

BAB III

METODE PENELITIAN

3.1 Dasar Penelitian

Penelitian ini dilakukan berdasarkan rumusan masalah yang telah dijabarkan pada bab sebelumnya yaitu untuk mengklasifikasikan kelayakan kredit calon debitur PD.BPR “BKK” Purwodadi - Grobogan 2014.

3.2 Instrumen Penelitian

Berdasarkan rumusan masalah yang telah dijabarkan, maka alat dan bahan yang dibutuhkan untuk penelitian ini antara lain:

3.2.1 Bahan

Dalam penelitian ini menggunakan data debitur di PD.BPR “BKK” Purwodadi Grobogan.

3.2.2 Peralatan

Pada penelitian yang akan dilakukan untuk menyelesaikan tugas akhir ini, instrument peralatannya yang berupa *software* dan *hardware* meliputi :

1. Perangkat lunak (*software*)

a) Sistem Operasi Windows 7

Menggunakan Windows 7 32 bit sebagai sistem operasi pada penelitian ini.

a) RapidMiner

RapidMiner merupakan perangkat lunak yang akan digunakan penulis untuk mengimplementasikan metode yang digunakan pada penelitian.

2) Perangkat Keras (*Hardware*)

Perangkat keras yang digunakan untuk menyelesaikan penelitian ini berupa Personal Computer dengan spesifikasi:

- CPU: Core™ i3 2,30 Ghz
- VGA: Nvidia GT 540 2GB
- RAM 4 GB

3.3 Sumber Data dan Jenis Data

Pada penelitian ini digunakan data yang berasal dari data historis debitur PD.BPR “BKK” Purwodadi-Grobogan 2014 dengan jumlah data sebanyak 627 *record*. Terdiri atas 17 atribut yaitu No, Cabang, Nama Debitur, No Kredit, Alamat, Jenis Kelamin, Umur, Agama, Status, Pekerjaan, karakter, Penghasilan, jumlah tanggungan, Nilai anggunan, Jangka Waktu Kredit, Angsuran, Keterangan (layak atau tidak layak).

3.3.1 Jenis Data

Pada penelitian ini akan dilakukan menggunakan dua jenis data yaitu data kuliatif dan data kuantitatif.

a) Data kualitatif

Data kualitatif merupakan data yang berbentuk kata – kata, data kualitatif dalam data yang akan digunakan pada penelitian ini yaitu cabang bank, nama debitur, alamat, jenis kelamin, agama, karakter status, pekerjaan, dan termasuk label target yaitu kriteria Layak dan Tidak layak.

b) Data kuantitatif

Data kuantitatif merupakan data yang berbentuk angka dan dapat dihitung nilainya dengan statistik ataupun matematika secara pasti. Fungsi dari data kuantitatif adalah mengetahui jumlah atau besaran objek yang akan diteliti. Data kuantitatif merupakan data yang menunjukkan probabilititas , dalam data yang digunakan untuk penelitian yaitu No Kredit, Penghasilan, umur, jumlah tanggungan, Jangka Waktu Kredit, Nilai Anggunan, Angsuran.

3.3.2 Sumber Data

Selama penelitian di PD.BPR “BKK” Purwodadi Grobogan penulis dapat mengumpulkan beberapa data, antara lain :

a) Data Primer

Data primer adalah data yang secara langsung diambil dari berbagai objek penelitian. Didapatkan data dari database dan Wawancara pegawai PD.BPR “BKK” Purwodadi-Grobogan.

a) Data Sekunder

Data yang diperoleh secara tidak langsung dari objek disebut dengan data sekunder. Dengan memperoleh data dari studi pustaka (*Library Research Method*) diantaranya buku ataupun penelitian yang membahas data mining pada perbankan, Jurnal yang membahas pengajuan kredit dan jurnal mengenai algoritma C4.5 .

3.4 Teknik Analisa data

Data debitur yang didapatkan dari proses pengumpulan data sebanyak 627 *reccord* 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 [10]:

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 konsiten 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 27data dari 327 tidak diberi umur sebagai kurang lengkap yang tersisa 300 *record*

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

2) *Data Integration and Transformation*

Teknik yang digunakan untuk menganalisis data kolerasi, atribut yang redundan disebut *integration*, sedangkan *transformation* berguna untuk meningkatkan akurasi dan efisiensi algoritma. Algoritma C4.5 memiliki kelebihan memproses data yang bernilai nominal, ortodinal maupun kontinyu. Sehingga nilai – nilai setiap atribut yang terdapat pada dataset tidak perlu ditransformasikan.

3) *Data Reduction*

Data set dapat direduksi dengan mengurangi jumlah atribut dan *record* supaya menjadi lebih sedikit tetapi tetap bersifat informatif. Memperoleh representasi dalam bentuk *volume* data yang telah berkurang jumlahnya namun tetap mendapatkan hasil analisis yang sama. Deskretasi data merupakan bagian dari reduksi data bagian penting untuk data numerik. Berikut adalah keseluruhan atribut sebelum dilakukan *reduction* data. Berdasarkan Prinsip *The Five C's of Credit Analysis* atau 5C, data yang berkaitan dengan prinsip tersebut parameternya adalah Umur, Penghasilan, Karakter, Jumlah tanggungan, Nilai Anggunan, dan keterangan sebagai Label. Kemudian menghapus atribut yang tidak digunakan, penghapusan atribut selain atribut yang digunakan. Dapat dilihat pada tabel 3. 1

Tabel 3.1 Atribut Data Penelitian

NO	Atribut	Proses	Penjelasan
1	No	Data Cleaning	No anggota debitur
2	Cabang	Data Cleaning	Nama cabang BPR
3	Nama Debitur	Data Cleaning	Nama dari debitur
4	No Kredit	Data Cleaning	Nomer pengajuan kredit
5	Alamat	Data Cleaning	Alamat dari debitur
6	Jenis Kelamin	Data Cleaning	Jenis kelamin debitur (L/P)
7	Umur	Digunakan Atribut	Usia dari debitur
8	Agama	Data Cleaning	Agam dari debitur
9	Status (perkawinan)	Data Cleaning	Status perkawinan debitur
10	Pekerjaan	Data Cleaning	Pekerjaan dari debitur
11	Karakter (Pribadi)	Digunakan Atribut	Kepribadian dari debitur
12	Penghasilan	Digunakan Atribut	Jumlah penghasilan dari debitur
13	Jumlah tanggungan	Digunakan Atribut	jumlah anak yang ditanggung dari debitur
14	Nilai Anggunan	Digunakan Atribut	Nilai kolekteral yang digunakan untuk jaminan
15	Jangka Waktu Kredit	Data Cleaning	Lama waktu pelunasan kredit
16	Angsuran	Data Cleaning	Jumlah nominal yang harus dibayarkan oleh debitur
17	Keterangan	Digunakan sebagai Label	Sebagai variabel dependen (Layak dan tidak layak)

Setelah dilakukan proses diatas yaitu melalui (*preparation data*), yaitu data *cleaning* , dimana yang keputusan pemberian kredit dilihat Umur, Penghasilan, Jumlah tanggungan dan Nilai anggunan dari penghasilan, dan tabel dibawah ini

merupakan data yang telah mengalami *reduction* data, dengan mengurangi jumlah atribut data namun tetap menghasilkan analitis yang sama atau serupa.

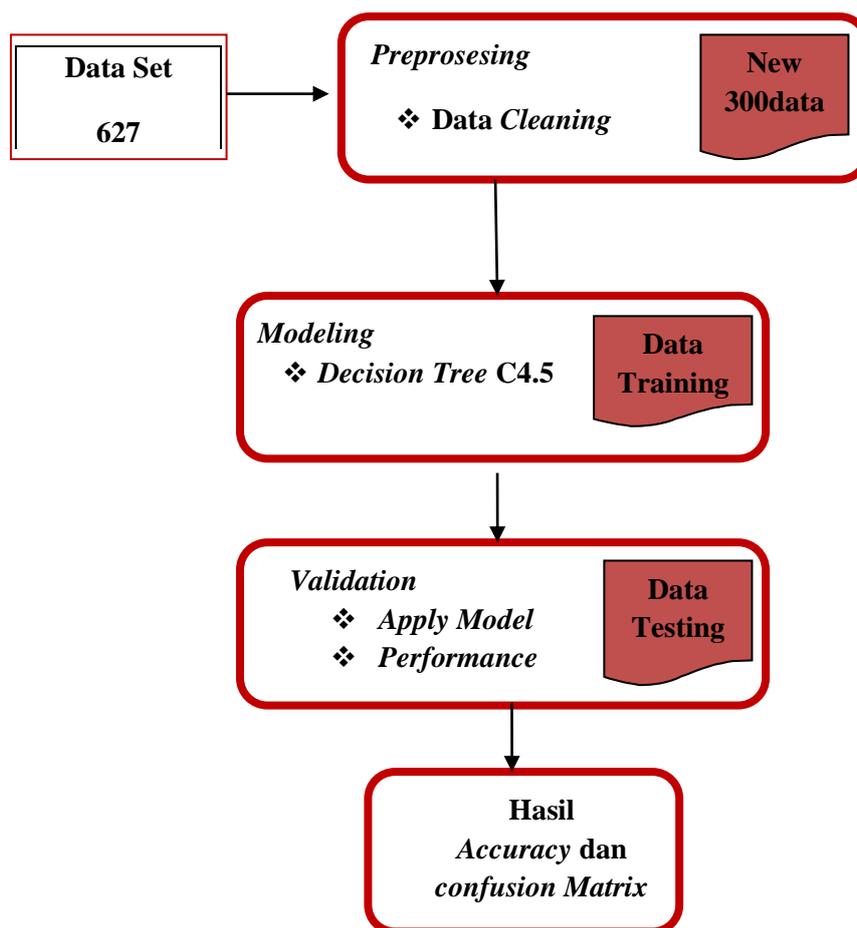
Tabel 3.2 Atribut yang akan digunakan

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)

3.5 Model Atau Metode yang Diusulkan

a) Permodelan (*Modeling*)

Metode yang akan digunakan pada penelitian adalah Algoritma C4.5 . Dalam permodelan ini Algoritma C4.5 akan dicari performanya yaitu *Performance Vector* (*accuracy*) dan *confusion matrix*. Data yang digunakan sudah melalui proses *cleaning* data, untuk melakukan pengukuran dalam penelitian ini menggunakan *tool* RapidMiner.

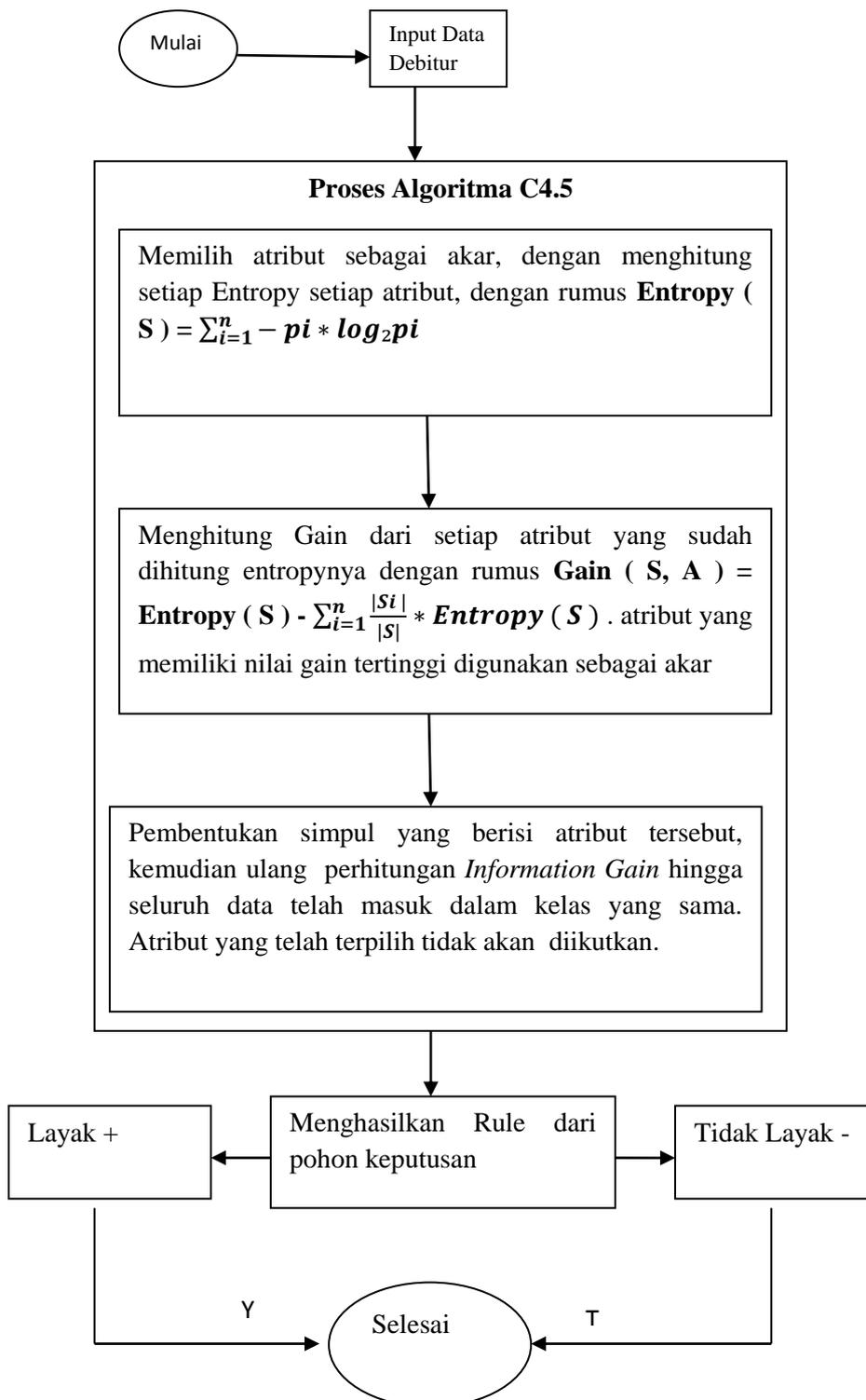


Gambar 3.1 Model tahap menemukan *Accuracy* dan *Confusion Matrix*

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% [11]. 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 .

b) Pemodelan Algoritma C4.5



Gambar 3.2 Permodelan Algoritma C4.5

Keterangan :

Entropy (s)

S = Himpunan Kasus

A = Atribut

N= Jumlah Partisi S

|Si| = jumlah Jumlah kasus pada partisi ke i

pi = Proporsi dari Si terhadap S

Identifikasi sampel dari *data set* baca data. Kemudian menghitung *entropy* (S) dari keseluruhan atribut, kemudian menghitung *gain* tertinggi dari seluruh atribut, selanjutnya didapatkan atribut yang akan digunakan sebagai akar / *node* . buat cabang untuk tiap- tiap nilai, bagi kasus dalam cabang, ulangi perhitungan *Gain* sampai semua data telah termasuk dalam kelas yang sama. Atribut yang telah dipilih tidak lagi diikuti dalam perhitungan. Dalam kasus *data set* pada penelitian ini terdiri dari 2 kelas yaitu debitur dengan kelas Layak yang dinyatakan (+) dan debitur dengan kelas Tidak Layak yang dinyatakan (-).

3.6 Cara Pengujian Model / Metode

Dalam perhitungan Algoritma C4.5 *data set* dibagi menjadi 70 *data training* dan 1 *data testing*.

3.6.1 Model Algoritma C4. 5

1. Menghitung jumlah kasus, yang di bagi menjadi 2 kelas yaitu Layak dan Tidak Layak , dan nilai entropy dari semua atribut Umur, Karakter, penghasilan, Jumlah tanggungan , Anggunan. Serta mencari nilai *gain* dari tiap atribut.
2. Sebelum melakukan perhitungan *gain* , hitung entropy dari tiap atribut dengan rumus sebagai berikut

$$Entropy = \sum_{i=1}^n - p_i * \log_2 p_i$$

Dengan perhitungan yang sama, dilakukan pada atribut berdasarkan pengelompokan jumlah kasus.

3. Kemudian menghitung nilai *gain*, dan tentukan nilai tertinggi untuk menentukan *node* atau akar yang pertama. Berikut rumus perhitungan *gain* :

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

4. Ulangi langkah Ke- 3 hingga semua atribut dan subset terhitung.
5. Proses pembentukan pohon keputusan terhenti jika sudah tidak ada atribut yang dipartisi dan semua *tuple* dalam *node* N telah memiliki kelas yang sama.

3.6.2 Tahapan Evaluasi Perhitungan Akurasi dan *Confusion Matrix*

Dalam tahapan ini mengevaluasi keakuratan hasil yang dicapai oleh model yang digunakan. Menggunakan *confusion matrix* yang telah disediakan dalam *framework* RapidMiner.

Tabel 3.3 *Confusion Matrix*

<i>Classification</i> C4. 5	Layak (+)	Tidak Layak (-)
Layak (+)	TN	FN
Tidak Layak (-)	FP	TP

Dari tabel di atas dapat hitung akurasinya sebagai berikut:

$$Akurasi = \frac{TP+TN}{TP+FP+FN+TN} \times 100\%$$