

## KLASIFIKASI DATA MINING MENGGUNAKAN ALGORITMA C4.5 TERHADAP KEPUASAN PELANGGAN SEWA KAMERA CIKARANG

Sutan Faisal  
sutan.faisal@ubpkarawang.ac.id  
Universitas Buana Perjuangan Karawang  
Karawang, Indonesia

### ABSTRAK

*Kepuasan pelanggan merupakan salah satu tujuan dari perusahaan dalam memberikan pelayanan kepada pelanggannya baik perusahaan jasa maupun non jasa. Salah satu perusahaan jasa penyedia sewa kamera yang berkomitmen untuk kepuasan pelanggannya adalah Sewa Kamera Cikarang. Penelitian ini bertujuan untuk menganalisa kepuasan pelanggan sewa kamera cikarang dengan menggunakan algoritma K-Nearest Neighbour (KNN). Atribut masukan kepuasan pelanggan dalam penelitian ini mencakup harga, fasilitas, pelayanan dan loyalitas. Keluaran hasil dari atribut masukan di atas adalah puas dan tidak puas. Penelitian ini diharapkan untuk membantu Sewa Kamera Cikarang untuk meningkatkan kepuasan pelanggan dan meningkatkan laba pada Sewa Kamera Cikarang. Hasil penelitian yang dapat dicapai menggunakan algoritma C4.5 adalah accuracy = 95%, classification recall = 100%, Classification precision = 92.31% dan AUC = 0.948. Hasil penelitian ini bisa menjadikan referensi untuk membangun sebuah aplikasi yang dapat memudahkan perusahaan dalam memperoleh informasi mengenai kepuasan pelanggannya.*

Kata Kunci: *Data mining, classification, Algoritma C 4.5, kepuasan pelanggan*

### 1. PENDAHULUAN

Untuk mencapai keberhasilan dalam sebuah usaha bidang jasa, kepuasan pelanggan harus menjadi dasar dari keputusan manajemen, sehingga manajemen harus menjadikan peningkatan kepuasan pelanggan sebagai suatu sasaran yang mendasar. Dalam rangka untuk memberikan pelayanan yang berkualitas perusahaan secara *continue* harus mengadakan peningkatan kualitas SDM dan peralatan yang disewakannya. Langkah ini penting untuk memperbaiki pelayanan dari waktu ke waktu.

Namun perkembangan yang pesat dalam dunia usaha ini mengharuskan perusahaan menghadapi ketatnya persaingan. Pada umumnya banyak cara untuk mempertahankan pelanggannya selama-lamanya, didalam persaingan usaha yang sangat ketat ini sangat sulit Untuk mewujudkan hal itu mengingat banyaknya perubahan yang dapat terjadi setiap saat. Seperti perubahan pada diri pelanggan.pesaing maupun perubahan kondisi secara luas yang selalu berubah secara dinamis. Hal ini membuat para pengambil kebijakan untuk terus mengembangkan suatu strategi yang dapat mencapai sasaran pertumbuhan penyewaan, peningkatan porsi pasar , serta pencapaian kemampuan sebagai dasar pertumbuhan berkelanjutan (Tan S, Kumar P, Steinbach M. 2005).

Sementara itu, *Data mining* adalah proses pengekstrakan informasi dari kumpulan data yang besar (Basuki, Ahmad dan Syarif, Iwan , 2003).. Pengekstrakan informasi dilakukan berdasarkan metode *data mining* yang akan digunakan. Ada beberapa metode serta algoritma *data mining* yang digunakan untuk mengekstrak informasi antara lain: metode klasifikasi, metode asosiasi, metode *clustering*, metode prediksi, dan metode estimasi.

### 2. DATA MINING

Menurut Gartner Group, Bahwa *data mining* adalah proses menemukan hubungan baru yang mempunyai arti, pola dan kebiasaan dengan memilih sebagian besar data yang disimpan dalam media penyimpanan dengan menggunakan teknologi pengenalan pola seperti Teknik statistic dan matematika. *Data mining* merupakan gabungan dari beberapa disiplin ilmu yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, *database*, dan isualisasi untuk penanganan permasalahan pengambilan informasi dari *database* yang besar [1]

*Data mining* menurut David Hand, Heikki Mannila, dan Padhraic Smyth dari MIT adalah analisa terhadap *data mining (big data)* untuk menemukan hubungan yang jelas serta menyimpulkannya yang belum diketahui sebelumnya dengan cara terkini di pahami dan berguna bagi pemilik data tersebut[2]. *Data mining* adalah proses yang menggunakan teknik statistik,

matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. *Data mining* merupakan serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [2].

Jadi *data mining* dapat diartikan sebagai berikut:

1. *Data mining* merupakan suatu proses otomatis untuk menganalisa data.
2. Data yang akan diproses merupakan data yang sangat besar, sehingga sulit untuk memproses dengan manual
3. Tujuan *data mining* adalah mendapatkan hubungan atau pola/ Teknik yang memberikan indikasi yang bermanfaat.

*Data mining* bukanlah hal yang baru, satu kesulitan untuk mendefinisikan *data mining* adalah kenyataan bahwa *data mining* mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang dulu sudah mapan terlebih dulu. *data mining* memiliki akar yang Panjang dari bidang ilmu yang berbeda seperti kecerdasan buatan (*artificial ntelligent*), *machine learning*, statistik, *database*, dan juga information retrieval [2]

*Database* yang tersimpan di media penyimpanan jarang sekali dimanfaatkan oleh sebagian besar penggunaanya dan bahkan jangka waktu tertentu data-data tersebut dihapus karena dianggap sampah dan hanya memenuhi media penyimpanan saja. Anggapan tersebut tidak sepenuhnya benar, karena sesungguhnya *database* dalam ukuran yang besar dapat memberikan informasi yang dibutuhkan untuk berbagai kepentingan, baik untuk kepentingan bisnis dalam mengambil keputusan maupun untuk ilmu pengetahuan dan penelitian.

### 3. CLASSIFICATION

Salah satu tugas yang dapat dilakukan dengan *data mining* adalah pengklasifikasian. Klasifikasi pertama kali diterapkan pada bidang tanaman yang mengklasifikasi suatu spesies tertentu, seperti yang dilakukan oleh Carolus von Linne (atau dikenal dengan nama Carolus Linnaeus) yang pertama kali mengklasifikasi spesies berdasarkan karakteristik fisik. Selanjutnya dia dikenal sebagai bapak klasifikasi [3].

Dalam klasifikasi terdapat target variable kategori. Metode-metode / model-model yang telah dikembangkan oleh periset untuk menyelesaikan kasus klasifikasi antara lain Pohon keputusan, Jaringan saraf tiruan. Dll.

Operator ini digunakan untuk evaluasi kinerja statistik dari tugas klasifikasi binominal yaitu tugas klasifikasi dimana atribut label memiliki tipe binominal. Operator ini memberikan daftar nilai kriteria kinerja dari tugas klasifikasi binominal.

Operator ini harus digunakan secara khusus untuk evaluasi kinerja hanya tugas klasifikasi binominal yaitu tugas klasifikasi dimana atribut label memiliki tipe binominal. Banyak operator evaluasi kinerja lainnya juga tersedia di RapidMiner misalnya. operator Kinerja, operator Kinerja (Klasifikasi), operator Kinerja (Regresi), dan lain-lain. Operator Kinerja (Binominal Klasifikasi) digunakan hanya dengan tugas klasifikasi binominal saja. Di sisi lain, operator Kinerja secara otomatis menentukan jenis tugas pembelajaran dan menghitung kriteria yang paling umum untuk tipe itu. Anda dapat menggunakan operator Performance (User-Based) jika Anda ingin menulis ukuran kinerja Anda sendiri.

### 4. DECISION TREE

Pohon keputusan adalah salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia. Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Konsep dari pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan.

Manfaat utama dari penggunaan pohon keputusan adalah kemampuannya untuk mem-break down proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambilan keputusan akan lebih menginterpretasikan solusi dari permasalahan. Pohon Keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi.

*Decision tree* (pohon keputusan) adalah sebuah diagram alir yang mirip dengan struktur pohon, dimana setiap internal *node* menotasikan atribut yang diuji, setiap cabangnya merepresentasikan hasil dari atribut tes tersebut, dan *leaf node* merepresentasikan kelas-kelas tertentu atau distribusi dari kelas-kelas (Han J, Kamber M, 2001).

Klasifier pohon keputusan merupakan teknik klasifikasi yang sederhana yang banyak digunakan. Bagian ini membahas bagaimana pohon keputusan bekerja dan bagaimana pohon keputusan dibangun. Seringkali untuk mengklasifikasikan obyek, kita ajukan urutan pertanyaan sebelum bisa kita tentukan kelompoknya.

## 5. ALGORITMA C 4.5

Algoritma C4.5 adalah salah satu algoritma untuk mengubah fakta yang besar menjadi pohon keputusan (*decision tree*) yang merepresentasikan aturan (*rule*). Tujuan dari pembentukan pohon keputusan dalam algoritma C4.5 adalah untuk mempermudah dalam penyelesaian permasalahan.

Dalam menggunakan algoritma C4.5 terdapat beberapa tahapan yang umum yaitu pertama mengubah bentuk data dalam tabel menjadi model pohon kemudian mengubah model pohon menjadi aturan (*rule*) dan terakhir menyederhanakan rule (Basuki, Ahmad dan Syarif, Iwan , 2003).

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

- Hitung jumlah data, jumlah data berdasarkan anggota atribut hasil dengan syarat tertentu. Untuk proses pertama syaratnya masih kosong.
- Pilih atribut sebagai *Node*.
- Buat cabang untuk tiap-tiap anggota dari *Node*.
- Periksa apakah nilai entropy dari anggota *Node* ada yang bernilai nol. Jika ada, tentukan daun yang terbentuk. Jika seluruh nilai entropy anggota *Node* adalah nol, maka proses pun berhenti.
- Jika ada anggota *Node* yang memiliki nilai entropy lebih besar dari nol, ulangi lagi proses dari awal dengan *Node* sebagai syarat sampai semua anggota dari *Node* bernilai nol. *Node* adalah atribut yang mempunyai nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung nilai gain suatu atribut digunakan rumus seperti yang tertera dalam persamaan berikut:

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

Keterangan:

S: himpunan kasus

A : atribut

n : jumlah partisi atribut A

|S<sub>i</sub>| : jumlah kasus pada partisi ke-i

|S| : jumlah kasus dalam S

penghitungan nilai entropi dapat dilihat pada persamaan berikut (kusrini, Emha Taufiq Lut :2009).

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

Keterangan:

S: himpunan kasus

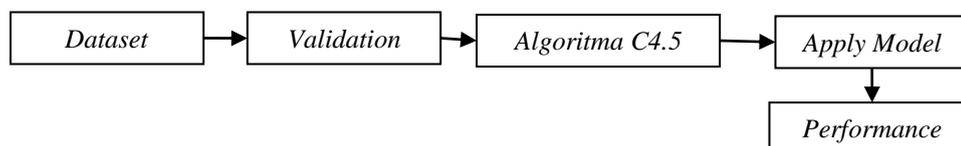
A : fitur

n : jumlah partisi S

p<sub>i</sub> : proporsi dari S<sub>i</sub> terhadap S

## 6. METHODOLOGY

Pada penelitian ini menggunakan *tools* pengujian rapidminer studio 9.0, dengan menggunakan methodology sebagai berikut:



Gambar 1. Methodology yang digunakan

### 6.1. Dataset

*Dataset* merupakan kumpulan data, yang mana biasanya satu dataset merepresentasikan satu tabel *database*, atau bisa juga suatu matriks data yang mana tiap kolom mewakili variabel tertentu, tiap baris merepresentasikan banyaknya data.

Dalam penelitian ini menggunakan Operator Retrieve memuat Objek RapidMiner ke dalam Proses. Objek ini berupa *ExampleSet* namun bisa juga berupa *Collection* atau Model. Mengambil data dengan cara ini juga menyediakan data meta dari RapidMiner *Object*.

## 6.2. Validation

*Validation* sebagai Operator untuk melakukan validasi sederhana yaitu secara acak membagi *ExampleSet* menjadi satu set pelatihan dan menetapkan tes dan mengevaluasi modelnya. Operator ini melakukan validasi *split* untuk memperkirakan kinerja operator pembelajaran (biasanya pada kumpulan data yang tidak terlihat). Hal ini terutama digunakan untuk memperkirakan seberapa akurat sebuah model (dipelajari oleh operator pembelajaran tertentu) akan tampil dalam praktik.

## 6.3. Algoritma C 4.5

Dalam penelitian ini dilakukan eksperimen menggunakan metode klasifikasi *data mining decision tree* algoritma C 4.5 terhadap data kuisisioner kepuasan pelanggan pada Sewa Kamera Cikarang. Data akan diolah dengan menggunakan algoritma C4.5 dan menghasilkan model, maka terhadap model yang dihasilkan tersebut dilakukan pengujian *Cross Validation* yang menghasilkan *accuracy*, *precision*, *recall* dan AUC.

## 6.4. Apply Model

*Apply Model* adalah suatu model pertama dilatih pada *ExampleSet* oleh Operator lain, yang sering merupakan algoritma pembelajaran. Setelah itu, model ini bisa diaplikasikan pada *ExampleSet* yang lain. Biasanya, tujuannya adalah untuk mendapatkan prediksi pada data yang tidak terlihat atau untuk mentransformasikan data dengan menerapkan model preprocessing. *ExampleSet* dimana model diterapkan, harus kompatibel dengan Atribut model. Ini berarti, bahwa *ExampleSet* memiliki jumlah, urutan, jenis, dan peran Atribut yang sama seperti *ExampleSet* yang digunakan untuk menghasilkan model.

## 6.5. Performance

Operator ini digunakan untuk evaluasi kinerja statistik dari tugas klasifikasi binomial yaitu tugas klasifikasi dimana atribut label memiliki tipe binomial. Operator ini memberikan daftar nilai kriteria kinerja dari tugas klasifikasi binomial.

Untuk mengukur hasil penelitian ini dengan menggunakan *confusion matrix* (*accuracy*, *Classification recall*, *Classification precision*) dan kurva ROC.

## 7. HASIL DAN PEMBAHASAN

### 7.1. Hasil Analisis Data Set

*Dataset* yang digunakan adalah *dataset* kuisisioner kepuasan pelanggan Sewa Kamera Cikarang, *dataset* ini berisi data – data informasi tentang kuisisioner kepuasan pelanggan mengenai harga, fasilitas, pelayanan dan loyalitas. Keseluruhan data pada *dataset* ini berjumlah 100 *records* yang setiap record nya memiliki 10 atribut diantaranya :

1. No (integer, role : id)
2. Nama (polynomial)
3. Harga X1 (integer)
4. Fasilitas X2 (integer)
5. Pelayanan X3 (integer)
6. Loyalitas X4 (integer)
7. Hasil (binomial : puas & tidak puas)
8. Confidence tidak puas (real)
9. Confidence puas (real)
10. Prediksi hasil (binomial)

Dari atribut *dataset* di atas (1 sampai 10) akan dilakukan proses *training* dan *test* menggunakan metode *deep learning*, sedangkan atribut ke – 10 akan menjadi target hasil dari proses klasifikasi. dan di sini kami akan mencoba menganalisis perbedaan dari akurasi dan eror yang didapat dengan membandingkan prediksi hasil dan hasil.

Kuisisioner dapat digambarkan di bawah ini

**KUISISIONER PENELITIAN  
KEPUASAN PELANGGAN TERHADAP SEWA KAMERA CIKARANG**

Subjungan dengan penelitian dosen untuk melakukan tridharma perguruan tinggi dan masukan untuk sewa kamera cikarang. Kami mengharapkan bantuan anda untuk memberikan masukan dengan mengisi kuisisioner ini dengan keadaan sebenarnya. Bantuan anda sangat berharga untuk penelitian ini. Terima kasih atas waktu dan kebaikan anda untuk menanggapi kuisisioner ini.

Nama :  
Jenis Kelamin :  
Umur :

Isilah kuisisioner ini sesuai dengan penilaian anda, dengan tanda (√)

1 : 0 – 40 (sangat Kurang)      2: 41 – 60 (Kurang)      3 : 61 – 70 (Cukup)  
4 : 71 – 80 (Baik)      5 : 81 – 90 (Cukup Baik)      6: 91 – 100 (Sangat Baik)

Beri tanda silang (X) pada kotak yang sesuai.

No	Aspek yang dinilai	Skor					
		1	2	3	4	5	6
<b>Harga</b>							
1	Sewa Kamera Cikarang harganya relative terjangkau	<input type="checkbox"/>					
<b>Fasilitas</b>							
2	Fasilitas peralatan pada sewa kamera cikarang lengkap dan berfungsi dengan baik (kamera, tripod, lensa dll)	<input type="checkbox"/>					
3	Tempat duduk bersih dan rapih	<input type="checkbox"/>					
4	Metode pembayaran yang diterapkan pada sewa kamera cikarang	<input type="checkbox"/>					
<b>Pelayanan</b>							
5	Karyawan memberikan penjelasan yang baik mengenai produk sewa kamera cikarang	<input type="checkbox"/>					
6	Persyaratan pemijaman pada sewa kamera cikarang	<input type="checkbox"/>					
7	Item buka operasional sewa kamera cikarang	<input type="checkbox"/>					
8	Promo yang sering dilakukan pada sewa kamera cikarang	<input type="checkbox"/>					
<b>Loyalitas</b>							
9	Secara keseluruhan anda lebih sering memilih sewa kamera cikarang dari pada tempat yang lain	<input type="checkbox"/>					
10	Jika ada seseorang yang meminta saran ke anda, anda akan merekomendasikan sewa kamera cikarang	<input type="checkbox"/>					
11	Anda akan mengatakan hal hal yang positif mengenai sewa kamera cikarang kepada orang lain	<input type="checkbox"/>					
12	Jika anda sewa kamera anda akan menempatkan sewa kamera cikarang sebagai pilihan anda	<input type="checkbox"/>					

Bagaimana kepuasan anda tentang harga, fasilitas dan , pelayanan sewa kamera cikarang ?

PUAS      8  
TIDAK PUAS

Gambar 2 Form Kuisisioner Data kuisisioner yang telah diolah menggunakan MS Excel

No	Nama	Jenis Kelamin (L/P)	X1 =		X2 =			X3 =				X4 =				Rata-Rata X4	Hasil			
			harga	Rata X1	fasilitas			pelayanan				loyalitas								
			1		2	3	4	5	6	7	8	9	10	11	12					
1	Dian	P	5	5	2	3	3	2,67	3	4	3	4	3,5	4	3	3	3	3,25	TIDAK PUAS	
2	Wiwini	L	5	5	5	3	4	4,00	4	4	4	4	4	4	4	4	3	5	4	PUAS
3	Ivan	L	5	5	5	3	4	4,00	4	4	4	4	4	4	4	4	3	4	3,75	PUAS
4	Anton	L	3	3	4	3	3	3,33	4	2	2	4	3	4	2	2	3	2,75	TIDAK PUAS	
5	Taufik	L	5	5	4	3	4	3,67	4	3	4	4	3,75	4	4	4	4	4	PUAS	
6	Doni	L	3	3	4	3	4	3,67	4	3	4	5	4	4	4	4	4	4	PUAS	
7	Didi	L	4	4	3	2	4	3,00	4	2	2	3	2,75	3	2	2	4	2,75	TIDAK PUAS	
8	Adi	L	5	5	5	4	4	4,33	4	3	4	5	4	4	3	4	4	3,75	PUAS	
9	Joko	L	6	6	5	4	5	4,67	4	3	4	5	4	4	3	4	4	3,75	PUAS	
10	Gina	P	5	5	5	4	5	4,67	4	3	4	5	4	4	3	4	3	3,5	PUAS	
11	Ahmad	L	3	3	4	4	3	3,67	4	2	2	5	3,25	4	2	2	3	2,75	TIDAK PUAS	
12	Yanto	L	5	5	4	3	5	4,00	4	3	4	3	3,5	4	3	4	3	3,5	PUAS	
13	Fifi	P	3	3	3	4	2	3	3,00	4	2	2	3	2,75	4	2	2	3	2,75	TIDAK PUAS
14	Fika	P	3	3	4	3	4	3,67	3	2	2	3	2,5	3	2	2	3	2,5	TIDAK PUAS	
15	Rendi	L	6	6	4	3	4	3,67	5	2	3	3	3,25	4	4	3	4	3,75	PUAS	
16	Dede	L	4	4	4	4	4	4,00	5	2	3	3	3,25	5	4	3	4	4	PUAS	
17	Santi	P	3	3	3	4	3	3,33	5	2	2	4	3,25	5	4	3	4	4	TIDAK PUAS	
18	Dadan	L	5	5	5	4	5	4,67	6	3	3	4	4	5	4	3	4	4	PUAS	
19	Jamilah	P	2	2	5	5	5	5,00	3	3	4	4	3,5	6	4	4	6	5	PUAS	
20	Fauzan	L	3	3	3	3	3	3,00	4	2	3	4	3,25	3	2	2	4	2,75	TIDAK PUAS	
21	Vina	P	3	3	3	4	3	3,33	3	2	2	4	2,75	3	2	2	4	2,75	TIDAK PUAS	
22	Lukman	L	5	5	5	2	3	3,33	4	2	4	4	3,5	3	5	4	4	4	PUAS	
23	Bayu	L	5	5	6	4	3	4,33	4	4	4	6	4,5	3	4	3	4	3,5	PUAS	
24	Antik	L	5	5	6	4	4	4,67	4	4	4	4	4	3	4	3	4	3,5	PUAS	

Gambar 3 Data kuisisioner yang telah diolah

Data kuisisioner yang telah diolah lagi

No	Nama	Harga X1	Fasilitas X2	Pelayanan X3	Loyalitas X4	Hasil
1	Dian	5	2.67	3.50	3.25	TIDAK PUAS
2	Wiwin	5	4.00	4.00	4.00	PUAS
3	Iwan	5	4.00	4.00	3.75	PUAS
4	Anton	3	3.33	3.00	2.75	TIDAK PUAS
5	Taufik	5	3.67	3.75	4.00	PUAS
6	Doni	3	3.67	4.00	4.00	PUAS
7	Didi	4	3.00	2.75	2.75	TIDAK PUAS
8	Adi	5	4.33	4.00	3.75	PUAS
9	Joko	6	4.67	4.00	3.75	PUAS
10	Gina	5	4.67	4.00	3.50	PUAS
11	Ahmad	3	3.67	3.25	2.75	TIDAK PUAS
12	Yanto	5	4.00	3.50	3.50	PUAS
13	Fifi	3	3.00	2.75	2.75	TIDAK PUAS
14	Fika	3	3.67	2.50	2.50	TIDAK PUAS
15	Rendi	6	3.67	3.25	3.75	PUAS
16	Dede	4	4.00	3.25	4.00	PUAS
17	Santi	3	3.33	3.25	4.00	TIDAK PUAS
18	Deden	5	4.67	4.00	4.00	PUAS
19	Jamilah	2	5.00	3.50	5.00	PUAS
20	Fauzan	3	3.00	3.25	2.75	TIDAK PUAS
21	Vina	3	3.33	2.75	2.75	TIDAK PUAS
22	Lukman	5	3.33	3.50	4.00	PUAS
23	Bayu	5	4.33	4.50	3.50	PUAS
24	Antik	5	4.67	4.00	3.50	PUAS

Gambar Data kuisisioner yang telah diolah lagi

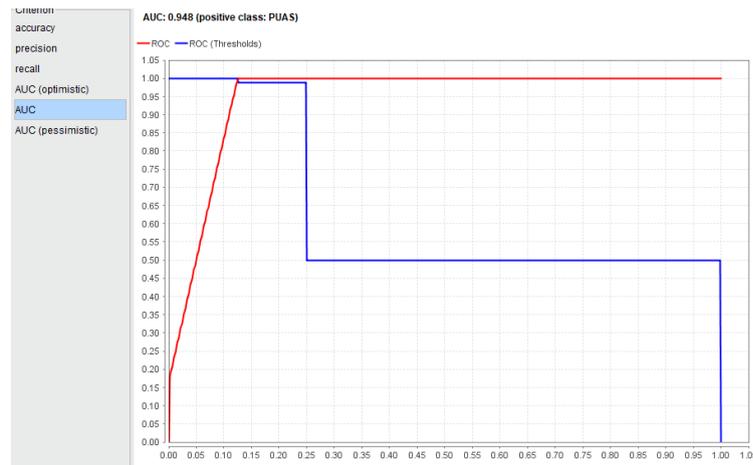
## 7.2. Hasil Percobaan dan Evaluasi

Pada percobaan ini terdapat 10 atribut yang akan di training dan 2 nilai yang menunjukkan target (klasifikasi) pada atribut ke – 10 yang berarti pada algoritma C4.5 diinisialisasi kan 9 atribut *input* dan 1 atribut *output*.

Hasil dari penelitian ini adalah :

### 1. Tabel *Confusion Matrix*

	true TIDAK PUAS	true PUAS	class precision
pred. TIDAK PUAS	7	0	100.00%
pred. PUAS	1	12	92.31%
class recall	87.50%	100.00%	



## 2. Pervormance Vector

### PerformanceVector

```
PerformanceVector:
accuracy: 95.00%
ConfusionMatrix:
True:  TIDAK PUAS      PUAS
TIDAK PUAS:  7         0
PUAS:  1         12
precision: 92.31% (positive class: PUAS)
ConfusionMatrix:
True:  TIDAK PUAS      PUAS
TIDAK PUAS:  7         0
PUAS:  1         12
recall: 100.00% (positive class: PUAS)
ConfusionMatrix:
True:  TIDAK PUAS      PUAS
TIDAK PUAS:  7         0
PUAS:  1         12
AUC (optimistic): 1.000 (positive class: PUAS)
AUC: 0.948 (positive class: PUAS)
AUC (pessimistic): 0.896 (positive class: PUAS)
```

## 8. KESIMPULAN DAN SARAN

### 8.1. Kesimpulan

Metode klasifikasi dengan menggunakan algoritma C 4.5 sangat baik untuk menentukan kebenaran dari klasifikasi dalam *data mining*. Terbukti dengan hasil *accuracy* = 95%, *Classification recall* = 100%, *Classification precision* = 92.31% dan *AUC* = 0.948.

### 8.2. Saran

Dari hasil pengujian yang telah dilakukan dan hasil kesimpulan yang didapatkan maka ada saran yang diberikan :

1. Untuk penelitian selanjutnya bisa digunakan algoritma yang lain
2. Hasil penelitian ini bisa digunakan untuk keperluan perusahaan dalam memprediksi kepuasan pelanggannya.

### Daftar Pustaka

- [1] Barber, I. "Bayesian Opinion Mining." *Tersedia di: <http://phpir.com/bayesian-opinion-mining> [diunduh: 10 Nov 2010]* (2010).
- [2] Bramer, Max. *Principles of data mining*. Vol. 180. London: Springer, 2007.
- [3] Basuki, Achmad dan Syarif, Iwan. 2003. Modul Ajar *Decision tree*. Surabaya : PENS-ITS.
- [4] Giudici, Paolo, and Silvia Figini. *Front Matter*. John Wiley & Sons, Ltd, 2009. *Applied Data mining for Business and Industry*
- [5] Gorunescu, Florin. *Data mining: Concepts, models and techniques*. Vol. 12. Springer Science & Business Media, 2011.
- [6] Han J, Kamber M. 2001. *Data mining : Concepts and Techniques*. Simon Fraser University, Morgan Kaufmann Publishers.
- [7] Larose, D.T, 2006. *Discovering Knowledge in Data: An Introduction to Data mining*. John Willey & Sons, Inc.
- [8] Maimon, Oded, and Lior Rokach, eds. *Data mining and knowledge discovery handbook*. Vol. 2. New York: Springer, 2005.
- [9] Powers, David Martin. "Evaluation: from *precision*, *recall* and F-measure to ROC, informedness, markedness and correlation." (2011).
- [10] Rapid-I GmbH. (2008). *Rapidminer-4.2-tutorial*. Germany: Rapid-I.

- [11] Sachdeva, M., Zhu, S., Wu, F., Wu, H., Walia, V., Kumar, S., ... & Mo, Y. Y. (2009). p53 represses c-Myc through induction of the tumor suppressor miR-145. *Proceedings of the National Academy of Sciences*, 106(9), 3207-3212.
- [12] Santosa, Budi. 2007. *Data mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis, Teori dan Aplikasi*. Graha Ilmu Yogyakarta
- [13] Tan S, Kumar P, Steinbach M. 2005. Introduction To *Data mining*. Addison Wesley.
- [14] Vitaro, F., Brendgen, M., Larose, S., & Trembaly, R. E. (2005). Kindergarten Disruptive Behaviors, Protective Factors, and Educational Achievement by Early Adulthood. *Journal of educational psychology*, 97(4), 617.
- [15] M Rizki Ilham, Purwanto. 2016. Implementasi *Data mining* Menggunakan Algoritma C 4.5 Untuk Prediksi Kepuasan Pelanggan. UDINUS Semarang