

# Penerapan Algoritma Klasifikasi C4.5 dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus : PT Atria Artha Persada

Muhammad Fauzul Arifin<sup>1</sup>, Devi Fitrihanah<sup>2</sup>

*Teknik Informatika, Universitas Mercu Buana, Jakarta*  
41513110073@student.mercubuana.ac.id<sup>1</sup>, devi.fitrihanah@mercubuana.ac.id<sup>2</sup>

## **Abstrak**

Permasalahan yang sering muncul dalam bisnis pada penjualan dengan sistem pembayaran kredit tempo adalah antara lain seperti kredit macet, order fiktif dan penipuan. Jika tidak menggunakan prosedur yang benar dalam penerimaan mitra penjualan bukan tidak mungkin masalah tersebut akan timbul dalam proses bisnis. Untuk mengembangkan bisnis tentu perlu meningkatkan penjualan dan dibarengi dengan pembayaran yang lancar. Oleh karena itu, pada proses penerimaan mitra perlu dibuatkan prosedur yang baik agar dapat meminimalisir kemungkinan masalah yang ada. Pada prosedur penerimaan mitra penjualan perlu dilakukan proses analisa untuk menentukan calon mitra penjualan tersebut diterima atau ditolak. Dari data mitra yang ada dapat digunakan untuk dilakukan pengolahan data. Salah satu teknik pengolahan data yang dapat digunakan dalam proses tersebut adalah klasifikasi. Klasifikasi adalah teknik pengolahan data yang membagi objek menjadi beberapa kelas sesuai dengan jumlah kelas yang diinginkan. Dan menggunakan Algoritma C4.5 dalam proses pengklasifikasi-annya. Yang dapat menentukan apakah calon mitra dapat diterima sebagai mitra atau tidak. Kemudian, Hasil dari pengklasifikasiannya divalidasi dengan ten-fold cross validation dengan tingkat akurasi 96,26 %, presisi 100% dan recall 71,43%.

**Keywords:** Algoritma C4.5, Klasifikasi, Pohon Keputusan, *Ten-fold cross Validation, Confusion Matrix*

Received December 2017

Accepted for Publication January 2018

DOI: 10.22441/incomtech.v8i1.2198

## **1. PENDAHULUAN**

PT. Atria Artha Persada merupakan perusahaan yang termasuk kedalam tipe usaha kecil menengah yang bergerak pada bidang distribusi ban kendaraan.

Dalam bisnisnya, perusahaan ini sering kali dihadapkan pada masalah penipuan yang dilakukan oleh mitra usahanya. Masalah seperti penipuan, order fiktif, pembayaran macet sampai ditinggal oleh mitra tanpa kabar pun pernah dialami. Sejauh ini dalam menentukan kelayakan mitra, PT. Atria Artha Persada memiliki standar yang ditetapkan pada saat pemesanan awal, yaitu dengan mendata toko-toko calon mitra yang mengajukan. Data calon mitra tersebut kemudian dipelajari dan dijadikan sebagai dasar untuk memutuskan apakah menerima atau menolak toko tersebut sebagai mitra. Hampir semua keputusan mengacu kepada data calon yang dimiliki, padahal data yang diminta dari calon mitra sudah dikatakan cukup lengkap. Hal ini terjadi dikarenakan analis (orang yang mempelajari data calon mitra dan memutuskan hasilnya) hanya menentukan calon mitra berdasarkan data mitra dan tidak melakukan analisis terhadap data calon yang pernah ada. Analisis tidak mencoba untuk melihat dokumentasi sejarah dari setiap mitra yang baik dan yang tidak baik. Keputusan yang diambil hampir 40% membawa kepada mitra yang merugikan.

Berdasarkan masalah tersebut, perlu dilakukan analisa untuk memberikan sebuah solusi terhadap permasalahan dalam menentukan calon mitra yang menggunakan pendekatan berbasis kepada data atau biasa disebut sebagai *data driven method*. Dengan menggunakan *data driven method*, pengambilan keputusan didasari kepada pembelajaran data calon mitra yang sudah ada sebelumnya. Dalam terminologi *data mining*, hal ini disebut sebagai *supervised learning*. Algoritma yang digunakan untuk dapat mengklasifikasi data yang sudah ada adalah algoritma C4.5.

Algoritma ini sudah banyak digunakan dalam implementasi penentuan keputusan karena memiliki banyak kelebihan. Salah satu 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[1], dapat juga digunakan untuk mengetahui kelompok data yang terkait[2] dan dapat memprediksi kelas objek yang tidak diketahui[3]. Algoritma klasifikasi terdapat berbagai pendekatan, ada yang basisnya pohon keputusan, induksi berbasis aturan, jaringan syaraf tiruan, algoritma genetika dan jaringan Bayesian[4]. Algoritma C4.5 adalah algoritma yang sangat populer untuk klasifikasi dalam *machine learning* dan pengolahan data[5]. Dalam algoritma C4.5 terjadi proses pengklasifikasian data dalam bentuk tingkatan yang berbeda mulai dari akar ke daun. Proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi rule, dan menyederhanakan rule[6]. Dari rule atau aturan yang terbentuk ini nantinya dapat digunakan dalam menentukan kelayakan penerimaan calon mitra.

## 2. ALGORITMA C4.5

Algoritma C4.5 adalah algoritma yang digunakan untuk menghasilkan sebuah pohon keputusan yang dikembangkan oleh Ross quinlan[7]. Ide 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[8]. Pada tahapannya algoritma C4.5 memiliki 2 prinsip kerja[9], yaitu: Membuat pohon keputusan, dan membuat aturan-aturan (*rule model*). Aturan aturan yang terbentuk dari pohon keputusan akan membentuk suatu kondisi dalam bentuk *if then*.

Terdapat empat langkah dalam proses pembuatan pohon keputusan pada algoritma C4.5, yaitu:

- Memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada.
- Membuat cabang untuk masing-masing nilai, artinya membuat cabang sesuai dengan jumlah nilai variabel gain tertinggi.
- Membagi setiap kasus dalam cabang, berdasarkan perhitungan nilai gain tertinggi dan perhitungan dilakukan setelah perhitungan nilai gain tertinggi awal dan kemudian dilakukan proses perhitungan gain tertinggi kembali tanpa meyeritakan nilai variabel gain awal.
- Mengulangi proses dalam setiap cabang sehingga semua kasus dalam cabang memiliki kelas yang sama, mengulangi semua proses perhitungan gain tertinggi untuk masing-masing cabang kasus sampai tidak bisa lagi dilakukan proses perhitungan.

Algoritma C4.5 secara rekursif mengunjungi setiap simpul keputusan, memilih pembagian yang optimal, sampai tidak bisa dibagi lagi. Dari ketiga peneliti yang telah dilakukan tersebut, klasifikasi dengan Algoritma C4.5 digunakan oleh para peneliti sebagai solusi untuk mengambil keputusan yang diharapkan mampu membantu dalam pengambilan keputusan dengan lebih mudah dan cepat[9].

Dalam penerapan dan penggunaan algoritma C4.5, dapat digunakan untuk melakukan prediksi dan klasifikasi terhadap calon pegawai yang berpotensi untuk masuk ke dalam perusahaan dengan cara membuat pohon keputusan berdasarkan data-data yang sudah ada dan melakukan prediksi terhadap calon pegawai baru yang ingin masuk ke perusahaan[10]. Selain itu algoritma ini digunakan untuk klasifikasi predikat keberhasilan mahasiswa disebuah universitas. Variabel yang memiliki prioritas utama terhadap predikat keberhasilan mahasiswa adalah mahasiswa yang memilih sesi perkuliahan pada pagi hari[11].

## 2.1 Entropy Dan Information Gain

*Entropy* adalah ukuran dari teori informasi yang dapat mengetahui karakteristik dari impurity dan homogeneity dari kumpulan data. Dari nilai *Entropy* tersebut kemudian dihitung nilai *information gain* masing-masing atribut[8]. Penghitungan nilai *Entropy* digunakan rumus seperti dalam Persamaan (1)[12].

$$Entropy(S) = \sum_{i=0}^n - p_i * \log_2(p_i) \dots\dots\dots(1)$$

Rumus (1) merupakan rumus yang digunakan dalam perhitungan entropy yang digunakan untuk menentukan seberapa informatif atribut tersebut. Berikut keterangannya :

- s : Himpunan kasus
- n : Jumlah partisi
- $p_i$  : Jumlah kasus pada partisi ke- $i$

*Information 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[8].

$$GAIN(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S) \dots (2)$$

Rumus (2) merupakan rumus yang digunakan dalam perhitungan *information gain* setelah melakukan perhitungan *entropy*. Berikut keterangannya :

- s : Himpunan kasus
- n : Jumlah partisi atribut A
- $|S_i|$  : Jumlah kasus pada partisi ke- $i$
- $|S|$  : Jumlah kasus dalam s

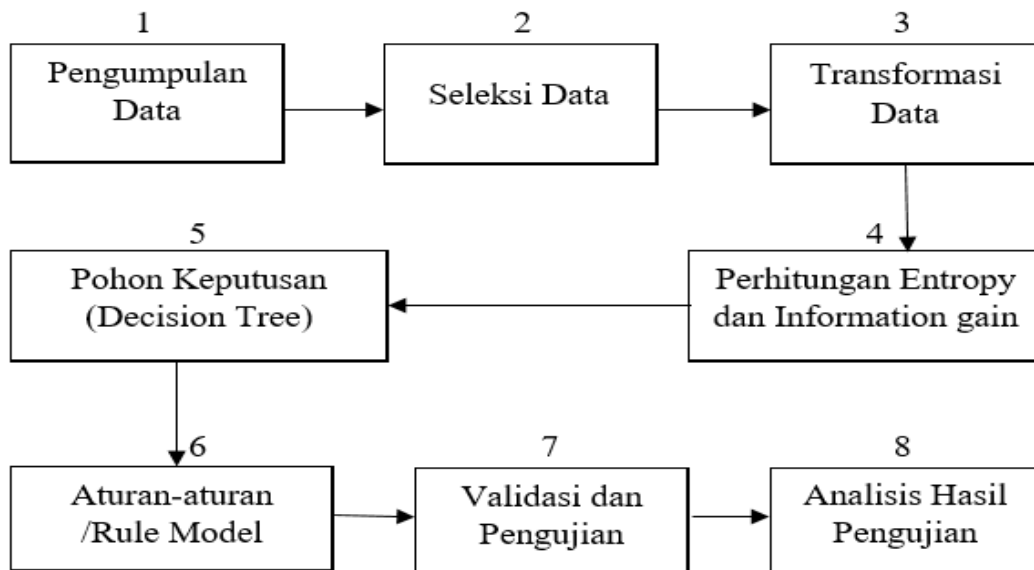
Dengan mengetahui rumus-rumus diatas, data yang telah diperoleh dapat dimasukkan dan diproses dengan algoritma C4.5 untuk proses pembuatan *decision tree*.

## 2.2 Pohon Keputusan(*Decision Tree*)

Pohon keputusan adalah sebuah struktur yang dapat digunakan untuk mengubah data menjadi pohon keputusan yang akan menghasilkan aturan-aturan keputusan besar menjadi himpunan-himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan[13]. Pohon keputusan yang dihasilkan oleh algoritma C4.5 dapat digunakan untuk klasifikasi[14]. Pohon keputusan adalah salah satu teknik penambangan data yang paling populer untuk penemuan pengetahuan. Secara sistematis menganalisis dan mengekstrak aturan untuk tujuan Klasifikasi / prediksi[15].

## 3. METODE PENELITIAN

Metode Penelitian yang digunakan dalam penerapan algoritma C4.5 untuk rekomendasi penerimaan calon mitra penjualan baru, menggunakan rancangan penelitian yang ditunjukkan pada Gambar 1 berikut :



Gambar 1. Flowchart Rancangan Proses Penelitian

- **Pengumpulan Data**

Pengumpulan data adalah mengumpulkan data-data yang akan digunakan dalam proses algoritma klasifikasi C4.5.

- **Seleksi Data**

Seleksi data adalah memilih data yang akan digunakan dalam proses algoritma klasifikasi C4.5. Tujuan dari seleksi data adalah menciptakan himpunan data target, pemilihan himpunan data, atau memfokuskan pada subset variabel atau sampel data, dimana penemuan (discovery) akan dilakukan[16].

- **Transformasi Data**

Transformasi data adalah proses mentransformasi atau mengubah data ke dalam bentuk yang sesuai, agar dapat di proses dengan perhitungan algoritma C4.5. Transformasi Data adalah proses mentransformasi atau menggabungkan data ke dalam bentuk yang sesuai untuk penggalian lewat operasi summary atau aggregation[16].

- **Perhitungan Entropy dan Information Gain**

Perhitungan semua atribut/variabel, *entropy* menggunakan rumus (1) dan *information gain* menggunakan rumus (2) untuk mengetahui *information gain* tertinggi yang akan dijadikan simpul akar pada pembuatan pohon keputusan.

- **Pohon Keputusan(Decision Tree)**

Pohon keputusan adalah hasil dari proses perhitungan *entropy* dan

*information gain*, setelah perhitungan berulang-ulang sampai semua atribut pohon memiliki kelas dan tidak bisa lagi dilakukan proses perhitungan.

- **Aturan-aturan/ Rule Model**

Aturan-aturan/ Rule model adalah uraian penjelasan yang merepresentasikan sebuah pohon keputusan.

- **Validasi dan Pengujian**

Validasi dan pengujian adalah Pengujian yang dilakukan untuk mengetahui semua fungsi bekerja dengan baik atau tidak. Validasi dilakukan dengan *Ten-fold Cross Validation*. *Ten-fold Cross Validation* adalah validasi yang dilakukan dengan cara membagi suatu set data menjadi sepuluh segmen yang berukuran sama besar dengan cara melakukan pengacakan data. Validasi dan pengujian dilakukan untuk mengetahui tingkat akurasi, presisi, dan recall dari hasil prediksi klasifikasi. Akurasi adalah persentase dari catatan yang diklasifikasikan dengan benar dalam pengujian dataset. Presisi adalah persentase data yang diklasifikasikan sebagai model baik yang sebenarnya juga baik. Recall adalah pengukuran tingkat pengenalan positif sebenarnya[17].

- **Analisis Hasil Pengujian**

Analisa yang dilakukan untuk memastikan bahwa hasil pengujian benar-benar sesuai dengan pembahasan. Analisa dilakukan dengan melakukan perhitungan kembali hasil validasi dan pengujian(akurasi, presisi, dan recall) secara manual. Apakah perhitungan yang dilakukan akan menghasilkan nilai yang sama atau tidak, dibantu dengan *Confusion matrix*. *Confusion Matrix* adalah model yang akan membentuk matrix yang terdiri dari true positif atau tupel positif dan true negatif atau tupel negatif[16]. *Confusion matrix* berisi informasi aktual (actual) dan prediksi (predicted) pada sistem klasifikasi[18]. *Confusion matrix* dapat memvisualisasi kinerja algoritma klasifikasi[5].

#### 4. HASIL DAN PEMBAHASAN

Adapun hasil dan tahapan proses pembahasan dalam algoritma klasifikasi C4.5 adalah sebagai berikut :

##### 4.1 Pengumpulan Data

Pengumpulan data dilakukan dengan cara meminta data calon mitra penjualan baru bulan September sampai dengan Oktober 2017 di PT. Atria Artha Persada. Data ini mencakup beberapa variabel, antara lain : No, Nama, No. Tlp, Alamat, Status Tempat, Tempo, Cara Pembayaran, Omset/Bulan, Target Pembelian/ Bulan, yang akan diseleksi dan diolah menggunakan algoritma C4.5. Jumlah data yang akan digunakan adalah 107 data calon mitra penjualan baru bulan September - Oktober 2017.

Tabel 1. Struktur Data calon mitra penjualan yang akan dijadikan data, dalam proses algoritma klasifikasi C4.5.

No.	Customer Name	No.Tlp	Alamat	Status Tempat	Tempo	Cara Pembayaran	Omset (Juta) /Bulan	Target Pembelian (Pcs) /Bulan
1	Alie Bintang Motor	081295xxxxx	Ciledug	Hak Milik	45 Hari	Tunai	20	20
2	Aneka Ban Cinere	021-7470xxxx/ 08170xxxxxx	Cinere	Kontrak	45 Hari	Transfer	50	50
3	Pojok Ban	08569xxxxxx / 082211xxxxx	Pondok Aren	Kredit	45 Hari	Transfer	75	50
4	Anugrah Motor		Ciputat	Hak Milik	45 Hari	Transfer	40	20
5	Asia Motor	0813180xxxx	Kramat Jati	Hak Milik	45 Hari	Transfer	50	20
6	Bahagia Motor	0812811xxxx, 0821258xxxx	Bekasi	Hak Milik	30 Hari	Transfer	60	20
7	Den Motor		Bitung	Kredit	45 Hari	Transfer	30	20
...	...	...	...	...	...	...	...	...
107	Raya Motor	8815186232	Serang	Kontrak	45 Hari	Transfer	30	20

#### 4.2 Seleksi Data

Pemilihan atribut/variabel tersebut dengan pertimbangan bahwa jumlah nilai variabelnya tidak banyak sehingga diharapkan calon mitra yang masuk dalam satu klasifikasi nilai variabel tersebut cukup banyak. Dari data yang ada, variabel diambil sebagai atribut/variabel dalam pembentukan pohon keputusan adalah : No. Tlp, Alamat, Status Tempat, Tempo, Cara Pembayaran, Omset/Bulan, Target Pembelian/Bulan.

#### 4.3 Transformasi Data

Transformasi data dilakukan dengan mengubah beberapa nilai atribut yang awalnya bernilai nominal menjadi nilai-nilai atribut yang sesuai dengan data pada Tabel 2 agar dapat dilakukan proses perhitungan algoritma klasifikasi C4.5.

#### 4.4 Perhitungan *Entropy* dan *Information Gain*

Dengan menggunakan rumus (1), kita dapat menghitung nilai *Entropy*. Menghitung *Entropy* Total dilakukan dengan cara menghitung jumlah keputusan “DITERIMA” dan “DITOLAK” dari seluruh kasus yang ada.

$$- Entropy(\text{Total}) \left( -\frac{93}{107} * \log_2 \left( \frac{93}{107} \right) \right) + \left( -\frac{14}{107} * \log_2 \left( \frac{14}{107} \right) \right) = 0.559740$$

*Entropy* Total adalah menghitung nilai total keputusan DITERIMA (93) dan menunggu (14), sedangkan 107 adalah jumlah keseluruhan kasus. Kemudian hitung masing-masing *Entropy* dari semua nilai variabel yang ada. Dan dengan

menggunakan rumus (2) kita dapat menghitung nilai *Information Gain* tiap-tiap Atribut.

Tabel 2. Tabel Transformasi data mitra penjualan yang diubah menjadi beberapa nilai atribut/variabel menjadi nilai-nilai atribut yang lebih cocok untuk melakukan proses klasifikasi.

No	No.Tlp	Alamat	Status Tempat	Tempo	Cara Pembayaran	Omset (Juta) /Bulan	Target Pembelian (Pcs) /Bulan	Rekomendasi
1	Valid	Tervalidasi	Hak Milik	45 Hari	Tunai	20	20	Diterima
2	Valid	Tervalidasi	Kontrak	45 Hari	Transfer	50	50	Diterima
3	Valid	Tervalidasi	Kredit	45 Hari	Transfer	75	50	Diterima
4	Tidak Valid	Tervalidasi	Hak Milik	45 Hari	Transfer	40	20	Ditolak
5	Valid	Tervalidasi	Hak Milik	45 Hari	Transfer	50	20	Diterima
6	Valid	Tervalidasi	Hak Milik	30 Hari	Transfer	60	20	Diterima
7	Tidak Valid	Tervalidasi	Kredit	45 Hari	Transfer	30	20	Ditolak
...	...	...	...	...	...	...	...	...
107	Valid	Tervalidasi	Kontrak	45 Hari	Transfer	30	20	Diterima

- Gain(Total, No. Tlp) =  $0.559740 - \left( \left( \frac{100}{107} * 0.365924 \right) + \left( \frac{7}{107} * 0 \right) \right) = 0.217755$
- Gain(Total, Alamat) =  $0.559740 - \left( \left( \frac{104}{107} * 0.487021 \right) + \left( \frac{3}{107} * 0 \right) \right) = 0.086373$
- Gain(Total, Status Tempat) =  $0.559740 - \left( \left( \frac{68}{107} * 0.430552 \right) + \left( \frac{23}{107} * 0.666578 \right) + \left( \frac{14}{107} * 0.749595 \right) + \left( \frac{2}{107} * 1 \right) \right) = 0.026066$
- Gain(Total, Tempo) =  $0.559740 - \left( \left( \frac{8}{107} * 0.811278 \right) + \left( \frac{96}{107} * 0.482066 \right) + \left( \frac{3}{107} * 0.918296 \right) \right) = 0.040829$
- Gain(Total, Cara Pembayaran) =  $0.559740 - \left( \left( \frac{7}{107} * 0.863121 \right) + \left( \frac{98}{107} * 0.506650 \right) + \left( \frac{2}{107} * 1 \right) \right) = 0.020547$
- Gain(Total, Omset/Bulan) =  $0.559740 - \left( \left( \frac{21}{107} * 0.791858 \right) + \left( \frac{56}{107} * 0.543564 \right) + \left( \frac{25}{107} * 0.402179 \right) + \left( \frac{5}{107} * 0 \right) \right) = 0.025879$
- Gain(Total, Target Pembelian/Bulan) =  $0.559740 - \left( \left( \frac{86}{107} * 0.583019 \right) + \left( \frac{13}{107} * 0.391243 \right) + \left( \frac{7}{107} * 0.591673 \right) + \left( \frac{1}{107} * 0 \right) \right) = 0.026066$



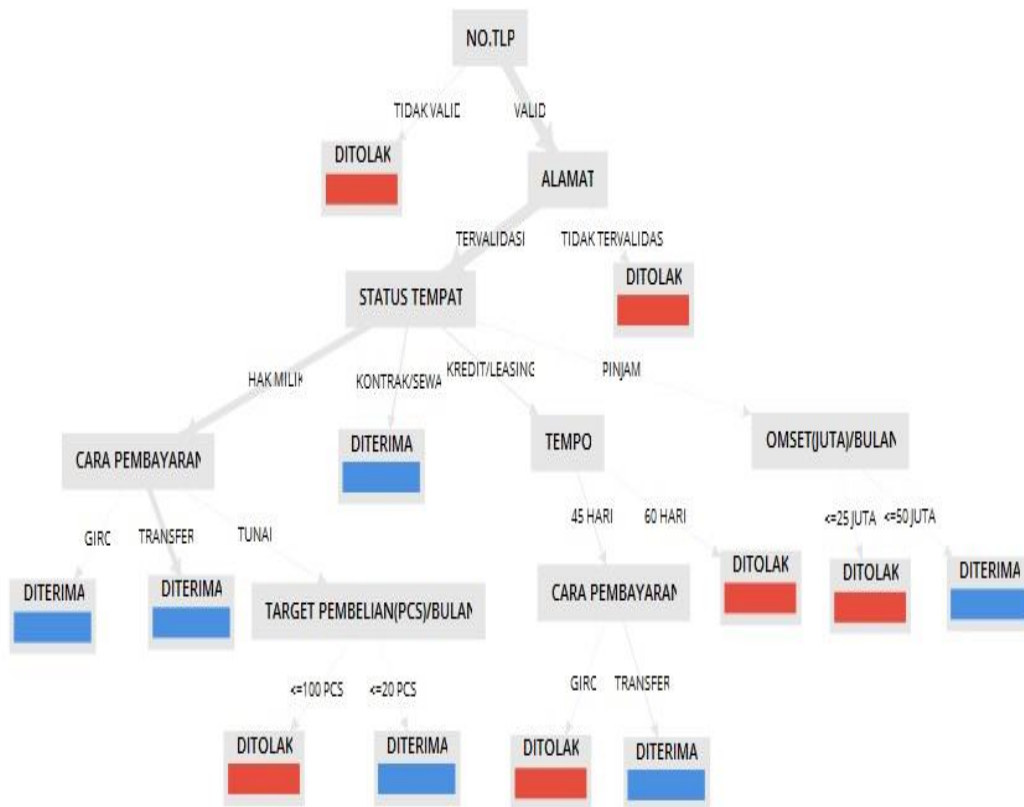
Setelah semua nilai *Entropy* dan *Information Gain* dihitung, kemudian hasil dari perhitungan tersebut dimasukkan ke dalam tabel 3. Dari perhitungan nilai *entropy* dan *information gain* diatas, dapat diketahui bahwa nilai *information gain* terbesar adalah variabel No. Tlp dengan nilai 0,217755 dan yang terkecil adalah Cara Pembayaran, dengan nilai 0,020547. Kemudian, ambil nilai *Information Gain* terbesar yaitu atribut No. Tlp dan jadikan simpul akar. Hilangkan atribut yang dipilih sebelumnya dan ulangi perhitungan nilai *Entropy*, *Information Gain*, dengan memilih *Information Gain* terbesar dan dijadikan simpul internal pohon. Ulangi perhitungan tersebut hingga semua atribut pohon memiliki kelas.

Tabel 3. Tabel Hasil Perhitungan *Entropy* dan *Information Gain* dari semua atribut sebagai acuan pembuatan pohon keputusan dengan pemilihan atribut/variabel dengan *information gain* terbesar sebagai simpul akar.

Node	Variabel	Nilai	Jumlah Kasus	Diterima	Ditolak	Entropy	Information Gain
1	Total		107	93	14	0.559740	
	No. Tlp						0.217755
		Valid	100	93	7	0.365924	
		Tidak Valid	7	0	7	0	
	Alamat						0.086373
		Tervalidasi	104	93	11	0.487021	
		Tidak Tervalidasi	3	0	3	0	
	Status Tempat						0.026066
		Hak Milik	68	62	6	0.430552	
		Kontrak/Sewa	23	19	4	0.666578	
		Kredit/Leasing	14	11	3	0.749595	
		Pinjam	2	1	1	1	
	Tempo						0.040829
		30 Hari	8	6	2	0.811278	
		45 Hari	96	86	10	0.482066	
		60 Hari	3	1	2	0.918296	
	Cara Pembayaran						0.020547
		Tunai	7	5	2	0.863121	
		Transfer	98	87	11	0.506650	
		Giro	2	1	1	1	
	Omset						0.025879
		<=25 Juta	21	16	5	0.791858	
		<=50 Juta	56	49	7	0.543564	
		<=100 Juta	25	23	2	0.402179	
		>100 Juta	5	5	0	0	
	Target Pembelian						0.004903
		<=20 Pcs	86	74	12	0.583019	
		<=50 Pcs	13	12	1	0.391243	
		<=100 Pcs	7	6	1	0.591673	
		>100 Pcs	1	1	0	0	

#### 4.5 Pohon Keputusan (*Decision Tree*)

Dari hasil Perhitungan *entropy* dan *information gain* yang didapat kemudian diolah kedalam *Decision Tree*. Berikut hasil Pohon Keputusan (*Decision Tree*) :



Gambar 2. Pohon Keputusan(*Decision Tree*)

#### 4.6 Aturan-Aturan/*Rule Model*

Dari pohon keputusan(*decision tree*) yang terbentuk pada Gambar 2 diatas, didapat aturan—aturan/*rule model* dalam penentuan rekomendasi penerimaan mitra penjualan. Ada 14 aturan yang terbentuk, dapat dilihat sebagai berikut ;

1. *If* No.Tlp = Tidak Valid *Then* Rekomendasi : Ditolak
2. *If* No.Tlp = Valid *And* Alamat = Tervalidasi, Cara Pembayaran = Giro, Status Tempat = Hak Milik *Then* Rekomendasi = Diterima
3. *If* No.Tlp = Valid *And* Alamat = Tervalidasi, Cara Pembayaran = Giro, Status Tempat = Kredit/Leasing *Then* Rekomendasi = Ditolak.
4. *If* No.Tlp = Valid *And* Alamat = Tervalidasi, Cara Pembayaran = Transfer, Tempo = 30 Hari *Then* Rekomendasi = Diterima.
5. *If* No.Tlp = Valid *And* Alamat = Tervalidasi, Cara Pembayaran = Transfer, Tempo = 45 Hari, Status Tempat = Hak Milik *Then* Rekomendasi = Diterima.

6. If No.Tlp = Valid *And* Alamat = Tervalidasi, Cara Pembayaran = Transfer, Tempo = 45 Hari, Status Tempat = Kontrak/Sewa Then Rekomendasi = Diterima.
7. If No.Tlp = Valid *And* Alamat = Tervalidasi, Cara Pembayaran = Transfer, Tempo = 45 Hari, Status Tempat = Kredit/Leasing Then Rekomendasi = Diterima.
8. If No.Tlp = Valid *And* Alamat = Tervalidasi, Cara Pembayaran = Transfer, Tempo = 45 Hari, Status Tempat = Pinjam, Omset(Juta)/Bulan =  $\leq 25$  Juta Then Rekomendasi = Ditolak.
9. If No.Tlp = Valid *And* Alamat = Tervalidasi, Cara Pembayaran = Transfer, Tempo = 45 Hari, Status Tempat = Pinjam, Omset(Juta)/Bulan =  $\leq 50$  Juta Then Rekomendasi = Diterima.
10. If No.Tlp = Valid *And* Alamat = Tervalidasi, Cara Pembayaran = Transfer, Tempo = 60 Hari, Status Tempat = Kontrak/Sewa Then Rekomendasi = Diterima.
11. If No.Tlp = Valid *And* Alamat = Tervalidasi, Cara Pembayaran = Transfer, Tempo = 60 Hari, Status Tempat = Kredit/Leasing Then Rekomendasi = Ditolak.
12. If No.Tlp = Valid *And* Alamat = Tervalidasi, Cara Pembayaran = Tunai, Target Pembelian(Pcs)/Bulan =  $\leq 100$  Pcs Then Rekomendasi = Ditolak.
13. If No.Tlp = Valid *And* Alamat = Tervalidasi, Cara Pembayaran = Tunai, Target Pembelian(Pcs)/Bulan =  $\leq 20$  Pcs Then Rekomendasi = Diterima.
14. If No.Tlp = Valid *And* Alamat = Tidak Tervalidasi: Ditolak.

Pohon keputusan dan aturan-aturan yang dihasilkan dari penelitian ini berkaitan dengan hal-hal berikut, antara lain :

- a. Pada aturan 1, variabel No. Tlp dengan nilai Tidak Valid. Nilai dari variabel lain tidak akan berpengaruh dengan hasil rekomendasi. Begitu pula pada aturan 14, variabel Alamat dengan nilai Tidak Tervalidasi. Nilai dari variabel lain tidak akan berpengaruh dengan hasil rekomendasi.
- b. Untuk aturan-aturan yang tidak menampilkan seluruh variabel dengan nilai variabelnya berarti variabel yang tidak ditampilkan juga tidak mempengaruhi hasil rekomendasi, artinya apapun nilai variabel yang di masukkan kedalam aturan-aturan tersebut tidak akan mempengaruhi hasil rekomendasi.

#### 4.7 Validasi Dan Pengujian

Pengujian dilakukan dengan validasi silang. Salah satu jenis validasi silang adalah *ten-fold cross validation*. Berikut hasil dari *ten-fold cross validation* :

## PerformanceVector

```

PerformanceVector:
accuracy: 96.36% +/- 6.03% (mikro: 96.26%)
ConfusionMatrix:
True:  DITERIMA      DITOLAK
DITERIMA:    93       4
DITOLAK:     0       10
precision: 100.00% (positive class: DITOLAK)
ConfusionMatrix:
True:  DITERIMA      DITOLAK
DITERIMA:    93       4
DITOLAK:     0       10
recall: 71.43% (positive class: DITOLAK)
ConfusionMatrix:
True:  DITERIMA      DITOLAK
DITERIMA:    93       4
DITOLAK:     0       10
AUC (optimistic): 0.950 +/- 0.150 (mikro: 0.950) (positive class: DITOLAK)
AUC: 0.500 +/- 0.000 (mikro: 0.500) (positive class: DITOLAK)
AUC (pessimistic): 0.750 +/- 0.335 (mikro: 0.750) (positive class: DITOLAK)

```

Gambar 3. Hasil validasi Ten-fold cross validation

Berdasarkan pengujian dengan menggunakan metode *Ten-Fold Cross Validation* menghasilkan nilai akurasi yang hampir sempurna sebesar 96,26 %, presisi sebesar 100 % dan recall sebesar 71,43 %. Hal ini menunjukkan bahwa, dari proses klasifikasi yang dilakukan akan dapat diterapkan pada rekomendasi penerimaan calon mitra penjualan.

Row No.	REKOMEND...	prediction(R...	confidence(...	confidence(...
1	DITERIMA	DITERIMA	0.966	0.034
2	DITOLAK	DITOLAK	0	1
3	DITERIMA	DITERIMA	0.966	0.034
4	DITERIMA	DITERIMA	0.966	0.034
5	DITERIMA	DITERIMA	0.966	0.034
6	DITERIMA	DITERIMA	0.966	0.034
7	DITERIMA	DITERIMA	0.966	0.034
8	DITOLAK	DITERIMA	0.966	0.034
9	DITERIMA	DITERIMA	0.966	0.034
10	DITERIMA	DITERIMA	0.966	0.034

Name	Type	Missing	Statistics	Filter (11/11 attributes)
REKOMENDASI	Binominal	0	Least: DITOLAK (14) Most: DITERIMA (93)	
prediction(REKOMENDASI)	Binominal	0	Least: DITOLAK (10) Most: DITERIMA (97)	
Confidence_DITERIMA confidence(DITERIMA)	Real	0	Min: 0 Max: 0.977	
Confidence_DITOLAK confidence(DITOLAK)	Real	0	Min: 0.023 Max: 1	

Gambar 4. Hasil dari perhitungan variabel Rekomendasi

Perhitungan variabel rekomendasi *true* diterima dan juga diprediksi diterima adalah 93 *record*, rekomendasi *true* diterima tapi diprediksi ditolak adalah 0 *record*, rekomendasi *true* ditolak tetapi diprediksi diterima adalah  $14 - 10 = 4$  *record*, dan rekomendasi *true* ditolak dan juga diprediksi ditolak adalah 10 *record*.

#### 4.8 Analisis Hasil Pengujian

Analisis hasil pengujian dilakukan dengan melakukan perhitungan secara manual dengan *Confusion Matrix*. Perhitungan menggunakan model *confusion matrix*. Berikut ini merupakan hasil dari perhitungan *confusion matrix* pada algoritma C4.5:

Tabel 3. *Confusion Matrix*

	TRUE DITERIMA	TRUE DITOLAK
Pred. DITERIMA	93	4
Pred. DITOLAK	0	10

- Accuracy =  $\left(\left(\frac{93+10}{107}\right) * 100 \%\right) = 96,26 \%$
- Precision =  $\left(\left(\frac{10}{0+10}\right) * 100 \%\right) = 100 \%$
- Recall =  $\left(\left(\frac{10}{4+10}\right) * 100 \%\right) = 71,43 \%$

Dari Perhitungan diatas, dapat disimpulkan bahwa hasil dari perhitungan *accuracy*, *precision*, dan *recall* tersebut sama dengan hasil perhitungan yang ditampilkan pada Gambar 3. Berdasarkan pengujian dan analisa hasil pengujian yang dilakukan, dengan tingkat akurasi 96,26 %, presisi 100 %, dan *recall* 71,43 % menunjukkan nilai yang hampir seratus persen akurat, presisi yang sangat akurat dan recall yang masih dalam kategori baik menyimpulkan bahwa peneliti berhasil dalam mengimplementasikan algoritma klasifikasi C4.5 dengan baik dan akan membantu PT. Atria Artha Persada dalam merekomendasikan penerimaan calon mitra penjualan, apakah diterima atau tidak.

#### 5. KESIMPULAN

Kesimpulan dari penelitian yang dilakukan adalah sebagai berikut:

1. Penerapan algoritma klasifikasi C4.5 akan dapat diimplementasikan pada rekomendasi penerimaan mitra penjualan di PT. Atria Artha Persada, dilihat dari tingkat *accuracy* yang mencapai 96.26 % dan *recall* 71.43%., yang menyatakan bahwa perhitungan yang dilakukan akan mampu memprediksi dan me-rekomendasikan penerimaan mitra penjualan dengan baik.

2. Terbentuk 14 aturan-aturan/rule model dapat digunakan sebagai acuan dalam perancangan dan pembuatan aplikasi GUI.
3. Dengan algoritma ini bisa digunakan untuk menentukan calon mitra yang potensial berdasarkan atribut/variabel dengan nilai-nilai atribut yang dapat meyakinkan untuk diterima sebagai mitra penjualan.

## **6. STUDI LANJUTAN**

Studi lanjutan dari penelitian ini adalah mengembangkan aplikasi GUI untuk dapat menentukan rekomendasi penerimaan mitra penjualan yang sesuai dengan aturan/rule model yang terbentuk.

## REFERENCES

- [1] J. Purnomo, W. Laksito YS, and Y. Retno Wahyu U, "Implementasi Algoritma C 4.5 Dalam Pembuatan Aplikasi Penunjang Keputusan Penerimaan Pegawai Cv. Dinamika Ilmu," *J. Teknol. Inf. dan Komun.*, vol. 2, no. 1, 2014.
- [2] S. S. Nikam, "A Comparative Study of Classification Techniques in Data Mining Algorithms," *Orient. J. Comput. Sci. Technol.*, vol. 8, no. 1, pp. 13–19, 2015.
- [3] M. Aryuni and E. D. Madyatmadja, "Feature Selection in Credit Scoring Model for Credit Card Applicants in XYZ Bank : A Comparative Study," *Int. J. Multimed. Ubiquitous Eng.*, vol. 10, no. 5, pp. 17–24, 2015.
- [4] A. Mahajan and A. Ganpati, "Performance Evaluation of Rule Based Classification Algorithms," *Int. J. Adv. Res. Comput. Eng. Technol.*, vol. 3, no. 10, pp. 3546–3550, 2014.
- [5] A. Wijaya and A. S. Girsang, "The Use Of Data Mining For Prediction Of Customer Loyalty," *CommIT (Communication Inf. Technol.*, vol. 10, no. 1, pp. 41–47, 2016.
- [6] D. Wira and T. Putra, "Algoritma c4.5 untuk menentukan tingkat kelayakan motor bekas yang akan dijual," *J. TEKNOIF*, vol. 4, no. 1, 2016.
- [7] S. A. Lusinia, "Algoritma C4.5 Dalam Menganalisa Kelayakan Kredit(Studi Kasus Di Koperasi Pegawai Republik Indonesia (KP-RI) Lengayang Pesisir Selatan, Painan, Sumatera Barat)," *J. KomTekInfo Fak. Ilmu Komput.*, vol. 1, no. 2, pp. 6–10, 2014.
- [8] L. N. Rani, "Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit," *J. KomTekInfo Fak. Ilmu Komput.*, vol. 2, no. 2, pp. 33–38, 2015.
- [9] H. Widayu, S. Darma, N. Silalahi, and Mesran, "Data Mining untuk memprediksi jenis transaksi nasabah pada koperasi simpan pinjam dengan ...," *Media Inform. Budidarma*, vol. Vol 1, no. June, p. 7, 2017.
- [10] F. Ferdian and S. Hansun, "Penerapan Algoritma C4 . 5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE," *Jatisi*, vol. Vol. 3 No., 2017.
- [11] Y. S. Luvia, D. Hartama, A. P. Windarto, and Solikhun, "Penerapan algoritma c4.5 untuk klasifikasi predikat keberhasilan mahasiswa di amik tunas bangsa," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. Volume 1, no. April 2017, p. 6, 2016.
- [12] Kusri and E. T. Luthfi, *Algoritma Data Mining. Yogyakarta : Andi.* 2009.
- [13] G. Syahputra, "Penerapan Algoritma C4 . 5 Dalam Analisa Kelayakan Penerima Bonus Tahunan Pegawai ( Studi Kasus : PT . Multi Pratama Nauli Medan )," *J. Mantik Penusa*, vol. 16, no. 2, 2015.
- [14] K. Adhatrao, A. Gaykar, D. Amiraj, R. Jha, and Honroa Vipul, "Predicting Students' Performance Using Id3 And C4.5 Classification Algorithms," *Int. J. Data Min. Knowl. Manag. Process*, vol. 3, no. 5, pp. 39–52, 2013.
- [15] P. Premchand and M. ghous. Mohiuddin, "Performance Analysis of Decision Tree Classifiers," *Int. J. Comput. Sci. Trends Technol.*, vol. 2, no.

- 2, pp. 62–70, 2014.
- [16] N. Iriadi and N. Nuraeni, “Kajian Penerapan Metode Klasifikasi Data Mining Algoritma C4.5 Untuk Prediksi Kelayakan Kredit Pada Bank Mayapada Jakarta,” *J. Tek. Komput. AMIK BSI*, vol. 2, 2016.
- [17] Y. Altujjar, W. Altamimi, I. Al-turaiki, and M. Al-razgan, “Predicting Critical Courses Affecting Students Performance: A Case Study,” *Procedia - Procedia Comput. Sci.*, vol. 82, no. March, pp. 65–71, 2016.
- [18] Abdussomad and W. Gata, “Algoritma Klasifikasi C4.5 Berbasis Particle Swarm Optimization Untuk Prediksi Kelayakan Kredit,” *SNIPTEK*, pp. 398–405, 2014.