grobogan 2014 dengan jumlah data sebanyak 627 record dengan 17 atribut. Dari 17 atribut kemudian direduksi menjadi 5 atribut yaitu umur, karakter, penghasilan, jumlah tanggungan, dan nilai anggunan dan menghasilkan data bersih 300 record . Pada pengujian cross validation dan split validation, menggunakan rapid minner dihasilkan akurasi tertinggi pada pengujian cross validation sebesar 93.33%.

Kata kunci : *The Five C's of Credit Analysis, data mining, algoritma c4.5, klasifikasi, kredit.*

ABSTRACT

Credit is one of the services offered by banks, credit risk that occurs in the granting of loans, in case the customer is unable to pay the loans received are monitored by the bank, and monitoring the customer to reduce the risk. The primary risk to banks and financial institutions are Credit scoring that have potential for bad credit, this crisis to the attention of financial institutions on credit risk. In taking the decision to give credit to the applicant used the principle of The Five C's of Credit Analysis. In this study used data mining techniques with C4.5 algorithm, the source of the data obtained from historical data PD.BPR debtor "BKK" Grobogan Purwodadi-2014, with the amount of data as much as 627 records with 17 attributes. From 17 attributes later reduced to five attributes such as age, character, income, number of dependents, and the value of collateral and generate clean data 300 record. In cross validation testing and validation split, using rapid minner generated the highest accuracy in the cross validation testing by 93.33%.

Keywords: *The Five C's of Credit Analysis, data mining, C4.5 algorithm, classification, credit.*

1. PENDAHULUAN

1.1 Latar Belakang

Kredit adalah salah satu pelayanan yang diberikan oleh bank, resiko kredit yang terjadi dalam pemberian pinjaman kredit, dalam halnya nasabah tidak mampu membayar pinjaman yang diterima selalu diperhatikan oleh bank, dan melakukan pengawasan terhadap nasabah guna mengurangi resiko (Pardede, 2004)

Resiko utama untuk bank dan lembaga keuangan adalah membedakan kreditur yang memiliki potensi terjadi kredit macet, krisis ini menjadi perhatian lembaga keuangan tentang resiko kredit. Data 3 tahun terakhir tahun 2012 sebanyak 505 debitur, tahun 2013 sebanyak 562 debitur dan tahun 2014 sebanyak 627 debitur pada BPR BKK Purwodadi.

Dalam pengambilan Keputusan untuk memberikan kredit ke pemohon digunakan prinsip *The Five C's of Credit Analysis*. *The Five C's of Credit Analysis* merupakan prinsip dalam analisis kredit terdiri atas : *Charakter* (kepribadian nasabah), *Capacity* (kemampuan nasabah, *Capital* (modal), *Collateral* (Nilai Barang

Jaminan) ,serta *Condition* (Kondisi nasabah) bank menilai dampak sektor ekonomi dan politik terhadap resiko terhadap usaha nasabah (Gunardi, 2009).

Didalam Prinsip *The Five C's of Credit Analysis* mencakup data yang dibutuhkan dalam proses analisis kredit. Sedangkan metode *decision tree* C4.5 merupakan sebuah metode dalam data mining yang menangani kasus yang bersifat hipotesis, untuk mengeksplorasi data, Seperti pada kasus klasifikasi resiko pemberian kredit nasabah bank atau lembaga keuangan lainnya. Metode yang dipakai seperti naïve bayes yang Menangani kuantitatif dan data diskrit. Diikuti oleh *Support vector machine* dan *Decision Tree* C4.5 dalam pengambilan keputusan mengeliminasi perhitungan yang tidak dibutuhkan, karena metode *decision tree* sampel yang diuji berdasarkan pada kriteria atau kelas tertentu (Sucipto, 2015). Metode C4.5 memiliki kelebihan yaitu diinterpretasikan dan dapat divisualisasikan dalam pohon keputusan, memiliki tingkat akurasi yang lebih tinggi dari metode yang

lain. Berdasarkan penelitian yang telah ada, menyatakan bahwa algoritma C4.5 memiliki tingkat akurasi yang lebih tinggi dari metode Naïve bayes dan *Support Machine vector*.

Berdasarkan perbandingan tingkat akurasi serta kelebihan algoritma C4.5. C4.5 dipilih sebagai metode dalam melakukan Klasifikasi Kelayakan Kredit Calon Debitur Bank Menggunakan Algoritma Decision Tree C4.5. karena telah terbukti pada penelitian sebelumnya memiliki akurasi yang cukup tinggi yaitu 87 % dan 92%.

1.2 Rumusan Masalah

Bagaimana menerapkan Algoritma Decision Tree C4.5 pada calon debitur PD. BPR BKK Purwodadi – Grobogan dan akurasi Algoritma Decision Tree C4.5 dalam mengklasifikasikan kelayakan kredit calon debitur PD. BPR BKK Purwodadi – Grobogan berdasarkan Prinsip *The Five C's of Credit Analysis* atau 5C .

2. Tinjauan Pustaka

2.2.1 Bank

Menurut UU RI No. 21 tahun 2008 tentang Perbankan Syariah yang

dimaksud dengan bank adalah “Bank adalah badan usaha yang menghimpun dana dari masyarakat dalam bentuk Simpanan dan menyalurkannya kepada masyarakat dalam bentuk kredit dan/atau bentuk lainnya dalam rangka meningkatkan taraf hidup rakyat” (Gunardi, 2009).

2.2.2 Kredit

Menurut Undang - Undang No10 tahun 1998, Kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu berdasarkan persetujuan atau kesepakatan panjam meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam untuk melunasi hutangnya setelah jangka waktu tertentu dengan jumlah bunga (Gunardi, 2009)

2.2.3 Prinsip-prinsip Perkreditan

Prinsip perkreditan disebut juga sebagai konsep 5C, antara lain:

1. *Character*

Menilai karakter untuk mengetahui apakah calon debitur merupakan orang yang jujur dan punya etikat baik untuk melakukan kewajiban yang harus dilakukan oleh debitur.

2. *Capacity*

Menilai kemampuan debitur untuk melakukan pelunasan kewajibannya dari usaha yang dilakukan atau usaha yang akan dibiayai oleh kredit bank.

3. *Capital*

Menilai besar atau kecilnya modal seorang calon debitur, serta menilai modal akan didistribusikan.

4. *Collateral*

Menilai jaminan fisik harta benda yang akan digunakan sebagai penjamin apabila seorang calon debitur tidak bisa memenuhi kewajiban, misal terjadi sesuatu seperti kecelakaan dan hal lain yang mengakibatkan peminjam tidak dapat melakukan pelunasan.

5. *Condition of Economy*

Melihat bagaimana situasi, kondisi ekonomi dan sektor usaha calon debitur atau peminjam. Agar mengurangi resiko yang muncul yang disebabkan kondisi ekonomi.

2.2.4 Klasifikasi

suatu pekerjaan dalam menilai objek data kemudian dimasukkan kedalam kelas

tertentu dari sekelompok kelas yang tersedia disebut klasifikasi. Pekerjaan utama yang dilakukan dalam mengklasifikasikan suatu objek antara lain, pembangunan model sebagai *prototype* yang akan disimpan dalam bentuk memori dan penggunaan model ini berfungsi untuk mengenalkan atau mengklasifikasikan suatu objek data lain untuk mengetahui dikelas mana objek data tersebut didalam model yang sudah disimpan. (Prasetyo, 2012)

empat komponen komponen dalam klasifikasi dilakukan berdasarkan [10]:

A.Kelas

Variabel dependen berupa kategorikal yang mempresentasikan “label” yang ada dalam obyek. Contohnya: Kelayakan pemberian kredit (Layak atau Tidak Layak)

B.Variabel Predictor

Variabel Independen direpresentasikan sebagai karakteristik data (atribut).

Contohnya:, umur, Karakter, Penghasilan, Jumlah Tanggungan ,Nilai Anggunan.

C.Dataset Training

Suatu data set berisi nilai dari kedua komponen diatas yang digunakan untuk menentukan kelas yang cocok berdasarkan predictor.

D. Testing dataset

Memuat data baru untuk diklasifikasikan oleh model yang telah dibuat kemudian dievaluasi akurasi klasifikasinya.

2.2.5 Algoritma C4.5

Di akhir tahun 1970 hingga di awal tahun1980-an, Quinlan[5] seorang peneliti di bidang mesin pembelajaran mengembangkan sebuah model pohon keputusan yang dinamakan ID3 (Iterative Dichotomiser), walaupun sebenarnya proyek ini telah dibuat sebelumnya oleh E.B. Hunt, J. Marin, dan P.T. Stone. Kemudian Quinlan membuat algoritma dari pengembangan ID3 yang

dinamakan C4.5 yang berbasis supervised learning. Salah satu metode klasifikasi menarik yang melibatkan konstruksi pohon keputusan, koleksi node keputusan, terhubung oleh cabang-cabang, memperpanjang bawah dari simpul akar sampai berakhir di node daun. Dimulai dari node root, yang oleh konvensi ditempatkan dibagian atas dari diagram pohon keputusan, atribut diuji pada node keputusan, dengan setiap hasil yang mungkin menghasilkan cabang. Berikut adalah rumus pada algoritma C4.5.

$$Info(D) = -\sum_{i=1}^m p_i \log_2(p_i)$$

$$Gain(A) = Info(D) - Info_A(D)$$

$$split_point = \frac{a_i + a_{i+1}}{2}$$

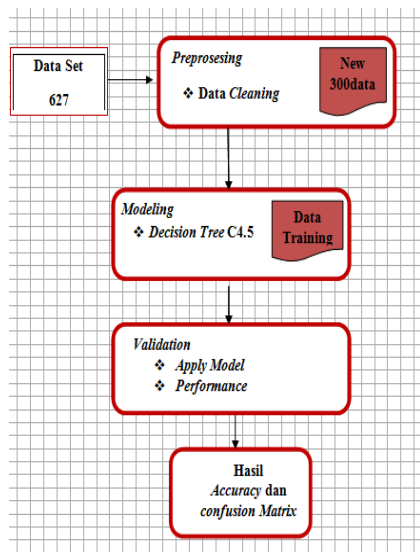
Dimana :

D = Dataset

m = banyaknya partisi D

pi = probabilitas

3. Metode Penelitian



Dalam pengujian data set pertama dilakukan adalah melakukan proses preprosesing yaitu Data cleaning, dari 627 data setelah dilakukan data cleaning menghasilkan 300 data set baru.

Pada pengujian validasi menggunakan cross validation dengan jumlah (k) dalam pengujian adalah 2, 5, 7, 10 dan serta pengujian menggunakan split validation dengan 4 kali percobaan dengan menggunakan data training 60%, 70%, 80%, dan 90% (Pramunendar, Dewi, & Asari, 2013). akan dilakukan proses pengujian cross validation dan split validation dalam rapidMiner, yang dengan melakukan permodelan dengan metode decision Tree C4.5, selanjutnya memasukan apply model dan performa. Validation merupakan

proses untuk mengevaluasi keakuratan prediksi dari model. Teknik validasi membagi data secara acak ke dalam k bagian disebut dengan cross validation. Dalam cross validation jumlah lipatan atau partisi dari data dilakukan sendiri, Pada proses cross validation menghasilkan confusion matrix dan akurasi dari model

3.1 Analisis Data

Data debitur yang didapatkan dari proses pengumpulan data sebanyak 627 record dan memiliki 17 atribut. Setelah data didapatkan untuk melakukan penelitian maka ada beberapa tahapan preparation data. Preparation data merupakan tahapan untuk mendapatkan data yang berkualitas, maka dapat dilakukan beberapa teknik sebagai berikut (Mahendra, 2015):

1) Data Cleaning

Data yang berkualitas sangat mempengaruhi kualitas keputusan yang akan diperoleh. Data yang tidak berkualitas adalah data yang nilai atributnya hilang, error dan data yang tidak konsisten dalam pengisian atributnya. Untuk mendapatkan data yang berkualitas dilakukan tahapan dalam data Cleaning sebagai berikut:

a) Menghilangkan data yang tidak lengkap terdapat 200 data dari 627 yang tergolong sangat tidak lengkap, 100 data dari 427 tergolong tidak lengkap, dan 27 data dari 327 tidak diberi umur sebagai kurang lengkap yang tersisa 300 *record*

b) mengisi nilai- nilai yang hilang, data yang tidak lengkap (*missing value*)

4 Hasil dan Pembahasan

4.1 Pembersihan Data Atau *Data cleaning*

Menghapus data yang tidak yang tergolong tidak lengkap dimana record dari atribut tersebut hanya ada 3 atribut yang terisi terdapat 200 data dari 627 data, kemudian yang tergolong sangat tidak lengkap dimana setiap record hanya satu atribut yang terisi, terdapat 100 dari 427, dan 27 data dari 327 tidak memiliki atribut umur sebagai data kurang lengkap sehingga menghasilkan data 300 data dari pemberihan data.

Atribut yang digunakan untuk penelitian:

NO	Atribut	Penjelasan
1	Umur	(20 – 29) , (30 -40), (> 40)
2	Karakter	Baik, Cukup , Tidak Baik .(karakter dapat dinilai , dengan melakukan cek pada BI <i>checking</i> atau dengan menanyakan kolega dan orang disekitar calon debitur, sehingga pegawai dapat menilai karakter calon debitur)
3	Penghasilan	Tinggi (Penghasilan > 4 kali dari angsuran), Cukup (Penghasilan 3 kali angsuran), Rendah (penghasilan < 2 kali angsuran
4	Jumlah Tanggungan	Tidak Ada (jumlah tidak mempunyai anak) anak < 2 (jumlah anak < 2), > 3 (jumlah anak > 3)
5	Nilai Anggunan	Tinggi (> 3 kali jumlah hutang), Cukup (2 kali Jumlah hutang) Rendah (1 kali jumlah hutang)
6	Keterangan	Layak (debitur akan dipertimbangkan dalam pemberian kredit) Tidak layak (pengajuan Kredit tidak diterima)

4.2 Validasi dan Evaluasi

Pada tahap ini akan dilakukan pengujian dengan menggunakan data yang telah melewati tahapan *preprocessing*, pada rapid miner untuk mengetahui akurasi dari model yang telah dibuat

4.3 Hasil dan Percobaan

Percobaan yang akan dilakukan adalah menggunakan data yang sudah dilakukan preprosesing dan data *cleaning*, dengan jumlah 300 data, dengan dengan 5 atribut yaitu umur, karakter, penghasilan, jumlah tanggungan, nilai anggunan dan satu atribut label yaitu layak dan tidak layak, pada percobaan *cross validation* jumlah pengujian dengan jumlah (k) adalah 2, 5, 7, dan 10. Dengan tipe sampling *Stratified*, Serta menggunakan *split validation* dengan data testing sebesar 60% training dan 40% data testing, kemudian pengujian ke – 2 menggunakan data training 70% dan data testing 30%, pengujian ke- 3 menggunakan data training 80% dan data testing 20% ,pengujian ke – 4 dengan data training 90% dan data testing 10

%. Pengujian validasi dilakukan dengan menggunakan tools Rapid Minner.

<i>Cross Validation</i>		<i>Split validation</i>	
<i>Sampling type : Stratified</i>		<i>Split: Absolute Sampling type: shuffled</i>	
<i>Number Validation</i>	<i>Accuracy %</i>	<i>Split data</i>	<i>Accuracy %</i>
K= 2	93.33	60% (<i>training</i>) 40% <i>Testing</i>	86.67
K= 5	89.33	70% (<i>training</i>) 30% <i>Testing</i>	87.78
K=7	88.63	80% (<i>training</i>) 20% <i>Testing</i>	90.00
K=10	88.00	90% (<i>training</i>) 10% <i>Testing</i>	86.67

4.5. Pembahasan

Percobaan yang dilakukan dengan menggunakan tools Rapid Minner, menggunakan metode klasifikasi decision tree c4.5 dan untuk validasi menggunakan *cross validation* dan *split validation* sedangkan untuk mengukur akurasinya menggunakan *Confusion Matrix*.

Dari pengujian yang dilakukan dengan menggunakan *cross validation* pengujian pada jumlah (k) 2 mendapatkan akurasi sebesar 93.33%, pada (k) 5 mendapatkan akurasi sebesar 89.33%, pada (k) 7 mendapatkan akurasi sebesar 88.63%, pada (k) 10 mendapatkan akurasi sebesar 88%, sedangkan pada pengujian menggunakan *split validation* dengan menggunakan data testing 60 % dan testing 40% mendapatkan akurasi sebesar 86.67% , dengan menggunakan data training 70% dan testing 30% mendapatkan akurasi sebesar 87.78 %, uji menggunakan data testing 80% dan testing 20% mendapatkan akurasi sebesar 90 %, dan uji dengan menggunakan data testing 90 % dan testing 10 % mendapatkan 86.67%. Berdasarkan pengujian validasi tersebut didapatkan uji validasi dengan akurasi tertinggi dengan menggunakan pengujian *cross Validation* yaitu sebesar 93.33%

4.6 Implementasi prototype

Untuk membantu pengklasifikasian kelayakan kredit calon debitur kredit bank, dirancanglah sebuah *prototype*, dengan menggunakan MatLab 2010, dimana pada inputan akan di inputkan, Umur, Karakter calon

debitur, Penghasilan, Jumlah Tanggungan dan Nilai Anggungan, dari data testing yang rule-nya telah diimplementasikan, kemudian akan mengklasifikasikan apakah seorang calon debitur layak atau tidak mendapatkan kredit.



5 Kesimpulan dan Saran

5.1 Kesimpulan

Dari penelitian yang dilakukan, dapat disimpulkan:

1. Klasifikasi kelayakan kredit berdasarkan prinsip *The five c 's Of Credit Analysis* atau 5C pada calon debitur PD. BPR. BKK Purwodadi Grobogan dapat diterapkan pada Algoritma Decision Tree C4.5,
2. Algoritma Decision Tree C4.5 dalam mengklasifikasikan Kelayakan kredit calon debitur kredit PD. BPR . BKK Purwodadi Grobogan pada pengujian *cross validation* dan *split validation* yang dilakukan sebanyak 4 kali,

dihasilkan akurasi tertinggi yaitu 93.33 %.

5. 2 Saran

Untuk meningkatkan kinerja dan menyempurnakan penelitian yang telah dibuat, peneliti memberikan saran sebagai berikut:

1. Penelitian selanjutnya dapat menambah parameter seperti prinsip 7P, sehingga lebih banyak informasi yang didapatkan.
2. Dapat dilakukan pengembangan dengan algoritma klasifikasi lain dalam data mining, agar dapat digunakan sebagai perbandingan.

Daftar Pustaka

1. Pardede, *Teori Model dan Kebijakan*. Yogyakarta, Indonesia: Andi, 2004.
2. Gunardi, "Analisis Kredit Bermasalah," STIE Malang Kucecwara, Malang, Magister Thesis 2009.
3. Adi Sucipto, "Prediksi Kredit macet Melalui Perilaku Nasabah Pada Koperasi Simpan Pinjam Dengan menggunakan Metode Algoritma Klasifikasi C4.5," *DISPOTEK*, vol. 6, no. 1, Januari 2015
4. Eko Prasetyo, *Data Mining : Konsep dan Aplikasi menggunakan MATLAB*, 1st ed. Yogyakarta, Indonesia: Andi, 2012.

5. Deny Cahya Mahendra, "Klasifikasi Data Debitur Untuk Menentukan Kelayakan Kredit Dengan Menggunakan Metode Naive Bayes," Universitas Dian Nuswantoro, Semarang, Skripsi 2015.

