

## ANALISIS REKOMENDASI CALON DEBITUR MOTOR PADA PT.XYZ MENGUNAKAN ALGORITMA C4.5

Lilis Nurellisa<sup>\*1</sup>, Devi Fitriana<sup>2</sup>

<sup>1,2</sup>Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana  
Email: <sup>1</sup>41515010004@student.mercubuana.ac.id, <sup>2</sup>devi.fitriana@mercubuana.ac.id

<sup>\*</sup>Penulis Korespondensi

(Naskah masuk: 11 Juni 2019, diterima untuk diterbitkan: 7 Agustus 2019)

### Abstrak

PT.XYZ merupakan perusahaan jasa pembiayaan atau *leasing* dengan berkonsentrasi kepada pembiayaan sepeda motor. Dalam bisnisnya PT.XYZ sering sekali dihadapkan oleh masalah kredit macet atau bahkan penipuan. Hal ini dikarenakan kesalahan dalam pemberian kredit kepada calon debitur yang tidak potensial. Jika tidak ditangani hal ini tentu saja akan berdampak buruk bagi perusahaan. Perusahaan mengalami penurunan kemampuan dalam membayar angsuran pinjaman ke perbankan bahkan dapat berdampak pada kebangkrutan. Dalam hal ini PT.XYZ perlu melakukan analisis untuk menentukan calon debitur yang potensial dengan menggunakan data driven method atau pendekatan berbasis kepada data. Yaitu pengambilan keputusan dengan melihat data pengajuan kredit yang pernah ada sebelumnya yang disebut juga sebagai *supervised learning*. Algoritma yang digunakan adalah algoritma C4.5 karena algoritma ini dapat mengklasifikasi data yang sudah ada sebelumnya. Dengan algoritma ini akan dihasilkan sebuah pohon keputusan yang akan membantu PT.XYZ dalam pengambilan keputusan. Dengan pengujian menggunakan 3587 sampel data pengajuan kredit dalam kurun waktu 1 tahun akurasi yang didapatkan ialah 97,96%. Dengan begitu hal ini menunjukkan bahwa metode klasifikasi menggunakan algoritma C4.4 berhasil diimplementasikan dengan baik. Hal ini diharapkan dapat membantu PT.XYZ dalam merekomendasikan calon debitur yang potensial.

**Kata kunci:** data mining, metode klasifikasi, algoritma C4.5, decision tree, supervised learning

## ANALYSIS OF MOTORCYCLE DEBITOR RECOMMENDATIONS IN PT. XYZ USING C4.5 ALGORITHM.

### Abstract

PT. XYZ is a finance or leasing service company by concentrating on motorcycle financing. In its business, PT. XYZ is often faced with problems of bad credit or even fraud. This is due to an error in giving credit to potential debtors. If it is not handled this, of course, will have a bad impact on the company. Companies experiencing a decline in the ability to repay loan installments to banks can even have an impact on bankruptcy. In this case, PT. XYZ needs to do an analysis to determine potential debtors using data-driven methods or data-based approaches. That is decision making by looking at credit application data that has never been before, which is also called supervised learning. The algorithm used is the C4.5 algorithm because this algorithm can classify pre-existing data. With this algorithm, a decision tree will be produced that will help PT. XYZ in decision making. By testing using 3587 samples of credit filing data within a period of 1 year the accuracy obtained was 97.96%. That way this shows that the classification method using the C4.4 algorithm is successfully implemented properly. This is expected to help PT. XYZ in recommending potential debtors.

**Keywords:** data mining, classification method, C4.5 algorithm, decision tree, supervised learning

### 1. PENDAHULUAN

PT.XYZ merupakan perusahaan jasa pembiayaan atau *leasing* dengan berkonsentrasi pada pembiayaan sepeda motor. PT XYZ bukan perusahaan *leasing* baru di Indonesia karena telah diresmikan sejak tahun 1997.

Dalam bisnisnya, PT.XYZ sering kali dihadapkan pada masalah banyak calon debitur yang seharusnya tidak layak kriteria untuk mendapatkan kredit dapat dengan mudah diterima pengajuan kreditnya sehingga dampak dari hal tersebut banyak terjadi adanya kredit macet. Kredit macet adalah suatu masalah dimana debitur tidak

dapat memenuhi kewajibannya untuk membayar angsuran akibatnya kredit tersebut menunggak. Atau keadaan di mana seseorang tidak dapat membayar sebagian atau seluruh kewajibannya kepada sebuah perusahaan pembiayaan atau leasing yang telah disepakati sebelumnya. (Isabella, Raisa Pratiw, 2013). Dampak dari adanya kredit macet bagi perusahaan adalah pendapatan perusahaan menurun. Perusahaan juga akan kesulitan dalam membayar angsuran pinjaman ke bank atau bahkan dapat terjadinya kebangkrutan pada perusahaan (Latif Adam, 2009).

Sejauh ini PT. XYZ memiliki prosedur dalam menentukan kelayakan calon debitur, yaitu dengan mendata calon debitur. Data tersebut kemudian dipelajari sebagai acuan untuk menentukan keputusan diterima atau tidaknya calon debitur tersebut. Selama ini penentuan calon debitur potensial hanya mengacu pada data pengajuan kredit saja walaupun dapat dikatakan data tersebut lengkap. Hal ini terjadi dikarenakan hanya menentukan calon debitur berdasarkan data pengajuan baru dan tidak melakukan analisis terhadap data calon debitur yang pernah ada sebelumnya yang bisa saja pernah bermasalah pada kredit sebelumnya.

Berdasarkan masalah tersebut, perlu dilakukan analisis untuk menentukan calon debitur yang potensial dengan menggunakan data *driven method* atau pendekatan berbasis kepada data, yaitu pengambilan keputusan dengan melihat data pengajuan kredit yang pernah ada sebelumnya *supervised learning*. Algoritma yang dapat digunakan untuk mengklasifikasi data yang sudah ada sebelumnya adalah algoritma C4.5.

Algoritma C4.5 telah banyak digunakan dalam penentuan keputusan karena menghasilkan sebuah rule dari hasil *decision tree* yang mudah diinterpretasikan, dapat mengolah data numerik dan diskret, dapat menangani data yang kosong atau *missing value*, dan performanya tercepat dibandingkan dengan algoritma lain.

Algoritma C4.5 adalah algoritma klasifikasi dalam pengolahan data dan *machine learning*. Dalam algoritma C4.5 dilakukan proses pengklasifikasian data dalam bentuk tingkatan mulai dari akar ke daun atau yang disebut juga dengan *decision tree*. Proses pada pohon keputusan adalah mengubah bentuk data tabel menjadi model pohon, dari pohon tersebut menghasilkan sebuah *rule*. Dari rule ini nantinya dapat diimplementasikan untuk menentukan kelayakan penerimaan calon debitur (Arifin & Fitriana, 2018).

## 2. STUDI LITERATUR

### 2.1. Klasifikasi

Klasifikasi adalah salah satu teknik prediksi dengan memetakan fakta ke dalam kelompok tipe

yang telah ditentukan. Klasifikasi juga dikenal sebagai supervised learning atau pembelajaran pada data sebelumnya dengan kelas yang telah ditentukan sebelumnya. (Paramita & Ely, 2015).

Komponen utama proses klasifikasi yaitu : (Sugianto, 2015) :

- Kelas : merupakan label dari hasil klasifikasi.
- Prediktor : aribut data yang akan diklasifikasi
- Data Training* : merupakan sekumpulan data yang berisi kelas dan predictor
- Data Testing : berisi data baru yang akan dikelompokkan untuk mengetahui akurasi dari model yang telah dibuat.

### 2.2. Algoritma C4.5

Algoritma ini adalah penerus dari algoritma sebelumnya yaitu algoritma ID3 yang digunakan untuk menghasilkan sebuah *rule* berupa pohon keputusan. Algoritma ini dapat digunakan untuk klasifikasi data dan disebut sebagai classifier statistik (Kalpesh Adhatrao, Aditya Gaykar, Amiraj Dhawan, 2013). Algoritma C4.5 merupakan algoritma klasifikasi atau pengelompokan yang bersifat prediktif. Pernyataan klasifikasi pada pohon keputusan (*decision tree*) terdapat pada cabang-cabangnya dan kelas-kelas atau segmen-segmennya terdapat pada daun-daunnya (Firmansyah & Rusito, 2016). Algoritma C4.5 adalah salah satu keluarga pohon keputusan yang dapat menghasilkan sebuah pohon keputusan dan aturan atau rule untuk tujuan meningkatkan prediksi Akurasi. Selain itu, model C4.5 mudah digunakan dan dipahami sebagai aturan yang diturunkan dan memiliki teknik interpretasi yang sangat mudah (Jantan, Hamdan, & Othman, 2011).

Berikut merupakan tahap pembuatan decision tree dengan Algoritma C4.5 (Budi Utami, 2016) :

- Menyiapkan *data training*. *Data training* yaitu data yang pernah terjadi sebelumnya yang telah dikelompokkan ke dalam kelas tertentu. Pada penelitian ini data pengajuan kredit motor pada satu tahun terakhir dijadikan sebagai *data training*.
- Menghitung nilai *entropy*. Perhitungan nilai *entropy* dijadikan sebagai awalan karena hasil perhitungan ini akan dijadikan acuan untuk perhitungan selanjutnya. Nilai entropy didapat dari jumlah data diterima dan ditolak pada *data training*. Semakin tinggi *entropy*, semakin tinggi pula potensi untuk meningkatkan proses klasifikasi (Kalpesh Adhatrao, Aditya Gaykar, Amiraj Dhawan, 2013). Untuk dapat menghitung nilai *entropy* dapat menggunakan rumus:

$$Entropy (S) = \sum_{i=1}^n - p_i \cdot \log_2 p_i \quad (1)$$

- c. Menghitung nilai *Information gain*. Nilai *gain* digunakan untuk memilih atribut tertentu untuk menjadi simpul pohon keputusan dalam algoritma C4.5 (Priyanka Saini, Sweta Rain, 2014). Hasil perhitungan nilai *Information Gain* tertinggi akan dijadikan akar simpul pohon keputusan atau *root node*. Hasil perhitungan *Information gain* juga dapat menentukan keseluruhan dari pohon keputusan untuk menentukan akar sampai kedaun. Untuk mencari nilai *gain* dapat menggunakan rumus:

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

- d. Ulangi langkah ke-2 hingga semua record terpartisi.  
 e. Proses partisi pohon keputusan akan berhenti saat (Putra & Chan, 2018):
- Semua record dalam simpul N mendapat kelas yang sama.
  - Tidak ada atribut di dalam record yang dipartisi lagi.
  - Tidak ada record di dalam cabang yang kosong.

### 2.3. Decision Tree

Pohon keputusan (*choice tree*) yaitu metode klasifikasi yang merepresentasikan kedalam struktur pohon. Konsep dari pohon keputusan ini adalah dengan mengumpulkan informasi, selanjutnya dibuatkan percabangan dari akar ke daun yang menghasilkan rule atau aturan untuk solusi permasalahan. (Elisa, 2017). Pohon keputusan adalah sebuah struktur pohon seperti *flowchart*, di mana setiap node internal (*non leaf node*) mewakili pada satu atribut, dan setiap simpul daun (atau terminal node) memegang label kelas. *Node* paling atas dalam sebuah pohon keputusan disebut dengan simpul akar (Gaurav L. Agrawal, 2013).

Pada *decision tree* terdapat 3 jenis *node*, yaitu (Sholihah, Satria, & Muslihudin, 2018) (Hadi, 2017) (Cahyono, 2010):

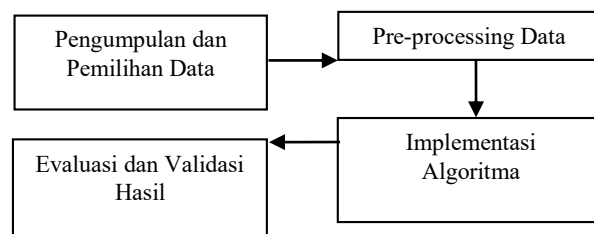
- Root Node* merupakan *node* paling atas atau disebut juga akar simpul pohon, pada *node* ini tidak ada input dan mempunyai *output* lebih dari satu.
- Internal Node* merupakan *node* percabangan, pada *node* ini hanya terdapat satu input dan mempunyai *output* minimal dua.
- Leaf Node* atau terminal *node* merupakan *node* akhir, pada *node* ini hanya terdapat satu input dan tidak mempunyai *output*.

### 2.4. Confusion Matrix

*Confusion Matrix* digunakan sebagai model penilaian untuk memprediksi yang benar atau salah. Biasanya digunakan untuk perhitungan akurasi pada sistem pendukung keputusan. Pada perhitungan kinerja

menggunakan *confusion matrix* akan terdapat 4 istilah antara lain : *True Positive*, *True Negative*, *False Negative* dan *False Positive* (Firmansyah & Rusito, 2016).

### 3. METODE PENELITIAN



Gambar 1. Block Diagram Penelitian

#### 3.1. Pengumpulan dan Pemilihan Data

Dataset yang digunakan dalam penelitian ini adalah data pengajuan kredit motor pada PT.XYZ dalam kurun waktu 1 tahun. Total data yang digunakan adalah 3587 data pengajuan kredit. Contoh dataset yang digunakan untuk penelitian ini adalah sebagai berikut :

Tabel 1. Dataset

No	Atribut	Deskripsi
1	Nama	Sample 1 Sample 2
2	No.Telp	Aktif Tidak Aktif
3	Status BI	Bermasalah Tidak Bermasalah
4	Penghasilan	Rendah Menengah Tinggi
5	Status Tinggal	Kontrak Bukan Milik Pribadi Milik Pribadi
6	Status Diri	Lajang Cera Menikah
7	Status Penghasilan Pasangan	Tidak Ada Tidak Bekerja Bekerja
8	Tanggungan	1 2 3 4 5
9	Tipe Order	MIO S MIO FINO MIO CW MIO SOUL XRIDE NMAX AEROX LEXI
10	Tenor	11 17 23 29 35
11	Keterangan	Diterima Ditolak

Dalam penelitian ini *pre-processing* yang dilakukan antara lain :

a. *Data Cleaning*

Pada tahap ini dilakukan pembersihan pada data pengajuan kredit kendaraan bermotor yakni memeriksa adanya duplikat data pada data 1 tahun terakhir, menghilangkan data yang kosong, dan menghilangkan *typografi*. Lalu setelah itu semua atribut kemudian akan dipilih untuk mendapatkan atribut yang mengandung nilai yang paling relevan (Sadikin & Alfiandi, 2018).

b. *Data Transformation*

Pada tahap ini data dikonversi ke dalam bentuk yang sesuai untuk diproses dalam data mining. Proses mengubah data dari satu bentuk ke bentuk lainnya disebut juga transformasi data. Dalam penelitian ini, data ditransformasikan dari nilai atribut numerik ke nilai atribut kategorial (Kalpesh Adhatrao, Aditya Gaykar, Amiraj Dhawan, 2013). Dalam penelitian ini atribut gaji yang berupa nominal akan dikonversi berupa *range* “rendah, menengah dan tinggi” berdasarkan arahan dari manajemen PT.XYZ. Hal ini dilakukan untuk membuat data lebih mudah diproses. Contoh proses transformasi data pada table 2

### 3.2. Implementasi Algoritma

Implementasi algoritma dilakukan dengan menghitung nilai *entropy* dan *information gain* menggunakan rumus. Berikut Merupakan contoh perhitungan salah satu atributnya.

$$\begin{aligned} \text{Entropy(Total)} &= \left( -\frac{3372}{3587} * \log_2 \left( \frac{3372}{3587} \right) \right) \\ &+ \left( -\frac{215}{3587} * \log_2 \left( \frac{215}{3587} \right) \right) \\ &= 0.327201404 \\ \text{Gain(Total, No. Telp)} &= 0.327201404 - \\ &\left[ \left( \frac{3372}{3587} * 0.316675157 \right) + \left( \frac{10}{3587} * 0 \right) \right] \\ &= 0.642993719 \end{aligned}$$

Tabel 2. Transformasi Data

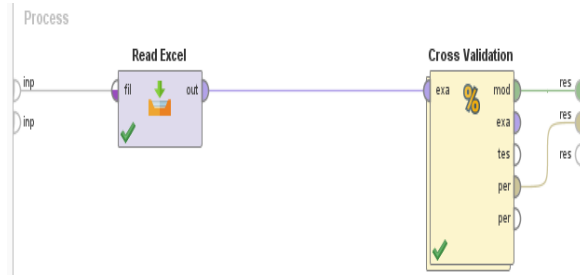
Value	Transformasi
1.000.000 sd < 3.000.000	Rendah
≥ 3.000.000 sd < 5.000.000	Menengah
≥ 5.000.000 sd seterusnya	Tinggi

Kemudian, setelah semua atribut dihitung nilai *entropy* dan *information gain*nya maka dari hasil perhitungan tersebut yang memperoleh nilai *information gain* tertinggi akan dijadikan simpul pohon keputusan. Dan hasil perhitungan tertinggi kedua akan dijadikan akar

pohon kedua setelah simpul pohon keputusan dan seterusnya. Hasil perhitungan nilai *entropy* dan *gain* secara keseluruhan pada tabel 3.

### 3.3. Implementasi Model

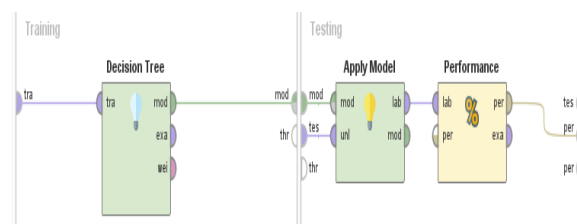
Setelah pengumpulan dan pemrosesan data, model diuji menggunakan tools Rapidminer sebagai berikut :



Gambar 2. Operator Pengujian

Berdasarkan Gambar 2. Operator Pengujian dapat dilihat operator *read excel* yaitu berisikan sampel data diuji menggunakan operator *cross validation* yaitu operator yang berfungsi membagi *data training* dan *data testing* dalam beberapa bagian. Didalam operator *cross validation*, akan ada operator lain yaitu salah satunya operator *Decision Tree* yang mana output yang dihasilkan dari pengujian menggunakan *cross validation* ini akan menghasilkan berupa *rule* dan pohon keputusan. Berikut penjelasan masing-masing operator :

- Operator *Read Excel*: Operator ini digunakan untuk mengimpor data yang akan digunakan untuk pengujian dari *Microsoft Excel* ke *RapidMiner*.
- Operator *Cross Validation* : Operator ini digunakan untuk pengujian dengan cara membagi data training dan data testing sebanyak k bagian. Pada pengujian ini dilakukan sebanyak k kali dengan mengolah data yang terpartisi sebagai data training dan data test (Jacobus & Winarko, 2014). Pada penelitian ini data training dibagi sebanyak 10 bagian berukuran sama. Untuk masing-masing dari 10 subset data tersebut, *Cross Validation* akan menggunakan 9 *fold* untuk pelatihan dan 1 *fold* untuk pengujian (Wibowo, 2017).



Gambar 3. Operator Pada Cross Validation

- Operator *Decision Tree* : Dalam operator ini, hasil dari perhitungan entropy dan gain akan menetapkan nilai *criterion, maximal depth, confidence, minimal gain, minimal leaf size, minimal size for split* dan *number of prepruning alternatives*.
- Operator *Apply Model* : Operator ini digunakan untuk membaca data yang akan diestimasi berdasarkan data training yang telah dipelajari.
- Operator *Performance* : Operator ini secara otomatis menentukan jenis tugas pembelajaran

Tabel 3. Hasil Perhitungan Entropy dan Information Gain

Atribut	Keterangan	Total	Diterima	Ditolak	Entropy	Information Gain
Total		3587	3372	215	0.327201404	
No.Telp						0.642993719
	Tidak Aktif	10	0	10	0	
	Aktif	3577	3372	205	0.316675157	
Status BI						0.646897226
	Bermasalah	15	0	15	0	
	Tidak Bermasalah	3572	3372	213	0.32103833	
Penghasilan						0.581056844
	Rendah	12	2	10	0.650022422	
	Menengah	3354	3204	150	0.263541624	
	Tinggi	221	216	5	0.155932691	
Status Tinggal						0.631397964
	Milik Pribadi	1461	1433	28	0.13672599	
	Kontrak	1402	958	444	0.90074606	
	Bukan Milik Pribadi	724	721	3	0.038762094	
Status Diri						0.271102841
	Menikah	2346	2226	120	0.291265906	
	Lajang	1029	956	73	0.369431678	
	Cerai	212	190	22	0.480844694	
Status Pekerjaan Pasangan						0.618389327
	Bekerja	1178	1172	6	0.046126533	
	Tidak Bekerja	1168	1029	139	0.526499331	
	Tidak Ada	1241	1146	95	0.389908878	
Tanggungans						0.620397825
	0	1071	1042	29	0.179516927	
	1	1202	1028	174	0.5965671	
	2	517	501	16	0.199122862	
	3	498	457	41	0.410333971	
	4	251	177	74	0.874867333	
	5	47	47	0	0	
	6	1	0	1	0	
Tipe Order						0.480070245
	AEROX	434	389	45	0.480574187	
	LEXI	721	702	19	0.175755128	
	MIO CW	330	325	5	0.113274303	
	MIO FINO	372	365	7	0.134746975	
	MIO S	193	168	25	0.556157909	
	MIO SOUL	419	401	18	0.255700278	
	NMAX	655	625	30	0.268287043	
	XRIDE	463	435	28	0.329328254	
Tenor						0.485129289
	11	652	601	24	0.283668814	
	17	789	747	42	0.29997051	
	23	1043	1003	40	0.234679404	
	29	548	530	18	0.208472993	
	35	555	528	27	0.280628542	

### 3.4 Evaluasi dan Validasi Hasil

Setelah mendapatkan model yang dihasilkan, berikutnya adalah melakukan evaluasi dan validasi dari hasil tersebut. Validasi dilakukan dengan menggunakan metode *Ten-Fold Cross Validation*. Berdasarkan hasil pengujian dapat diketahui semua fungsi bekerja dengan baik. Pengujian menggunakan *Ten-fold Cross Validation* dilakukan dengan cara membagi data set menjadi 10 bagian yang berukuran sama dengan cara melakukan pengacakan data.

Validasi dan pengujian bertujuan untuk mendapatkan tingkat akurasi, presisi, dan *recall* dari hasil prediksi. (Altujjar, Altamimi, Al-Turaiki, & Al-Razgan, 2016).

## 4. HASIL DAN PEMBAHASAN

Berdasarkan hasil perhitungan *entropy* dan *information gain* yang dapat dilihat pada Tabel Tabel 3. Hasil Perhitungan *Entropy* dan *Information Gain* atribut Status BI dipilih sebagai simpul akar pohon keputusan karena mendapatkan nilai *information gain*

tertinggi yaitu 0.646897226. Setelah itu dilanjutkan memilih node hingga semua semua atribut terpartisi. Jika semua atribut sudah memiliki kelas, maka pohon keputusan akan dibentuk seperti yang ditunjukkan pada Gambar 4. *Decision Tree*

Dari pohon keputusan yang terbentuk pada Gambar 5. Deskripsi Hasil *Decision Tree* maka didapat aturan

atau *rule model* dalam penentuan rekomendasi calon debitur motor antara lain :

1. *If* Status BI = Bermasalah *Then* Rekomendasi : Ditolak.
2. *If* Status BI = Tidak Bermasalah and No.Telp := Tidak Aktif *Then* Rekomendasi : Ditolak.



Gambar 4. Decision Tree

### Tree

```

STATUS BI = Bermasalah: Ditolak (Diterima=0, Ditolak=15)
STATUS BI = Tidak Bermasalah
| NO TELP = Aktif
| | STATUS TINGGAL = Bukan Milik Pribadi: Diterima (Diterima=721, Ditolak=3)
| | STATUS TINGGAL = Kontrak
| | | TANGGUNGAN > 3.500
| | | | STATUS PEKERJAAN PASANGAN = Bekerja: Diterima (Diterima=19, Ditolak=0)
| | | | STATUS PEKERJAAN PASANGAN = Tidak Ada: Ditolak (Diterima=0, Ditolak=20)
| | | | STATUS PEKERJAAN PASANGAN = Tidak Bekerja: Ditolak (Diterima=0, Ditolak=46)
| | | TANGGUNGAN ≤ 3.500
| | | | PENGHASILAN = Menengah
| | | | | TANGGUNGAN > 2.500
| | | | | | STATUS PEKERJAAN PASANGAN = Bekerja: Diterima (Diterima=28, Ditolak=0)
| | | | | | STATUS PEKERJAAN PASANGAN = Tidak Ada: Diterima (Diterima=10, Ditolak=1)
| | | | | | STATUS PEKERJAAN PASANGAN = Tidak Bekerja
| | | | | | | TIPE ORDER = LEXI: Diterima (Diterima=4, Ditolak=0)
| | | | | | | TIPE ORDER = MIO S: Ditolak (Diterima=0, Ditolak=8)
| | | | | | | TIPE ORDER = MIO SOUL: Diterima (Diterima=5, Ditolak=0)
| | | | | | | TIPE ORDER = NMAX: Ditolak (Diterima=0, Ditolak=23)
| | | | | | | TIPE ORDER = XRIDE: Diterima (Diterima=8, Ditolak=0)
| | | | | TANGGUNGAN ≤ 2.500
| | | | | TENOR > 14: Diterima (Diterima=936, Ditolak=26)
| | | | | TENOR ≤ 14
| | | | | | STATUS PEKERJAAN PASANGAN = Bekerja: Diterima (Diterima=110, Ditolak=0)
| | | | | | STATUS PEKERJAAN PASANGAN = Tidak Ada
| | | | | | | TIPE ORDER = AEROX: Ditolak (Diterima=2, Ditolak=3)
| | | | | | | TIPE ORDER = LEXI: Diterima (Diterima=35, Ditolak=11)
| | | | | | | TIPE ORDER = MIO CW: Diterima (Diterima=6, Ditolak=5)
| | | | | | | TIPE ORDER = MIO FINC: Diterima (Diterima=1, Ditolak=1)
| | | | | | | TIPE ORDER = MIO SOUL: Ditolak (Diterima=2, Ditolak=13)
| | | | | | | TIPE ORDER = XRIDE: Ditolak (Diterima=0, Ditolak=7)
| | | | | | STATUS PEKERJAAN PASANGAN = Tidak Bekerja: Ditolak (Diterima=0, Ditolak=13)
| | | | | PENGHASILAN = Rendah
| | | | | | STATUS PEKERJAAN PASANGAN = Bekerja: Diterima (Diterima=1, Ditolak=1)
| | | | | | STATUS PEKERJAAN PASANGAN = Tidak Ada: Ditolak (Diterima=0, Ditolak=6)
| | | | | | STATUS PEKERJAAN PASANGAN = Tidak Bekerja: Ditolak (Diterima=0, Ditolak=3)
| | | | | PENGHASILAN = Tinggi: Diterima (Diterima=47, Ditolak=0)
| | STATUS TINGGAL = Milik Pribadi: Diterima (Diterima=1437, Ditolak=0)
| NO TELP = Tidak Aktif: Ditolak (Diterima=0, Ditolak=10)

```

Gambar 5. Deskripsi Hasil Decision Tree

3. *If* Status BI = Tidak Bermasalah, No.Telp =Aktif, Status Tinggal = Milik Pribadi / Bukan Milik Pribadi *Then* Rekomendasi : Diterima.
4. *If* Status BI = Tidak Bermasalah, No.Telp =Aktif, Status Tinggal = Kontrak, Tanggungan = >3.500, Status Pekerjaan Pasangan = Tidak Ada / Tidak Bekerja *Then* Rekomendasi : Ditolak.

5. *If* Status BI = Tidak Bermasalah, No.Telp =Aktif, Status Tinggal = Kontrak, Tanggungan = >3.500, Status Pekerjaan Pasangan = Bekerja *Then* Rekomendasi : Diterima
6. *If* Status BI = Tidak Bermasalah, No.Telp =Aktif, Status Tinggal = Kontrak, Tanggungan = ≤ 3.500, Penghasilan = Rendah, Status Pekerjaan Pasangan = Tidak Bekerja / Tidak Ada *Then* Rekomendasi : Ditolak.
7. *If* Status BI = Tidak Bermasalah, No.Telp =Aktif, Status Tinggal = Kontrak, Tanggungan = ≤ 3.500, Penghasilan = Rendah, Status Pekerjaan Pasangan = Bekerja *Then* Rekomendasi : Diterima.
8. *If* Status BI = Tidak Bermasalah, No.Telp =Aktif, Status Tinggal = Kontrak, Tanggungan = ≤ 3.500, Penghasilan = Tinggi *Then* Rekomendasi : Diterima
9. *If* Status BI = Tidak Bermasalah, No.Telp =Aktif, Status Tinggal = Kontrak, Penghasilan = Menengah, Tanggungan = > 2.500, Status Pekerjaan Pasangan = Tidak Bekerja, Tipe Order = MIO S/ NMAX *Then* Rekomendasi : Ditolak.
10. *If* Status BI = Tidak Bermasalah, No.Telp =Aktif, Status Tinggal = Kontrak, Penghasilan = Menengah, Tanggungan = > 2.500, Status Pekerjaan Pasangan = Bekerja/ Tidak ada *Then* Rekomendasi : Diterima
11. *If* Status BI = Tidak Bermasalah, No.Telp =Aktif, Status Tinggal = Kontrak, Penghasilan = Menengah, Tanggungan = > 2.500, Status Pekerjaan Pasangan = Tidak Bekerja, Tipe Order = LEXI, MIO SOUL, XRIDE *Then* Rekomendasi : Diterima
12. *If* Status BI = Tidak Bermasalah, No.Telp =Aktif, Status Tinggal = Kontrak, Penghasilan = Menengah, Tanggungan = < 2.500, , Tenor = ≤ 14, Status Pekerjaan Pasangan = Tidak Bekerja *Then* Rekomendasi : Ditolak.
13. *If* Status BI = Tidak Bermasalah, No.Telp =Aktif, Status Tinggal = Kontrak, Penghasilan = Menengah, Tanggungan = < 2.500, Status Pekerjaan Pasangan = Tidak Ada, Tipe Order = XRIDE/ MIO SOUL *Then* Rekomendasi : Ditolak.
14. *If* Status BI = Tidak Bermasalah, No.Telp =Aktif, Status Tinggal = Kontrak, Penghasilan = Menengah, Tanggungan = < 2.500, Tenor = ≤ 14, Status Pekerjaan Pasangan = Tidak Ada, Tipe Order = XRIDE/ MIO SOUL *Then* Rekomendasi : Ditolak.
15. *If* Status BI = Tidak Bermasalah, No.Telp =Aktif, Status Tinggal = Kontrak, Penghasilan = Menengah, Tanggungan = < 2.500, Tenor = ≤ 14, Status Pekerjaan Pasangan = Bekerja *Then* Rekomendasi : Diterima
16. *If* Status BI = Tidak Bermasalah, No.Telp =Aktif, Status Tinggal = Kontrak, Penghasilan = Menengah, Tanggungan = < 2.500, Tenor = ≤ 14, Status Pekerjaan Pasangan = Tidak Ada, Tipe Order = MIO CW, MIO FINO, LEXI *Then* Rekomendasi : Diterima

### PerformanceVector

```

PerformanceVector:
accuracy: 97.96% +/- 0.55% (micro average: 97.96%)
ConfusionMatrix:
True:   Diterima   Ditolak
Diterima:   3371    72
Ditolak:     1    143
precision: 99.50% +/- 1.50% (micro average: 99.31%) (positive class: Ditolak)
ConfusionMatrix:
True:   Diterima   Ditolak
Diterima:   3371    72
Ditolak:     1    143
recall: 66.37% +/- 10.24% (micro average: 66.51%) (positive class: Ditolak)
ConfusionMatrix:
True:   Diterima   Ditolak
Diterima:   3371    72
Ditolak:     1    143
AUC (optimistic): 0.991 +/- 0.012 (micro average: 0.991) (positive class: Ditolak)
AUC: 0.949 +/- 0.027 (micro average: 0.949) (positive class: Ditolak)
AUC (pessimistic): 0.907 +/- 0.049 (micro average: 0.907) (positive class: Ditolak)

```

Gambar 6. Deskripsi hasil Performance Vector

Berdasarkan hasil pengujian menggunakan metode *Ten-Fold Cross Validation* pada Gambar 5. Deskripsi hasil *Performance Vector* pengujian ini menghasilkan

nilai akurasi sebesar 97,96 %, presisi sebesar 99.50% dan *recall* sebesar

66,37%. Evaluasi hasil tes dilakukan secara manual dengan menghitung menggunakan *Confusion Matrix*. Hasil perhitungan menggunakan *Confusion Matrix* dalam algoritma C4.5 dapat dilihat pada Tabel 4. Prediksi :

Tabel 4. Prediksi

	True Diterima	true Ditolak
pred. Diterima	3371	72
pred. Ditolak	1	143

$$\text{Accuracy} = \left( \left( \frac{3371 + 143}{3587} \right) * 100\% = 97,96\% \right)$$

$$\text{Precision} = \left( \left( \frac{143}{1 + 143} \right) * 100\% = 99,50\% \right)$$

$$\text{Recal} = \left( \left( \frac{143}{72 + 143} \right) * 100\% = 66,37\% \right)$$

Berdasarkan hasil perhitungan manual menggunakan rumus, dapat disimpulkan bahwa hasil dari perhitungan *accuracy*, *precision* dan *recall* mendapatkan hasil yang sama dengan hasil perhitungan menggunakan tools Rapidminer yang ada pada Gambar 5. Deskripsi hasil *Performance Vector*.

## 5. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan menggunakan 3587 sampel data, algoritma C4.5 dapat digunakan untuk menentukan calon debitur yang potensial. Data yang didapatkan terlebih dahulu dilakukan pra-pemrosesan yang antara lain membuang adanya duplikat data, tipografi dan melakukan transformasi data. Dan dari penelitian ini akan dihasilkan 2 klasifikasi yaitu diterima dan ditolak. Hasil perhitungan *confusion matrix* dan 10 *Ten-Fold Cross Validation* dihasilkan *Accuracy* yang baik sebesar 97,96%, *Precision* 99,50% dan *Recall* 66,37% dan menghasilkan 16 rule model yang dapat membantu PT.XYZ dalam pengambilan keputusan. Dapat disimpulkan bahwa Algoritma C4.5 dapat diimplementasikan sebagai rekomendasi kelayakan calon debitur yang potensial. Diharapkan dengan adanya penelitian ini dapat menjadi studi lanjutan untuk pembuatan aplikasi, atau bisa juga mengujinya menggunakan algoritma lain.

## DAFTAR PUSTAKA

- ALTUJJAR, Y., ALTAMIMI, W., AL-TURAIKI, I., & AL-RAZGAN, M. (2016). Predicting Critical Courses Affecting Students Performance: A Case Study. *Procedia Computer Science*, 82(March), 65–71. <https://doi.org/10.1016/j.procs.2016.04.010>
- ARIFIN, M. F., & FITRIANAH, D. (2018). Penerapan Algoritma Klasifikasi C4.5 Dalam

- Rekomendasi Penerimaan Mitra Penjualan Studi Kasus: PT Atria Artha Persada. *InComTech*, 8(2), 87–102. <https://doi.org/10.22441/incomtech.v8i1.2198>
- BUDI UTAMI, Y. R. (2016). Klasifikasi Penentuan Tim Utama Olahraga Hockey Menggunakan Algoritma C4.5. *Techno.COM*, 15(4), 364–368.
- CAHYONO, A. B. (2010). Analisis Pemanfaatan Small Disjunct Pada Decision Tree Dengan Algoritma Genetika. *Seminar Nasional Aplikasi Teknologi Informasi*, 6–10.
- ELISA, E. (2017). Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti. *Jurnal Online Informatika*, 2(1), 36. <https://doi.org/10.15575/join.v2i1.71>
- FIRMANSYAH, M. T., & RUSITO. (2016). Implementasi Metode Decision Tree Dan Algoritma C4.5 Untuk Klasifikasi Data Nasabah Bank. *Infokam*, XII(1), 1–12.
- GAURAV L. AGRAWAL, H. G. (2013). Optimization of C4.5 Decision Tree Algorithm for Data Mining Application. *International Journal of Emerging Technology and Advanced Engineering*, 3(3), 341–345. <https://doi.org/10.1109/ITAPP.2010.5566133>
- HADI, F. (2017). Penerapan Data Mining Dalam Menganalisa Pemberian Pinjaman Dengan Menggunakan Metode Algoritma C5.0 ( Studi Kasus: Koperasi Jasa Keuangan Syariah Kelurahan Lambung Bukik ). *Jurnal KomTekInfo*, 4(2), 214–223.
- ISABELLA, RAISA PRATIWI, F. A. (2013). Pengaruh Pengendalian Internal Terhadap Kredit Macet (Studi Kasus Pada Finance di Kota Palembang). *Jurusan Akuntansi STIE Multi Data Palembang*, 2013, (x), 1–13.
- JACOBUS, A., & WINARKO, E. (2014). Penerapan Metode Support Vector Machine pada Sistem Deteksi Intrusi secara Real-time. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 8(1), 13–24. <https://doi.org/10.22146/ijccs.3491>
- JANTAN, H. R., HAMDAN, A. A., & OTHMAN, Z. (2011). Human Talent Prediction in HRM using C4.5 Classification Algorithm. (*IJCSE*) *International Journal on Computer Science and Engineering*, 1(4), 2526–2534. <https://doi.org/10.4018/jtd.2010100103>
- KALPESH ADHATRAO, ADITYA GAYKAR, AMIRAJ DHAWAN, R. J. AND V. H. (2013). PREDICTING STUDENTS' PERFORMANCE USING ID3 AND C4.5 CLASSIFICATION ALGORITHMS Kalpesh. *International Journal of Data Mining & Knowledge Management Process (IJDMP)*, 3(5), 39–52. <https://doi.org/10.5121/ijdkp.2013.3504>
- LATIF ADAM. (2009). Kredit Bermasalah, Penyebab dan Dampaknya. Retrieved March 10, 2019,



- from <http://lipi.go.id/berita/kredit-bermasalah-penyebab-dan-dampaknya-/3997>
- PARAMITA, M., & ELY, R. (2015). Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining. *Sistem Informasi*, 11(November), 1–7.
- PRIYANKA SAINI, SWETA RAIN, A. K. J. (2014). Decision Tree Algorithm Implementation Using Educational Data. *International Journal of Computer-Aided Technologies (IJCAx)*, 1(1), 31–41.
- PUTRA, P. P., & CHAN, A. S. (2018). Pengembangan Aplikasi Perhitungan Prediksi Stock Motor Menggunakan Algoritma C 4 . 5 Sebagai Bagian dari Sistem Pengambilan Keputusan ( Studi Kasus di Saudara Motor ). *JURNAL INOVTEK POLBENG*, 3(1), 24–33.
- SADIKIN, M., & ALFIANDI, F. (2018). Comparative Study of Classification Method on Customer Candidate Data to Predict its Potential Risk. *International Journal of Electrical and Computer Engineering*, 8(6), 4763–4771.  
<https://doi.org/10.11591/ijece.v8i6.pp.4763-4771>
- SHOLIHAN, H., SATRIA, F., & MUSLIHUDIN, M. (2018). Implementasi Algoritma C4.5 Klasifikasi Nasabah Potensial ADIRA Dinamika Multi Finance Pringsewu. *Knsi 2018*, 608–612.
- SUGIANTO, C. A. (2015). Penerapan Teknik Data Mining Untuk Menentukan Hasil Seleksi Masuk Sman 1 Cibeber Untuk Siswa Baru Menggunakan Decision Tree. *Tedc*, 9(1), 39–43.
- WIBOWO, A. (2017). 10 FOLD-CROSS VALIDATION. Retrieved from <https://mti.binus.ac.id/2017/11/24/10-fold-cross-validation/>

*Halaman ini sengaja dikosongkan*