

# IMPLEMENTASI DATA MINING KLASIFIKASI NASABAH POTENSIAL MENGGUNAKAN ALGORITMA C4.5 (STUDI KASUS ADIRA DINAMIKA MULTI FINANCE)

Hidayatus Sholihah<sup>1</sup>, Fiqih Satria<sup>2</sup>

<sup>1</sup> Prodi Manajemen Informatika Pringsewu Lampung

<sup>1,2</sup> Jl. Wisma Rini No. 09 pringsewu Lampung

E-mail : dayhidas25@gmail.com<sup>1</sup>, [Fiqih.satria@gmail.com](mailto:Fiqih.satria@gmail.com)<sup>1</sup>

## ABSTRAK

Dengan pertumbuhan ekonomi yang secara pesat, kebutuhan akan informasi yang akurat sangat dibutuhkan dalam kehidupan sehari-hari. Salah satu keberhasilan perbankan adalah mengklasifikasi nasabahnya, karena nasabah merupakan aset terpenting dalam suatu perusahaan. Adira dinamika multi finance adalah perusahaan yang bergerak di bidang pembiayaan dana secara kredit, memberikan penawaran kembali terhadap nasabah potensial yang sebelumnya telah melakukan pembiayaan kredit merupakan suatu langkah untuk tetap mendapatkan nasabah kembali. Mengetahui nasabah yang berpotensi merupakan hal yang harus diketahui seorang analis kredit. Analisis kredit dengan teknik data mining perlu dilakukan untuk mempersingkat waktu analisis pemberian kredit. Penelitian ini bertujuan untuk membentuk model decision tree C4.5 untuk klasifikasi nasabah kredit berdasarkan data yang sudah ada di Adira Finance. Preprocessing data terdiri dari data cleaning, data integration, data selection, dan data transformation dilakukan untuk meningkatkan kualitas model klasifikasi. Proses pembentukan model decision tree C4.5 dilakukan menggunakan bantuan software WEKA 3.8.

**Kata kunci :** data mining, kredit, Adira Finance, decision tree C4.5, WEKA

## I. PENDAHULUAN

### 1.1 Latar Belakang

Dengan pertumbuhan ekonomi yang secara pesat, kebutuhan akan informasi yang akurat sangat dibutuhkan dalam kehidupan sehari-hari. Teknologi informasi menjadi kebutuhan sekunder bagi para penggunanya, sehingga informasi akan menjadi suatu elemen penting dalam perkembangan masyarakat untuk saat ini dan waktu yang akan datang.

Namun kebutuhan akan informasi yang tinggi kadang tidak diimbangi dengan penyajian informasi yang memadai, seringkali informasi harus digali ulang dari data yang jumlahnya sangat besar. Penggunaan teknik data mining diharapkan dapat memberikan pengetahuan-pengetahuan yang tersembunyi, sehingga dapat menjadi informasi yang berharga.

Adira dinamika multi finance adalah perusahaan yang bergerak di bidang pembiayaan dana secara kredit, memberikan penawaran kembali terhadap nasabah potensial yang sebelumnya telah melakukan pembiayaan kredit merupakan suatu langkah untuk tetap mendapatkan nasabah kembali, tentunya dengan klasifikasi kriteria yang sesuai.

Oleh karena itu keakuratan informasi sangat dibutuhkan, sehingga diperlukan suatu langkah atau metode yang dapat mempermudah dalam melakukan pemilihan nasabah yang bisa ditawarkan kembali untuk mengajukan kredit di Adira Finance.

Dengan adanya penerapan data mining dalam mengklasifikasi nasabah potensial dengan menggunakan algoritma C4.5 diharapkan dapat membantu mengklasifikasi nasabah, sehingga Adira Finance dapat lebih tepat dalam memberikan pembiayaan kredit terhadap nasabahnya.

### 1.2 Rumusan Masalah

Berdasarkan latar belakang masalah di atas, dapat dibuat suatu rumusan masalah yaitu bagaimana melakukan pengklasifikasian data nasabah Adira Finance menggunakan algoritma C4.5?

### 1.3 Tujuan dan manfaat penelitian

#### 1.3.1 Tujuan

Tujuan dari penelitian ini adalah :

1. Memahami konsep dasar algoritma data mining yaitu algoritma C4.5
2. Mengklasifikasikan data nasabah guna memberikan pembiayaan kredit

3. Mendapatkan pohon keputusan untuk menentukan nasabah potensial pada Adira Finance

### **1.3.2 Manfaat penelitian**

Manfaat dari penelitian ini adalah :

1. Memudahkan analys kredit dalam menentukan nasabah yang potensial.
2. Memberikan masukan kepada adira finance terkait dalam rangka mengambil kebijakan dalam menentukan nasabah potensial

## **II. Landasan Teori**

### **2.1 Konsep Data Mining**

#### **2.1.1 pengertian Data Mining**

menurut Pramudiono dalam bukunya Kusri (2009) mengatakan data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual.

Menurut Eko Prasetyo (2012) data mining juga diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah data mining knowledge discovery.

#### **2.1.2 Teknik Dalam Data Mining**

1. Association Rule Mining  
Association rule mining adalah teknik mining untuk menemukan aturan asosiatif antara suatu kombinasi item. Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, support yaitu persentase kombinasi item tsb. dalam database dan confidence yaitu kuatnya hubungan antar item dalam aturan asosiatif. Algoritma yang paling populer dikenal sebagai Apriori dengan paradigma generate and test, yaitu pembuatan kandidat kombinasi item yang mungkin berdasar aturan tertentu lalu diuji apakah kombinasi item tsb memenuhi syarat support minimum. Kombinasi item yang memenuhi syarat tsb. disebut frequent itemset, yang nantinya dipakai untuk membuat aturan-aturan yang memenuhi syarat confidence minimum. Algoritma baru yang lebih efisien bernama FP-Tree.

#### **2. Classification**

Classification adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Model itu sendiri bisa berupa aturan “jika-maka”, berupa decision tree, formula matematis atau neural network. Decision tree adalah salah satu metode classification yang paling populer karena mudah untuk diinterpretasi oleh manusia. Disini setiap

percabangan menyatakan kondisi yang harus dipenuhi dan tiap ujung pohon menyatakan kelas data. Algoritma decision tree yang paling terkenal adalah C4.5, tetapi akhir-akhir ini telah dikembangkan algoritma yang mampu menangani data skala besar yang tidak dapat ditampung di main memory seperti RainForest. Metode-metode classification yang lain adalah Bayesian, neural network, genetic algorithm, fuzzy, case-based reasoning, dan k-nearest neighbor. Proses classification biasanya dibagi menjadi dua fase : learning dan test. Pada fase learning, sebagian data yang telah diketahui kelas datanya diumpukan untuk membentuk model perkiraan. Kemudian pada fase test model yang sudah terbentuk diuji dengan sebagian data lainnya untuk mengetahui akurasi dari model tsb. Bila akurasinya mencukupi model ini dapat dipakai untuk prediksi kelas data yang belum diketahui

#### **3. Clustering**

Berbeda dengan association rule mining dan classification dimana kelas data telah ditentukan sebelumnya, clustering melakukan penge-lompokan data tanpa berdasarkan kelas data tertentu. Bahkan clustering dapat dipakai untuk memberikan label pada kelas data yang belum diketahui itu. Karena itu clustering sering digolongkan sebagai metode unsupervised learning. Prinsip dari clustering adalah memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar kelas/cluster. Clustering dapat dilakukan pada data yang memiliki beberapa atribut yang dipetakan sebagai ruang multidimensi. Banyak algoritma clustering memerlukan fungsi jarak untuk mengukur kemiripan antar data, diperlukan juga metode untuk normalisasi bermacam atribut yang dimiliki data. Beberapa kategori algoritma clustering yang banyak dikenal adalah metode partisi dimana pemakai harus menentukan jumlah k partisi yang diinginkan lalu setiap data dites untuk dimasukkan pada salah satu partisi, metode lain yang telah lama dikenal adalah metode hierarki yang terbagi dua lagi : bottom-up yang menggabungkan cluster kecil menjadi cluster lebih besar dan top-down yang memecah cluster besar menjadi cluster yang lebih kecil.

#### **2.1.3 Tahapan dalam Data mining**

Menurut Zainul Aras Z,dkk (2016) mengatakan bahwa terdapat 5 tahapan dalam data mining yaitu :

##### **1. Cleaning and integration**

Pembersihan (Cleaning) data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Integrasi (Integration) data

merupakan penggabungan data dari berbagai database kedalam satu database baru.

## 2. Selection and information

Seleksi data sangat diperlukan karena data yang ada pada database seringkali tidak semuanya dipakai. Oleh karena itu, hanya data yang sesuai untuk diproses yang akan diambil dari database. Transformasi data maksudnya adalah data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining.

## 3. Data mining

Merupakan proses utama saat metode diterapkan untuk menentukan pengetahuan berharga dan tersembunyi dari data.

## 4. Evaluation and presentasion

Evaluasi (evaluation) pola yang tepat dari teknik data mining. Serta mampu mempresantasikan (Presentation) apakah output yang ada memang sudah terapan.

## 5. Knowledge

Pengetahuan(Knowledge) yang dapat diketahui dari serangkaian proses penambangan data. Dan ini adalah bagian akhir dari data mining.

## 2.2 Pengertian klasifikasi

Klasifikasi (Classification) adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Model itu sendiri bisa berupa aturan “jika-maka”, berupa decision tree, formula matematis atau neural network. Decision tree adalah salah satu metode classification yang paling populer karena mudah untuk diinterpretasi oleh manusia. Disini setiap percabangan menyatakan kondisi yang harus dipenuhi dan tiap ujung pohon menyatakan kelas data. Algoritma decision tree yang paling terkenal adalah C4.5, tetapi akhir-akhir ini telah dikembangkan algoritma yang mampu menangani data skala besar yang tidak dapat ditampung di main memory seperti RainForest. Metode-metode classification yang lain adalah Bayesian, neural network, genetic algorithm, fuzzy, case-based reasoning, dan k-nearest neighbor. Proses classification biasanya dibagi menjadi dua fase : learning dan test. Pada fase learning, sebagian data yang telah diketahui kelas datanya diumpungkan untuk membentuk model perkiraan. Kemudian pada fase test model yang sudah terbentuk diuji dengan sebagian data lainnya untuk mengetahui akurasi dari model tsb. Bila akurasinya mencukupi model ini dapat dipakai untuk prediksi kelas data yang belum diketahui.

## 2.3 Algoritma C4.5

Menurut Sujana (2010) Algoritma C4.5 merupakan kelompok algoritma *Decision Tree*. Algoritma ini mempunyai input berupa *training samples* dan *samples*. *Training samples* berupa data contoh yang akan digunakan untuk membangun sebuah *tree* yang telah diuji kebenarannya. Sedangkan *samples* merupakan field-field data yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi data .

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan (Decision Tree). Pohon keputusan merupakan metode klasifikasi dan prediksi yang terkenal. Pohon keputusan berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain : ID3, CART, dan C4.5. Algoritma C4.5 merupakan pengembangan dari algoritma ID3, Proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi rule, dan menyederhanakan rule.

Algoritma ini sudah sangat terkenal dan disukai karena memiliki banyak kelebihan. Kelebihan ini misalnya dapat mengolah data numerik dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan performanya merupakan salah satu yang tercepat dibandingkan dengan algoritma lain

Ida dasar dari algoritma ini adalah pembuatan pohon keputusan berdasarkan pemilihan atribut yang memiliki prioritas tertinggi atau dapat disebut memiliki nilai gain tertinggi berdasarkan nilai entropy atribut tersebut sebagai poros atribut klasifikasi. Kemudian secara rekursif cabang-cabang pohon diperluas sehingga seluruh pohon terbentuk. Menurut kamus IGI Global (International Publisher of Progressive Academic), entropy adalah jumlah

data yang tidak relevan terhadap informasi dari suatu kumpulan data Gain adalah informasi yang didapatkan dari perubahan entropy pada suatu kumpulan data, baik melalui observasi atau bisa juga disimpulkan dengan cara melakukan partisipasi terhadap suatu set data

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

- Pilih atribut sebagai akar.
- Buat cabang untuk tiap-tiap nilai.
- Bagi kasus dalam cabang.

- Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Menurut Jiandi data yang dimiliki harus disusun menjadi sebuah tabel berdasarkan kasus dan jumlah responden sebelum dilakukan perhitungan untuk mencari nilai entropy dan gain dapat diselesaikan menggunakan formula sebagai berikut :

$$Entropi(S) = \sum_{j=1}^k -p_j \log_2 p_j$$

Keterangan :

- $S$  adalah himpunan (dataset) kasus
- $k$  adalah banyaknya partisi  $S$
- $p_j$  adalah probabilitas yang di dapat dari Sum(Ya) dibagi Total Kasus.

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|s_i|}{|S|} \times Entropy(s_i)$$

$S$  : Himpunan kasus

$A$  : Atribut

$n$  : Jumlah partisi atribut  $A$

$|S_i|$  : Jmlh kasus pada partisi ke  $i$

$|S|$  : Jumlah kasus dalam  $S$

## 2.4 ADIRA FINANCE

Adira Finance adalah perusahaan yang melayani beragam pembiayaan seperti kendaraan bermotor, baik baru maupun bekas. Saat ini perusahaan Adira menjadi salah satu pembiayaan otomotif terbesar di Indonesia. PT Adira Dinamika Multi Finance Tbk atau Adira Finance didirikan pada tahun 1990 dan mulai beroperasi pada tahun 1991

Adira Finance terus melebarkan sayapnya dengan mengoprasikan 558 jaringan usaha di seluruh Indonesia termasuk di kabupaten Pringsewu. Adira Finance pringsewu berkedudukan di JL.Ahmad Yani No.21-22 Sidoharjo kabupaten pringsewu provinsi lampung dengan setatus kedudukan sebagai satelit cabang lampung1.

## III. Metodologi penelitian

### 3.1 Pengumpulan Data

Dalam penelitian ini penulis metode pengumpulan data sebagai berikut :

1. Wawancara

Pada tahap ini penulis mewawancarai beberapa pihak yang terkait dalam hal ini berupa pegawai Adira Finance. Wawancara mengenai data kredit

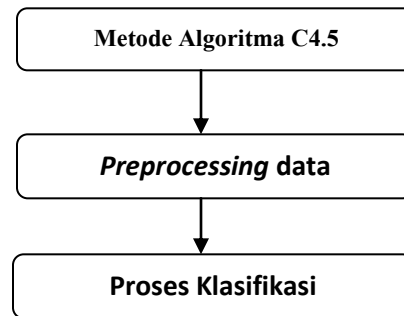
nasabah seperti pendapatan, pinjaman, angsuran, waktu, kategori.

2. Dokumentasi

Melalui metode ini, penulis mengumpulkan dokumen, arsip, dan data-data nasabah adira finance sebagai bahan penelitian.

3. Studi pustaka

Dalam hal ini penulis mengumpulkan dan mempelajari buku-buku maupun literatur-literatur yang berhubungan dengan masalah yang diteliti. Dari studi pustaka ini didapatkan banyak referensi yang akan mendukung penelitian ini, sekaligus untuk memperkuat pengetahuan dasar dan teori yang digunakan pada penelitian ini.



3.1 Gambar alur Penelitian

### 3.2 Metode Algoritma C4.5

Ada beberapa tahapan dalam membuat sebuah pohon keputusan dalam algoritma C4.5 yaitu :

1. Mempersiapkan data training, dapat diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon dengan menghitung nilai gain yang tertinggi dari masing-masing atribut atau berdasarkan nilai index entropy terendah. Sebelumnya dihitung terlebih dahulu nilai index entropy, dengan rumus :

$$Entropy(i) = - \sum_{j=1}^m f(i,j) \cdot \log_2 f(i,j)$$

Keterangan :

$i$  = himpunan kasus

$m$  = jumlah partisi  $i$

$f(i,j)$  = proporsi  $j$  terhadap  $i$

- 3) Hitung nilai gain dengan rumus :

$$Entropy\ split = - \sum_{i=1}^p \frac{n1}{n} \cdot IE(i)$$

Keterangan :

p = jumlah partisi atribut  
 ni = proporsi ni terhadap i  
 n = jumlah kasus dalam n

- 4) Ulangi langkah ke-2 hingga semua record terpartisi. Proses partisi pohon keputusan akan berhenti disaat :
  - a. Semua tupel dalam record dalam simpul m mendapat kelas yang sama. Tidak ada atribut dalam record yang dipartisi lagi
- 5) Tidak ada record di dalam cabang yang kosong. Algoritma C4.5 menggunakan konsep information gain atau entropy reduction untuk memilih pembagian yang optimal [4]. Tahapan dalam membuat sebuah pohon keputusan dengan algoritma C4.5 [2] yaitu :
  - a. Mempersiapkan data training, dapat diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas-kelas tertentu.
  - b. Atribut atau berdasarkan nilai index entropy terendah. Sebelumnya dihitung terlebih dahulu nilai index entropy, dengan rumus :

$$Entropy(i) = - \sum_{j=1}^m f(i,j) \cdot \log_2 f(i,j)$$

Keterangan :

i = himpunan kasus  
 m = jumlah partisi i  
 f(i,j) = proporsi j terhadap i

- c. Hitung nilai gain dengan rumus :

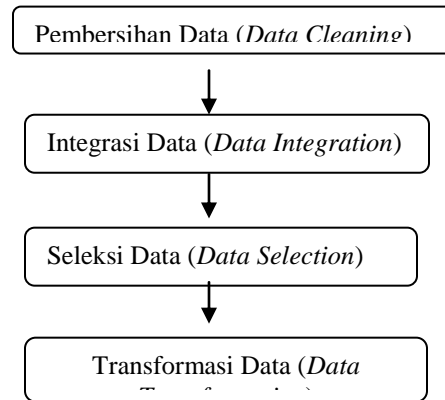
$$Entropy(i) = - \sum_{j=1}^m f(i,j) \cdot \log_2 f(i,j)$$

Keterangan :

p = jumlah partisi atribut  
 ni = proporsi ni terhadap i  
 n = jumlah kasus dalam n

- 6). Ulangi langkah ke-2 hingga semua record terpartisi. Proses partisi pohon keputusan akan berhenti disaat :
  - a. Semua tupel dalam record dalam simpul m mendapat kelas yang sama.
  - b. Tidak ada atribut dalam record yang dipartisi lagi.
  - c. Tidak ada record di dalam cabang yang kosong

### 3.3 Preprocessing data



3.2 Gambar alur Preprocessing data

#### 1. 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*

#### 2. 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.

#### 3. 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.

#### 4. Transformasi Data (Data Transformation)

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

### 3.4 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):  $Info(D) = -\sum p_i \log_2(p_i)$

dimana *m* merupakan banyaknya jenis kategori nilai pada atribut *C*,  $p_i = |C_i, D|/|D|$  merupakan probabilitas dari tupel *D* yang mempunyai kelas *C<sub>i</sub>*.

Misalkan terdapat atribut *A* yang memiliki *v* nilai yang berbeda {*a*<sub>1</sub>,*a*<sub>2</sub>,...*a<sub>v</sub>*}. Atribut *A* dapat digunakan untuk membagi *D* ke dalam *v* partisi {*D*<sub>1</sub>,*D*<sub>2</sub>,...*D<sub>v</sub>*}, dimana *D<sub>j</sub>* memuat tupel *D* yang memiliki nilai *a<sub>j</sub>* dari *A*. Persamaan untuk mencari nilai *entropy* dari subset *A* sebagai berikut (Han, *et al*, 2006: 298):  $InfoA(D) = E(A) = \sum |D_j|/|D| \times Info(D_j)$

dimana *E(A)* adalah *entropy* dari subset *A*, *v* merupakan banyaknya jenis kategori nilai pada subset *A*,  $|D_j|/|D|$  merupakan bobot dari subset *j* dan jumlah sampel pada subset yang mempunyai nilai *a<sub>j</sub>* 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:

$$split\ InfoA(D) = -\sum |D_j|/|D| \times \log_2 |D_j|/|D|$$

$$Gain\ Ratio(A) = Gain(A) / split\ Info(A)$$

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

## IV. Hasil dan Pembahasan

### 4.1 Data Penelitian

Data penelitian ini menggunakan data nasabah kredit Adira Finance. Setelah melakukan serangkaian tahapan data mining Penelitian ini menggunakan 100 dataset dengan 7 atribut yang terdiri 6 atribut *predictor* dan 1 atribut tujuan. Atribut *predictor* yang digunakan dalam data kredit adalah Umur, Jenis Kelamin, Status Perkawinan, Pekerjaan, Pendapatan dan Tagihan Listrik. Sedangkan atribut

terakhir adalah sebagai output berupa keputusan yang digunakan untuk membedakan hasil prediksi, yaitu Berpotensi atau Tidak Berpotensi.

Nama	Jenis Kelamin	Umur	Status Perkawinan	Pekerjaan	Penghasilan	Tagihan Listrik	Klasifikasi	
jamil	Pria	40	Laki-laki	menikah	pedagang	Besar	Sedang	Berpotensi
misran	Pria	35	Laki-laki	menikah	PNS	Besar	Sedang	Berpotensi
joko prasidingsuwa Timur	Pria	30	Laki-laki	Belum Menikah	Pegawai swasta	Besar	Sedang	Tidak berpotensi
karnoso	Pria	25	Laki-laki	menikah	Pegawai swasta	Besar	Sedang	Tidak berpotensi
Tagih andi paryatuk	Pria	20	Laki-laki	Belum Menikah	Pegawai swasta	Besar	Sedang	Berpotensi
Wijanto	Pria	15	Laki-laki	menikah	Pegawai swasta	Besar	Sedang	Tidak berpotensi
Shinta ma Bumihutan	Perempuan	10	perempuan	menikah	Pegawai swasta	Besar	Sedang	Tidak berpotensi
melki dan Ropani	Pria	5	Laki-laki	menikah	PNS	Besar	Sedang	Tidak berpotensi
joko supri Sidiyarto	Pria	45	Laki-laki	Belum Menikah	Pegawai swasta	Besar	Sedang	Berpotensi
marini	Perempuan	40	perempuan	menikah	pedagang	Besar	Sedang	Tidak berpotensi
Wiganto	Pria	35	Laki-laki	menikah	PNS	Besar	Sedang	Berpotensi
Isdiyah	Perempuan	30	perempuan	menikah	pedagang	Besar	Sedang	Berpotensi
Yennyati	Pria	25	Laki-laki	menikah	Buruh	Besar	Sedang	Tidak berpotensi
Agus Riyadi Podomoro	Pria	20	Laki-laki	Belum Menikah	Pegawai swasta	Besar	Sedang	Berpotensi
merudin	Pria	15	Laki-laki	Belum Menikah	pedagang	Besar	Sedang	Berpotensi
Ulowenter podomoro	Pria	10	Laki-laki	Belum Menikah	PNS	Besar	Sedang	Berpotensi
Tuti Sunda Pringsewu Utara	Perempuan	5	perempuan	Belum Menikah	PNS	Besar	Sedang	Tidak berpotensi
Tariati	Perempuan	45	perempuan	menikah	Buruh	Besar	Sedang	Tidak berpotensi
gandi mu Pringsewu Timur	Paruh Baya	40	menikah	Pegawai swasta	Besar	Sedang	Tidak berpotensi	
hukman	Pria	35	Laki-laki	menikah	Buruh	Besar	Sedang	Tidak berpotensi
Murtama popyatuk	Tua	40	Laki-laki	menikah	Pegawai swasta	Besar	Sedang	Tidak berpotensi
Dwi Caydi Podomoro	Laki-laki	20	Laki-laki	Belum Menikah	Pegawai swasta	Besar	Kecil	Berpotensi

4.1 Gambar Atribut Data Kredit Nasabah

## 4.2 Implementasi

### 4.2.1 Data Yang Akan Diolah Dengan WEKA

Ketentuan-ketentuan untuk data yang akan diolah dengan WEKA adalah :

- Jika data yang disiapkan dalam bentuk Excel maka tidak boleh lebih dari satu *sheet*.
- Lalu simpan dengan format *CSV (comma separated value)*.

Dalam penyimpanan tulis nama file dilanjutkan titik *CSV*. Contohnya : data akhir.CSV.

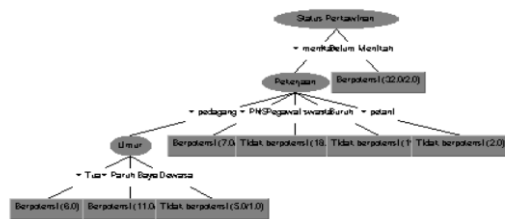
### 4.2.2 Analisa Decision Tree Menggunakan WEKA



Untuk mencari *knowledge* dengan algoritma C4.5 atau J48 maka dilakukan langkah-langkah sebagai berikut :

- Pada jendela utama WEKA klik tombol *explorer*. Lalu muncul jendela WEKA Explorer, kemudian pada *tab Preprocess* klik tombol *open file* untuk membuka data.
- Selanjutnya keluar kotak dialog WEKA dan klik tombol *All* untuk mengeksekusi data.
- Selanjutnya pada kotak dialog WEKA, aktifkan *tab classify*. Lalu klik tombol *Choose* pada *tab Classify* untuk mencari algoritma yang digunakan.
- Pada menu *classify* pilih *Root Tree* dan klasifikasi dengan pohon keputusan, yaitu metode klasifikasi dengan struktur pohon yang merepresentasikan kriteria pembagian kelas dan kelas-kelas dengan menggunakan metode algoritma J48.

- Setelah *classifiers*, pilihan-pilihan pengujian, dan *class* telah ditentukan, proses pembelajaran dapat dimulai dengan mengklik tombol *Start*, *user* dapat menghentikan proses ini sewaktu-waktu dengan tombol *Stop*. Saat *training* selesai, area *output classifiers* di sebelah kanan menampilkan *teks* yang menggambarkan hasil *training* dan pengujian. Sebuah *entry* baru saja muncul di kotak *Result list*.
- Untuk membuat pohon keputusan maka klik kanan *trees-J.48* pada menu *result list* kemudian pilih *Visualize tree*.
- Hasil *Visualize Tree* adalah hasil akhir dari pohon keputusan.



4.2 Gambar pohon keputusan

Di mana *rule* yang dihasilkan pada gambar 4.2 adalah :

- IF status perkawinan me nikah pekerjaan pedagang umur Tua, THEN Berpotensi
- IF status perkawinan menikah pekerjaan pedagang umur Paruh Baya, THEN Berpotensi
- IF status perkawinan me nikah pekerjaan pedagang umur Dewasa, THEN Tidak Berpotensi
- IF Menikah pekerjaan PNS, THEN Berpotensi
- IF Menikah pekerjaan Pegawai Swasta, THEN Tidak Berpotensi
- IF Menikah pekerjaan Buruh, THEN Tidak Berpotensi
- IF Menikah pekerjaan Petani, THEN Tidak Berpotensi
- IF Status Perkawinan belum menikah, THEN Berpotensi

#### 4.2.3 Pembahasan Hasil Pengujian

Dari pohon keputusan dan *rule* yang terbentuk dimana status berpotensi dan tidak berpotensi dilihat dari atribut yang ada pada data awal yaitu data nasabah Adira Finance dan hasilnya atribut Status Perkawinan yang menjadi akar, pekerjaan dan umur menjadi dahan atau node . rincian jumlah nasabah yang berpotensi dan tidak berpotensi terlihat pada tabel berikut :

NO	Rule	jumlah	
		Berpotensi	Tidak Berpotensi
1	Menikah, pedagang, tua	6	
2	Menikah, pedagang, paru baya	12	
3	menikah, pedagang, dewasa		5
4	menikah, PNS	6	
5	Menikah, pegawai swasta		18
6	Menikah, petani		2
7	Menikah, Buruh		20
8	belum menikah	31	

4.3 Tabel rincian jumlah berdasarkan Rule

## V. Penutup

### 5.1 Kesimpulan

- Analisis klasifikasi pada data kredit Adira Finance 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.
- Metode pohon keputusan (*decision tree*) yang diproses dengan *software* WEKA dapat mengidentifikasi kelayakan kredit dengan baik.

### 5.2 Saran

- Penelitian ini dapat dilanjutkan dengan menambahkan variabel-variabel yang memiliki hubungan dengan masalah pengajuan perpanjangan kredit agar nantinya tingkat pemberian keputusan yang lebih baik dapat tercapai.
- Penelitian ini dapat dilanjutkan dengan menggunakan metode-metode *data mining* lainnya untuk mendapatkan perbandingan metode yang lebih.

### DAFTAR PUSTAKA

- Kusrini dan Taufiq Luthfi Emha. "Algoritma Data Mining". Yogyakarta: Andi.
- Prasetyo, Eko, (2012) . *Data Mining*, Andi, Yogyakarta.
- sunjana. (2010) " *Klasifikasi Data Nasabah sebuah asuransi menggunakan algoritma 4.5.*" Seminar Nasional Aplikasi Teknologi Informasi. 1907-5022.
- Aras, Zainul., Sarjono. (2016). " *Analisis Data Mining untuk menentukan kelompok penerima bantuan bedah rumah menggunakan K-means*, Vol 1 No.2 ISSN 2540-8011.

- [5] Sholihah, Hidayatus,. Liani, Pipit,. Alita, Asti Ningrum. (2017). *“Laporan KKP Adira Finance”*
- [6] Han, J., Kamber, M., & Pei, J. (2006). *Data Mining: Concept and Techniques, Second Edition*. Waltham: Morgan Kaufmann.
- [7] Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concept and Techniques, Third Edition*. Waltham: Morgan Kaufmann.