

# ANALISIS KLASIFIKASI PADA NASABAH KREDIT KOPERASI X MENGGUNAKAN *DECISION TREE C4.5* DAN *NAÏVE BAYES*

## *CLASSIFICATION ANALYSIS OF CREDIT CUSTOMERS X COOPERATIVES USING DECISION TREE C4.5 AND NAÏVE BAYES*

Oleh: Ahadiyah Nurul Kholifah<sup>1)</sup>, Nur Insani<sup>2)</sup>  
 Program Studi Matematika, Jurusan Pendidikan Matematika FMIPA UNY  
[ahadiyah.ifah@gmail.com](mailto:ahadiyah.ifah@gmail.com)<sup>1)</sup>, [nurinsani.utomo@gmail.com](mailto:nurinsani.utomo@gmail.com)<sup>2)</sup>

### Abstrak

Salah satu penyebab kredit bermasalah adalah kurang telitinya pihak koperasi dalam survei dan analisis pemberian kredit. Oleh karena itu, analisis kredit dengan teknik *data mining* perlu dilakukan sehingga dapat meminimalisir nasabah terlambat membayar angsuran serta mempersingkat waktu analisis pemberian kredit. Penelitian ini bertujuan untuk membentuk model *decision tree C4.5* dan *naïve bayes* untuk klasifikasi nasabah kredit berdasarkan nilai kolektibilitasnya. *Preprocessing* data yang terdiri dari *data cleaning*, *data integration*, *data selection*, dan *data transformation* dilakukan untuk meningkatkan kualitas model klasifikasi. Proses pembentukan model *decision tree C4.5* dan *naïve bayes* dilakukan menggunakan bantuan *software* WEKA 3.6.13. Hasil akurasi dari model *decision tree C4.5* dengan pengujian *use training set*, *10-fold cross validation*, dan *percentage split* berturut-turut adalah 71,91%, 68,03%, dan 66,84% sedangkan pada *naïve bayes* sebesar 67,01%, 64,66%, dan 65,82%.

Kata kunci: kredit, *data mining*, *decision tree C4.5*, *naïve bayes*, *software* WEKA

### Abstract

*One of loans problems which often occur are inaccurate surveys and loans analysis from the credit providers. Therefore, credit analysis with data mining techniques need to be done to minimize the customers that are late on paying the installments and to shorten the credit analysis. In this research, the data mining technique used to build a C.45 decision tree and naïve bayes models to classify credit customers based on its collectibility values. Data preprocessing consisting data cleaning, data integration, data selection, and data transformation were done to increase the quality of the classification model. Decision tree C4.5 and naïve bayes model is created using WEKA 3.6.13 software. The accuracy results from the decision tree C4.5 model using use training set, 10-fold cross validation, and percentage split were 71.91%, 68.03% and 66.84% respectively, where for the naïve bayes models the accuracy results for those three measurement were 67.01%, 64.66% and 65.82% respectively.*

Keywords: credit, *data mining*, *decision tree C4.5*, *naïve bayes*, WEKA

### PENDAHULUAN

Menurut Undang-Undang Perbankan No.12 Tahun 1992 pasal 1, kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam-meminjam antara pihak bank dengan pihak lain, yang mewajibkan pihak peminjam untuk melunasi hutangnya setelah jangka waktu tertentu dengan bunga, imbalan, atau pembagian hasil keuntungan.

Koperasi merupakan salah satu lembaga keuangan yang melakukan transaksi kredit dalam bentuk pinjaman anggota. Koperasi memiliki

prinsip kehati-hatian sebagaimana yang diterapkan pada perbankan dengan melakukan analisis kredit. Secara umum analisis kredit dilakukan dengan cara penilaian 5C. Pada dasarnya konsep 5C adalah *Character* (kepribadian), *Capacity* (kemampuan), *Capital* (modal), *Condition* (kondisi ekonomi), dan *Collateral* (agunan) Penilaian dengan 5C dapat memberikan beberapa informasi mengenai seberapa baik nasabah akan melunasi pinjaman (Kasmir, 2012: 136). Penilaian 5C biasanya dilakukan secara manual dan butuh waktu yang lama. Namun demikian, masih sering terjadi

permasalahan seperti adanya nasabah yang terlambat membayar angsuran.

Analisis terhadap data kredit dengan teknik lain perlu dilakukan untuk meminimalisir masalah tersebut. Teknik *data mining* merupakan proses menentukan pola dan informasi dari data yang berjumlah besar (Han, *et al*, 2012: 8). Proses di dalam *data mining* untuk membedakan kelas data atau konsep yang bertujuan agar dapat digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui dikenal sebagai klasifikasi. Dalam klasifikasi terdapat beberapa metode diantaranya *decision tree C4.5* dan *naïve bayes*.

*Decision tree* adalah metode klasifikasi paling terkenal karena mudah untuk diinterpretasi oleh manusia (Gorunescu, 2011: 3). *Decision tree* merupakan salah satu metode yang dapat diterapkan pada data kredit karena dapat mengubah data seperti pekerjaan, umur, penghasilan, dan kolektibilitas menjadi pohon keputusan dan aturan-aturan keputusan yang kemudian dapat digunakan untuk klasifikasi nasabah kredit berdasarkan status kelancaran. *Decision tree* juga dapat mem-*break down* proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan. Algoritma C4.5 dipilih untuk membangun pohon keputusan karena secara rekursif mengunjungi setiap simpul keputusan, memilih cabang optimal, sampai tidak ada cabang lagi yang mungkin dihasilkan (Indri Rahmayuni, 2014: 42).

*Naïve bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik sederhana yang berdasar pada teorema *Bayes* dengan asumsi independensi yang kuat (Eko Prasetyo, 2012: 59). *Naïve bayes* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam *database* dengan data yang besar (Kusrini & Emha, 2009: 189). *Naïve bayes* juga merupakan salah satu metode yang dapat diterapkan pada klasifikasi nasabah kredit yang didasarkan pada teorema *Bayes* dengan menghitung peluang dari satu kelas pada masing-

masing kelompok atribut yang ada dan menentukan kelas mana yang paling optimal. Hasil yang diperoleh dari perhitungan *naïve bayes* dapat mengklasifikasikan nasabah kredit berdasarkan status kelancaran.

Analisis klasifikasi nasabah kredit pernah dilakukan oleh beberapa peneliti. Penelitian terdahulu yang berkaitan dengan klasifikasi nasabah kredit antara lain: Claudia Clarentia Ciptohartono (2014) menilai kelayakan kredit perusahaan BCA Finance Jakarta 2013 menggunakan algoritma *naïve bayes* dengan *software* bantu *Rapid Miner*, model tes yang digunakan untuk mengukur kinerja algoritma adalah *cross validation*; Rina Fiati (2015) membuat model klasifikasi kelayakan kredit koperasi karyawan berbasis *decision tree* dengan bantuan *software Rapid Miner*; Yogi Yusuf (2009) evaluasi pemohon kredit mobil di PT X dengan menggunakan teknik *data mining decision tree* dan model *credit scoring* dengan algoritma C5.0.

Permasalahan dari uraian di atas melatarbelakangi perlunya penelitian mengenai penerapan teknik *data mining* khususnya metode *decision tree C4.5* dan *naïve bayes* untuk mengklasifikasikan nasabah kredit berdasarkan status kelancaran dan untuk mengetahui model yang paling baik dalam menangani data kredit Koperasi X.

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah sebagai berikut: (1) Bagaimana analisis klasifikasi pada data nasabah kredit Koperasi X menggunakan *decision tree C4.5* dan *naïve bayes*? dan (2) Bagaimana perbandingan hasil klasifikasi menggunakan *decision tree C4.5* dengan *naïve bayes*?

Agar pembahasan tidak terlalu luas, batasan-batasan dalam penelitian ini adalah sebagai berikut: (1) data yang digunakan adalah data peminjam Koperasi X pada tahun 2011 - Maret 2016 dan data laporan nominatif pinjaman perbulan tahun 2014 - Maret 2016, (2) klasifikasi nasabah kredit pada Koperasi X yaitu lancar, kurang lancar, diragukan, dan macet, (3) data

yang digunakan berasal dari objek penelitian sehingga model hanya dapat digunakan pada objek penelitian, (4) pembentukan model klasifikasi menggunakan bantuan aplikasi *software* yaitu WEKA 3.6.13, (5) metode klasifikasi yang digunakan yaitu *decision tree* C4.5 dan *naïve bayes*.

Tujuan penelitian ini yaitu mengetahui prosedur analisis klasifikasi pada data nasabah kredit Koperasi X menggunakan *decision tree* C4.5 dan *naïve bayes* serta mengetahui perbandingan hasil klasifikasi menggunakan *decision tree* C4.5 dan *naïve bayes*.

Adapun manfaat yang diharapkan dari penelitian ini. Secara teoritis bagi penulis, penelitian ini diharapkan dapat menambah pengetahuan dan wawasan mengenai *decision tree* C4.5 dan *naïve bayes* yang digunakan untuk mengklasifikasikan data nasabah kredit Koperasi X, sedangkan secara praktis, penelitian ini diharapkan dapat digunakan oleh koperasi X untuk mempermudah dalam pengambilan keputusan pemberian kredit.

**METODE PENELITIAN**

**Jenis dan Sumber Data Penelitian**

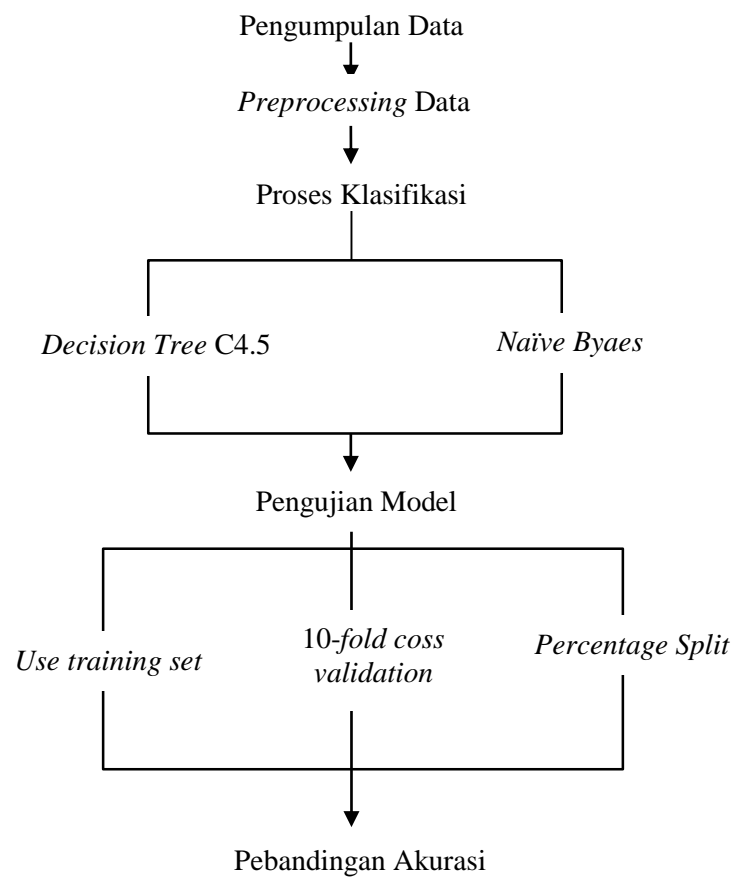
Sumber data diperoleh dari Koperasi X yang terdiri dari 3 file *excel* peminjam dengan jaminan sertifikat tanah, tunjuk, dan Buku Pemilik Kendaraan Bermotor (BPKB) serta 25 file *excel* laporan nominatif pinjaman perbulan dimana setiap file mewakili satu bulan.

Data yang digunakan adalah data peminjam yang telah melunasi pinjaman dan peminjam yang bermasalah. Pengumpulan data dilakukan secara manual yaitu pada data laporan nominatif pinjaman perbulan dipilih data yang mempunyai nilai bakidebet 0 artinya peminjam telah melunasi pinjaman dan data yang mempunyai nilai kolektibilitas 2-4 artinya peminjam bermasalah dalam mengangsur pinjaman. Hasil pemilihan pada data laporan nominatif pinjaman perbulan tersebut kemudian dihubungkan dengan file peminjam dengan *primary key* yaitu No.Anggota diperoleh sebanyak 1.076 *dataset*. Beberapa atribut yang

terpilih untuk proses data mining yaitu No.Anggota, Pekerjaan, No.KTP, Pekerjaan Istri/Suami, Pendapatan Keluarga, Pengeluaran Keluarga, Pinjaman, Kemampuan angsuran, Jangka Waktu, Agunan, Status Agunan, Nilai Jaminan, Tujuan, Tanggal Peminjaman, dan Kolektibilitas.

**Teknik Analisis Data**

Langkah-langkah analisis data dalam penelitian ini disajikan dalam Gambar 1.



Gambar 1. Langkah Analisis Data

Berikut adalah penjelasan tahapan-tahapan yang dilakukan setelah pengumpulan data:

**Preprocessing data**

a. Pembersihan Data (*Data Cleaning*)

Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten. Pada tahap ini data-data yang memiliki isian tidak sempurna seperti data yang tidak memiliki kelengkapan atribut yang dibutuhkan dan data yang tidak *valid* dihapus dari *database*.

b. Integrasi Data (*Data Integration*)

Integrasi data merupakan proses kombinasi beberapa sumber data ke dalam *database*. Pada tahap ini dilakukan penggabungan data dari berbagai sumber untuk dibentuk penyimpanan data yang koheren.

c. Seleksi Data (*Data Selection*)

Seleksi data merupakan pemilihan data yang digunakan untuk proses *data mining*. Data hasil seleksi yang akan digunakan untuk proses *data mining*, disimpan suatu berkas dan terpisah dari basis data operasional.

d. Transformasi Data (*Data Transformation*)

Transformasi data merupakan proses mentransformasikan dan mengkonsolidasikan data yang digunakan untuk proses *mining*. Pada tahap ini dilakukan perubahan format data menjadi format yang sesuai dengan teknik *data mining* yang digunakan.

**Proses Klasifikasi**

*Decision Tree* atau pohon keputusan merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon (*tree*) dimana setiap simpul internal (*internal node*) merupakan sebuah atribut, setiap cabang merupakan nilai atribut, dan setiap simpul daun (*leaf node*) atau simpul terminal merupakan label *class*, serta simpul yang paling atas adalah simpul akar (*root node*) (Han, et al, 2006: 291). Pohon keputusan memiliki beberapa cara dalam menentukan ukuran data dalam bentuk pohon, salah satunya adalah dengan algoritma C4.5. Algoritma C4.5 menggunakan *gain ratio* sebagai penentu simpul akar, internal, dan daun.

Perhitungan informasi yang dibutuhkan untuk mengklasifikasi pada tupel  $D$  dinyatakan sebagai berikut (Han, et al, 2006: 297):

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

dimana  $m$  merupakan banyaknya jenis kategori nilai pada atribut  $C$ ,  $p_i = \frac{|C_i, D|}{|D|}$  merupakan probabilitas dari tupel  $D$  yang mempunyai kelas  $C_i$ .

Misalkan terdapat atribut  $A$  yang memiliki  $v$  nilai yang berbeda  $\{a_1, a_2, \dots, a_v\}$ . Atribut  $A$  dapat digunakan untuk membagi  $D$  ke dalam  $v$  partisi  $\{D_1, D_2, \dots, D_v\}$ , dimana  $D_j$  memuat tupel  $D$  yang memiliki nilai  $a_j$  dari  $A$ . Persamaan untuk mencari nilai *entropy* dari subset  $A$  sebagai berikut (Han, et al, 2006: 298):

$$Info_A(D) = E(A) = \sum_{j=1}^v \frac{|D_j|}{|D|} \times Info(D_j)$$

dimana  $E(A)$  adalah *entropy* dari subset  $A$ ,  $v$  merupakan banyaknya jenis kategori nilai pada subset  $A$ ,  $\frac{|D_j|}{|D|}$  merupakan bobot dari subset  $j$  dan jumlah sampel pada subset yang mempunyai nilai  $a_j$  dari  $A$ , dibagi dengan jumlah tupel dari  $D$ .

Menurut Han, et al (2012: 298), nilai *information gain* dari atribut  $A$  pada subset  $D$  dapat dihitung dengan persamaan berikut:

$$Gain(A) = info(D) - E(A)$$

Nilai *split information* digunakan pada pencarian nilai *gain ratio* untuk mengatasi bias terhadap atribut yang memiliki banyak nilai unik. Persamaan *split information* dan *gain ratio* dinyatakan sebagai berikut:

$$Split Info_A(D) = - \sum_{j=1}^v \frac{|D_j|}{|D|} \times \log_2 \frac{|D_j|}{|D|}$$

$$Gain Ratio (A) = \frac{Gain(A)}{Split Info (A)}$$

Apabila atribut tersebut memiliki nilai *gain ratio* terbesar maka atribut tersebut terpilih sebagai atribut *split* pada konstruksi pohon keputusan (Han, et al, 2006: 301).

Selain klasifikasi *decision tree* C4.5 perlu juga mengetahui tahap klasifikasi *naïve bayes*. Menurut Han, et al (2012: 351) proses dari pengklasifikasian *naïve bayes* adalah sebagai berikut:

- a. Variabel  $D$  adalah kumpulan dari data dan label yang terkait dengan *class*. Setiap data diwakili oleh vektor atribut  $n$ -dimensi,  $X =$

$(x_1, x_2, \dots, x_n)$  dengan  $n$  dibuat dari data  $n$  atribut, berturut-turut,  $A_1, A_2, \dots, A_n$ .

- b. Misalkan terdapat  $i$  class,  $C_1, C_2, \dots, C_i$ . Diberikan sebuah data  $X$ , kemudian pengklasifikasian akan memprediksi  $X$  ke dalam kelompok yang memiliki probabilitas posterior tertinggi berdasarkan kondisi  $X$ . Artinya klasifikasi *naïve bayes* memprediksi bahwa  $X$  termasuk class  $C_i$  jika dan hanya jika:

$P(C_i|X) > P(C_j|X)$  untuk  $1 \leq j \leq m, j \neq i$   
 Maka nilai  $P(C_i|X)$  harus lebih dari  $P(C_j|X)$  supaya diperoleh hasil akhir  $P(C_i|X)$ .

- c. Ketika  $P(X)$  konstan untuk semua class maka hanya  $P(X|C_i)P(C_i)$  yang dihitung. Jika probabilitas class prior sebelumnya tidak diketahui, maka diasumsikan bahwa class-nya sama, yaitu  $P(C_1) = P(C_2) = \dots = P(C_m)$ , untuk menghitung  $P(X|C_i)$  dan  $P(X|C_i)P(C_i)$ . Perhatikan bahwa probabilitas class prior dapat diperkirakan oleh

$$P(C_i) = \frac{|C_{i,D}|}{|D|}$$

dimana  $|C_{i,D}|$  adalah jumlah data *training* dari kelas  $C_i$  dan  $D$  adalah jumlah total data *training* yang digunakan.

- d. Apabila diberikan kumpulan data yang mempunyai banyak atribut, maka mengurangi perhitungan  $P(X|C_i)$ , *naïve bayes* mengasumsikan pembuatan kelas independen yang bersyarat. Anggap bahwa nilai-nilai atribut tersebut bersifat independen satu sama lain dan diantara atribut tidak terdapat relasi dependensi, maka

$$P(X|C_i) = \prod_{k=1}^n P(x_k|C_i) \\ = P(x_1|C_i) \times P(x_2|C_i) \times \dots \\ \times P(x_n|C_i)$$

- e.  $P(X|C_i)P(C_i)$  dievaluasi pada setiap kelas  $C_i$  untuk memprediksi pengklasifikasian label kelas data  $X$  dengan menggunakan

$$P(X|C_i)P(C_i) > P(X|C_j)P(C_j) \\ \text{untuk } 1 \leq j \leq m, j \neq i$$

label kelas untuk data  $X$  yang diprediksi adalah kelas  $C_i$  jika nilai  $P(X|C_i)P(C_i)$  lebih dari nilai  $P(X|C_j)P(C_j)$

### Pengujian model

Model yang didapatkan dari kedua metode kemudian diuji menggunakan model tes *use training set*, *k-fold cross validation*, dan *percentage split*. Berikut penjelasan dari ketiga model tes:

- a. *Use training set*

Pengetesan dilakukan dengan menggunakan data *training* itu sendiri. Akurasi akan sangat tinggi, tetapi tidak memberikan estimasi akurasi yang sebenarnya terhadap data yang lain (data yang tidak dipakai untuk *training*).

- b. *Cross-validation*

*Cross-validation* adalah bentuk sederhana dari teknik statistik. Jumlah *fold* standar untuk memprediksi tingkat *error* dari data adalah dengan menggunakan *10-fold cross validation* (Witten, et al, 2011: 153).

- c. *Percentage split*

Hasil klasifikasi akan dites menggunakan k% dari data tersebut, dimana k adalah proporsi dari *dataset* yang digunakan untuk data *testing*. Pembagian data pada penelitian ini sebesar 80% untuk data *training* dan 20% untuk data *testing*.

### HASIL PENELITIAN DAN PEMBAHASAN

Penelitian ini menggunakan sebanyak 1.076 *dataset* yang terdiri dari 13 atribut. Sebelum pembentukan model *decision tree* C4.5 dan *naïve bayes* untuk klasifikasi nasabah kredit terlebih dahulu dilakukan tahap *preprocessing* data yang terdiri dari pembersihan data, integrasi data, seleksi data, dan transformasi data.

Pembersihan data dilakukan terhadap data yang tidak memiliki kelengkapan atribut dengan cara menghapus data tersebut sehingga jumlah data yang semula 1.076 *dataset* menjadi 979 *dataset*. Tahap selanjutnya dilakukan integrasi

data dengan penggantian atribut No.KTP dan tanggal peminjaman menjadi atribut umur. Selain itu integrasi data juga dilakukan pada awal ketika pemilihan data-data yang diperlukan yaitu menghubungkan data laporan nominatif pinjaman perbulan dengan data peminjam menggunakan *primary key* No.Anggota yang kemudian disimpan dalam satu file.

Setelah tahap integrasi data, dilanjutkan dengan tahap seleksi data dan transformasi data. Seleksi data yang dilakukan adalah penghapusan atribut No.Anggota dan penghapusan atribut-atribut yang tidak digunakan dalam analisis. Transformasi data dilakukan dengan mengubah tipe data numerik menjadi interval dan menginisialkan nilai atau isian karakter terlalu panjang pada beberapa atribut. Hasil transformasi data yang dilakukan dapat dilihat pada Tabel 1.

Tabel 1. Hasil Transformasi Data

| Atribut                              | Kategori               | Transformasi |
|--------------------------------------|------------------------|--------------|
| <b>Umur (th)</b>                     | <30                    | U1           |
|                                      | 30-39                  | U2           |
|                                      | 40-49                  | U3           |
|                                      | 50-59                  | U4           |
|                                      | ≥60                    | U5           |
| <b>Pendapatan Keluarga(rupiah)</b>   | 0-1000000              | D1           |
|                                      | 1000001-2000000        | D2           |
|                                      | 2000001-3000000        | D3           |
|                                      | 3000001-4000000        | D4           |
|                                      | 4000001-5000000        | D5           |
| <b>Pengeluaran Keluarga (rupiah)</b> | >5000000               | D6           |
|                                      | 0-1000000              | K1           |
|                                      | 1000001-2000000        | K2           |
|                                      | 2000001-3000000        | K3           |
|                                      | >3000000               | K4           |
| <b>Pinjaman(rupiah)</b>              | 0-2000000              | PJ1          |
|                                      | 2000001-4000000        | PJ2          |
|                                      | 4000001-6000000        | PJ3          |
|                                      | 6000001-8000000        | PJ4          |
|                                      | >8000000               | PJ5          |
| <b>Kemampuan Angsuran %</b>          | 0 – 40                 | KA1          |
|                                      | 41-60                  | KA2          |
|                                      | >60                    | KA3          |
| <b>Jangka Waktu (bulan)</b>          | 1-12                   | J1           |
|                                      | 13-24                  | J2           |
|                                      | >24                    | J3           |
| <b>Nilai Jaminan (rupiah)</b>        | 0-5000000              | N1           |
|                                      | 5000001-10000000       | N2           |
|                                      | 10000001-15000000      | N3           |
|                                      | 15000001-20000000      | N4           |
|                                      | >20000000              | N5           |
| <b>Pekerjaan</b>                     | Wiraswasta             | PK1          |
|                                      | Karyawan Swasta        | PK2          |
|                                      | Petani/pedagang/tukang | PK3          |
|                                      | Buruh                  | PK4          |

| Atribut                       | Kategori               | Transformasi |
|-------------------------------|------------------------|--------------|
| <b>Pekerjaan Suami/ Istri</b> | PNS                    | PK5          |
|                               | Tidak Bekerja          | PK6          |
|                               | Wiraswasta             | PS1          |
|                               | Karyawan Swasta        | PS2          |
|                               | Petani/pedagang/tukang | PS3          |
| <b>Agunan</b>                 | Buruh                  | PS4          |
|                               | PNS                    | PS5          |
|                               | Tidak Bekerja          | PS6          |
|                               | Sertifikat tanah       | A1           |
|                               | BPKB                   | A2           |
| <b>Status Agunan</b>          | Tunjuk                 | A3           |
|                               | Milik Sendiri          | SA1          |
|                               | Bukan Milik Sendiri    | SA2          |
| <b>Tujuan</b>                 | Modal                  | T1           |
|                               | Kesejahteraan          | T2           |
| <b>Kolektibilitas</b>         | 1 (Lancar)             | KL1          |
|                               | 2 (Kurang Lancar)      | KL2          |
|                               | 3 (Diragukan)          | KL3          |
|                               | 4 (Macet)              | KL4          |

Pembentukan model *decision tree* C4.5 dan *naïve bayes* menggunakan bantuan *software* WEKA 3.6.13 karena jumlah *dataset* yang besar sehingga akan sulit apabila diolah secara manual. *Output* dari *software* WEKA 3.6.13 pada pembentukan *model decision tree* C4.5 disajikan pada Gambar 2.

```

=== Classifier model (full training set) ===

J48 pruned tree
-----

Agunan = A1: KL1 (339.0/89.0)
Agunan = A2: KL1 (279.0/59.0)
Agunan = A3
|   Kemampuan Angsuran = KA1
|   |   Pendapatan Keluarga = D2
|   |   |   Nilai Jaminan = N5: KL3 (6.0/1.0)
|   |   |   Nilai Jaminan = N3
|   |   |   |   Pengeluaran Keluarga = K2:KL1 (2.0)
|   |   |   |   Pengeluaran Keluarga = K3:KL3 (0.0)
|   |   |   |   Pengeluaran Keluarga = K1:KL3 (3.0)
|   |   |   |   Pengeluaran Keluarga = K4:KL3 (0.0)
|   |   |   |   Nilai Jaminan = N1
|   |   |   |   |   Pengeluaran Keluarga = K2: KL3
|   |   |   |   |   (5.0/1.0)
|   |   |   |   |   |   Pengeluaran Keluarga = K3:KL3 (0.0)
|   |   |   |   |   |   Pengeluaran Keluarga = K1
|   |   |   |   |   |   Pekerjaan = PK1: KL2 (4.0/2.0)
|   |   |   |   |   |   Pekerjaan = PK2: KL1 (0.0)
|   |   |   |   |   |   Pekerjaan = PK3: KL1 (8.0/3.0)
|   |   |   |   |   |   Pekerjaan = PK4: KL2 (5.0/2.0)
|   |   |   |   |   |   Pekerjaan = PK6: KL1 (0.0)
|   .....

Number of Leaves :      41
Size of the tree :      53
Time taken to build model: 0.23 seconds
    
```

Gambar 2. Output WEKA Model Decision Tree C4.5

Berdasarkan Gambar 2, dapat diketahui bahwa model terbentuk dalam waktu 0,23 detik dengan ukuran pohon 53 dan memiliki 41 aturan

keputusan, beberapa diantaranya dapat dituliskan sebagai berikut:

1. **IF** Agunan = A1 **THEN** Kelas = KL1;
2. **IF** Agunan = A2 **THEN** Kelas = KL1;
3. **IF** Agunan = A3 **AND** Kemampuan Angsuran = KA1 **AND** Pendapatan Keluarga = D2 **AND** Nilai Jaminan = N5 **THEN** Kelas = KL3;
4. **IF** Agunan = A3 **AND** Kemampuan Angsuran = KA1 **AND** Pendapatan Keluarga = D2 **AND** Nilai Jaminan = N3 **AND** Pengeluaran Keluarga = K2 **THEN** Kelas = KL1;
5. **IF** Agunan = A3 **AND** Kemampuan Angsuran = KA1 **AND** Pendapatan Keluarga = D2 **AND** Nilai Jaminan = N3 **AND** Pengeluaran Keluarga = K3 **THEN** Kelas = KL3;

Output dari software WEKA 3.6.13 pada pembentukan model *naïve bayes* disajikan pada Gambar 3.

```

=== Classifier model (full training set) =
Naive Bayes Classifier

Attribute          Class
                   KL3      KL1      KL4      KL2
                   (0.16) (0.68) (0.06) (0.09)
=====
Pekerjaan
PK1                52.0    228.0   16.0    34.0
PK2                13.0    41.0    3.0     8.0
PK3                60.0    195.0   32.0    28.0
PK4                33.0    140.0   14.0    19.0
PK6                1.0     37.0    1.0     2.0
PK5                6.0     33.0    1.0     6.0
[total]           165.0   674.0   67.0    97.0

Umur
U1                23.0    66.0    10.0    12.0
U3                44.0    220.0   19.0    34.0
U2                51.0    202.0   16.0    22.0
U4                33.0    121.0   16.0    20.0
U5                13.0    64.0    5.0     8.0
[total]           164.0   673.0   66.0    96.0

.....
Time taken to build model: 0 seconds
    
```

Gambar 3. Output WEKA Model *Naïve Bayes*

Berdasarkan Gambar 3 dapat diketahui total data yang mempunyai nilai PK1 dan berada pada kelas KL1 sebanyak 52. Pada data

sebenarnya total data yang mempunyai nilai PK1 dan berada pada kelas KL1 sebanyak 51. Hal ini terjadi karena klasifikasi *naïve bayes* pada program WEKA menghindari frekuensi nol di setiap kelas sehingga terjadi penambahan 1 angka di setiap kelas. Berdasarkan Gambar 3 juga dapat diketahui bahwa model *naïve bayes* terbentuk dalam waktu mendekati 0 detik.

Evaluasi model perlu dilakukan untuk mengetahui tingkat akurasi dari kedua model. Evaluasi model dilakukan dengan pengujian menggunakan beberapa model tes yaitu *use training set*, *10-fold cross validation*, dan *percentage split*. Hasil pengujian yang diperoleh dapat dilihat pada Tabel 2 dan Tabel 3.

Tabel 2. Hasil Pengujian Model *Decision Tree* C4.5

|                                       | <i>Use Training Set</i> | <i>10-Fold Cross Validation</i> | <i>Percentage Split</i> |
|---------------------------------------|-------------------------|---------------------------------|-------------------------|
| <i>Correctly Classified</i>           | 71,91%                  | 68,03%                          | 66,84%                  |
| <i>Incorrect Classified</i>           | 28,09%                  | 31,97%                          | 33,16%                  |
| <b>Waktu komputasi (satuan:detik)</b> | 0,25                    | 0,14                            | 0,02                    |

Tabel 3. Hasil Pengujian Model *Naïve Bayes*

|                                       | <i>Use Training Set</i> | <i>10-Fold Cross Validation</i> | <i>Percentage Split</i> |
|---------------------------------------|-------------------------|---------------------------------|-------------------------|
| <i>Correctly Classified</i>           | 67,01%                  | 64,66%                          | 65,82%                  |
| <i>Incorrect Classified</i>           | 32,99%                  | 35,34%                          | 34,18%                  |
| <b>Waktu komputasi (satuan:detik)</b> | 0,01                    | 0                               | 0                       |

Berdasarkan Tabel 2 dan 3, akurasi yang diberikan dengan model tes *use training set*, *10-fold cross validation*, dan *percentage split* pada *decision tree* C4.5 berurutan adalah 71,91%, 68,03%, dan 66,84%, sedangkan pada *naïve bayes* berurutan adalah 67,01%, 64,66%, dan 65,82%. Dapat diambil informasi bahwa secara keseluruhan klasifikasi *decision tree* C4.5

memiliki kecenderungan tingkat akurasi yang lebih tinggi. Waktu komputasi yang diberikan dengan model tes *use training set*, *cross validation*, dan *percentage split* pada *decision tree* C4.5 berturut-turut adalah 0,25 detik, 0,14 detik, dan 0,02 detik, sedangkan pada *naïve bayes* berturut-turut adalah 0,01 detik, mendekati 0 detik, dan mendekati 0 detik. Dapat diambil informasi bahwa klasifikasi *naïve bayes* mempunyai waktu komputasi yang lebih cepat dibanding *decision tree* C4.5.

### SIMPULAN DAN SARAN

Analisis klasifikasi pada data kredit Koperasi X telah berhasil dilakukan dengan tahapan menyiapkan data, pengumpulan data, *preprocessing* yang terdiri dari *data cleaning*, *data integration*, *data selection*, dan *data transformation*, proses klasifikasi menggunakan metode *decision tree* C4.5 dan *naïve bayes*, dan evaluasi model. Hasil perbandingan kedua metode tersebut adalah klasifikasi *decision tree* C4.5 memiliki kecenderungan tingkat akurasi yang lebih tinggi yaitu sebesar 71,91%, 68,03%, dan 66,84%, sedangkan waktu komputasi *naïve bayes* lebih cepat dibanding *decision tree* C4.5 yaitu sebesar 0,01 detik, mendekati 0 detik, dan mendekati 0 detik.

Pengembangan dan perbaikan model perlu dilakukan guna memperoleh hasil yang lebih akurat dalam mengklasifikasikan nasabah kredit. Beberapa diantaranya, yaitu (1) menggunakan metode yang lain selain *decision tree* C4.5 dan *naïve bayes*, (2) mencari faktor lain yang dapat digali dari data yang tersedia, (3) memperbanyak jumlah data, (4) menggunakan *software* lain seperti Tanagra dan Matlab.

### DAFTAR PUSTAKA

Claudia Clarentia Ciptohartono. (2014). Algoritma Klasifikasi *Naïve Bayes* untuk Menilai Kelayakan Kredit. *Skripsi*. Fakultas Ilmu Komputer, Teknik

Informatika, Universitas Dian Nuswantoro.

- Eko Prasetyo. (2012). *Data Mining: Konsep dan Aplikasi menggunakan MATLAB*. 1st ed. Yogyakarta: Andi.
- Gorunescu, Florin. (2011). *Data Mining Concept, Models and Techniques*. Berlin Heidelberg: Springer.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concept and Techniques, Third Edition*. Waltham: Morgan Kaufmann.
- Han, J., Kamber, M., & Pei, J. (2006). *Data Mining: Concept and Techniques, Second Edition*. Waltham: Morgan Kaufmann.
- Indri Rahmayuni. (2014). Perbandingan Performasi Algoritma C4.5 dan CART dalam Klasifikasi Data Nilai Mahasiswa Prodi Teknik Komputer Politeknik Negeri Padang. *Jurnal TEKNOIF*, Vol.2, No. 1.
- Kasmir. (2012). *Dasar-dasar Perbankan*. Jakarta : Rajawali Press.
- Kusrini & Emha Taufiq Luthfi. (2009). *Algoritma Data Mining*. Yogyakarta: Andi.
- Rina Fiati & Putri Kurnia Handayani. (2015). Model Klasifikasi Kelayakan Kredit Koperasi Karyawan Berbasis *Decision Tree*. *Prosiding SNATIF Ke-2*. Universitas Muria Kudus.
- Undang-Undang Perbankan No.12 Tahun 1992 pasal 1.
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques Third Edition*. Burlington: Morgan Kaufmann.
- Yogi Yusuf W, F.Rian Pratikto, & Vivianne A. S. (2009). Evaluasi Pemohon Kredit Mobil di PT X dengan menggunakan Teknik Data Mining Decision Tree. *Simposium Nasional RAPI VIII*. Bandung: Universitas Katolik Parahyangan.