

Implementasi Algoritma Klasifikasi C4.5 Untuk Memprediksi Kelayakan Pembelian Kendaraan (Implementation of Classification Algorithm C4.5 To Predict the Eligibility of Vehicle Purchases)

Egih Sugiатna^{[1]*}, Anton Maulana Ibrahim^[2], Ichwansyah Abdul Hadi^[3]

^[1,2,3] Department of Computer Science, Faculty of Technology and Information Universitas Budi Luhur Jakarta, Indonesia

E-mail: [egihsgiatna, anton.jagalempeni, ichwansyahhadi]@gmail.com

KEYWORDS:

Data mining, Decision tree, Naive Bayes, C4.5, classification

ABSTRACT

To choose the right vehicle according to the needs and funds owned by consumers, requires careful analysis taking into account many criteria and factors. The criteria used as benchmarks in choosing a vehicle include price, maintenance costs, number of doors, room capacity, luggage area, and vehicle safety level. To process all of these criteria requires a way that can help consumers in choosing the most appropriate vehicle according to their needs. The methodology used is the literature study methodology to choose the best algorithm for processing existing datasets. C4.5 algorithm is the algorithm used in this study. C4.5 algorithm is then implemented to determine the car that will be taken by the buyer according to the background of buying criteria (purchase price), maint (maintenance costs), doors (number of doors), persons (capacity to load passengers), lug_boot (luggage area), safety (vehicle safety level). The results of the experiment were classified as good where the results shown were 92.4% accuracy, 92.4% Precision and 92.4% Recall.

KATA KUNCI:

Data mining, Decision tree, Naive Bayes, C4.5, klasifikasi

ABSTRAK

Untuk memilih kendaraan yang tepat sesuai dengan kebutuhan dan dana yang dimiliki oleh konsumen, membutuhkan analisis yang cermat memperhitungkan banyak kriteria dan faktor. Kriteria yang digunakan sebagai tolok ukur dalam memilih kendaraan antara lain harga, biaya perawatan, jumlah pintu, kapasitas ruangan, luas bagasi, dan tingkat keamanan kendaraan. Untuk memproses semua kriteria ini dibutuhkan cara yang dapat membantu konsumen dalam memilih kendaraannya yang paling tepat sesuai dengan kebutuhannya. Metodologi yang digunakan adalah metodologi studi literatur untuk memilih algoritma terbaik untuk memproses dataset yang ada. Algoritma C4.5 merupakan algoritma yang digunakan terpilih dalam penelitian ini. Algoritma C4.5 kemudian diimplementasikan untuk menentukan mobil yang akan diambil oleh pembeli sesuai dengan latar belakang kriteria buying (harga beli), maint (biaya perawatan), doors (jumlah pintu), persons (daya tampung untuk memuat penumpang), lug_boot (luas bagasi), safety (tingkat keamanan kendaraan). Hasil dari percobaan tergolong baik dimana hasil yang ditunjukkan adalah akurasi 92.4%, Precision 92.4% dan Recall 92.4%.

I. PENDAHULUAN

Perkembangan industri otomotif di Indonesia terus tumbuh sangat pesat, ini bisa dilihat dengan meningkatnya jumlah kendaraan di Indonesia bahkan brand baru bermunculan. Dengan sistem transportasi yang baik untuk efisiensi waktu, perpindahan individu dari satu lokasi ke lokasi lain menggunakan kendaraan pribadi adalah salah satu pilihan yang dianggap lebih nyaman dinami. Banyak alasan orang akhirnya memutuskan untuk memiliki kendaraan pribadi roda empat. Mulai dari faktor cuaca, faktor mobilitas, bahkan faktor gaya hidup.

Begitu banyak merek dan harga kompetitif yang ditawarkan oleh produsen otomotif, membuat dealer mengalami kesulitan membantu konsumen menentukan jenis kendaraan yang akan mereka beli. Untuk memilih kendaraan yang tepat sesuai dengan kebutuhan dan cocok dengan kemampuan dana yang dimiliki oleh konsumen, diperlukan analisis yang cermat yang memperhitungkan banyak kriteria dan faktor. Kriteria digunakan sebagai patokan dalam memilih kendaraan, antara lain harga, biaya perawatan, jumlah pintu, kapasitas ruangan, luas bagasi, dan tingkat keamanan kendaraan. Untuk

mengatasi masalah tersebut peneliti memilih C4.5. Algoritma C4.5 adalah salah satu algoritma yang diterapkan dalam proses data mining. Algoritma C4.5 merupakan perpanjangan dari algoritma ID3 Quinlan sendiri untuk menghasilkan pohon keputusan (*decision tree*). Sama seperti dengan CART, algoritma C4.5 secara rekursif mengunjungi setiap simpul keputusan, memilih pemisahan optimal, sampai tidak ada pemisahan lebih lanjut [1].

Fitriana et al. [2], Dalam penelitian ini, dibuat sebuah sistem untuk memilih kendaraan yang dengan Kriteria yang digunakan sebagai patokan dalam memilih kendaraannya harga, suku cadang, Volume silinder, kekuatan kendaraan. Algoritma Naive Bayes adalah algoritma yang digunakan dalam penelitian ini. Dari data pembelian 20 mobil yang digunakan dalam pengujian dengan metode Naive Bayes. Berdasarkan penelitian tersebut data pembelian mobil bekas sebagai data pelatihan, metode Naive Bayes berhasil diklasifikasikan 15 data dari 20 data diuji. Jadi metode Naive Bayes berhasil memprediksi pembelian mobil dengan persentase akurasi 75%.

Wajhillah, Rusda Nusa [3] telah melakukan penelitian terkait prediksi penyakit jantung dengan menggunakan algoritma C4.5. Hasil penelitian untuk nilai akurasi algoritma klasifikasi C4.5 senilai 81,25%, sedangkan untuk nilai akurasi algoritma klasifikasi C4.5 berbasis PSO sebesar 93,75% sehingga tampak selisih nilai akurasi yaitu sebesar 12,5%. Sedangkan evaluasi menggunakan ROC curve diperoleh hasil untuk algoritma klasifikasi C4.5 bernilai 0,718 dengan tingkat diagnosa Fair Classification dan algoritma klasifikasi C4.5 berbasis PSO bernilai 0,855 dengan tingkat diagnosa yang sama yaitu Good Classification, didapatkan selisih nilai AUC sebesar 0,137. Sehingga dapat disimpulkan bahwa penerapan teknik optimasi particleswarm optimization dapat meningkatkan nilai akurasi pada algoritma C4.5.

Diantara teknik *data mining* untuk mengklasifikasi, mengidentifikasi pola dan memprediksi tren dari data set adalah Naive Bayes dan C4.5. Namun peneliti memilih C4.5 karena pada pemelitan sebelumnya [4][5] menyatakan "algoritma C4.5 lebih baik daripada naive Bayes karena dalam algoritma C4.5 ada deskripsi lengkap dari data dan kegiatan pra-pemrosesan yang diperlukan".

Pohon keputusan (*decision tree*) merupakan metode klasifikasi dan prediksi yang sangat terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami, juga dapat diekspresikan dalam bentuk bahasa basis data seperti SQL (*Structured Query Language*) untuk mencari record pada kategori tertentu [6]. Ketentuan algoritma C4.5 untuk membangun pohon keputusan (*decision tree*) adalah sebagai berikut:

- Tentukan nilai *gain* tertinggi sebagai akar
- Buat cabang untuk masing-masing atribut
- Bagi kasus dalam cabang
- Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Tahapan dalam perhitungan algoritma *decision tree* C4.5 dengan beberapa tahapan :

1. Menyiapkan data training
2. Menentukan akar dari pohon yang didapat dari nilai *gain* tertinggi.
3. Menghitung nilai entropy [5]

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

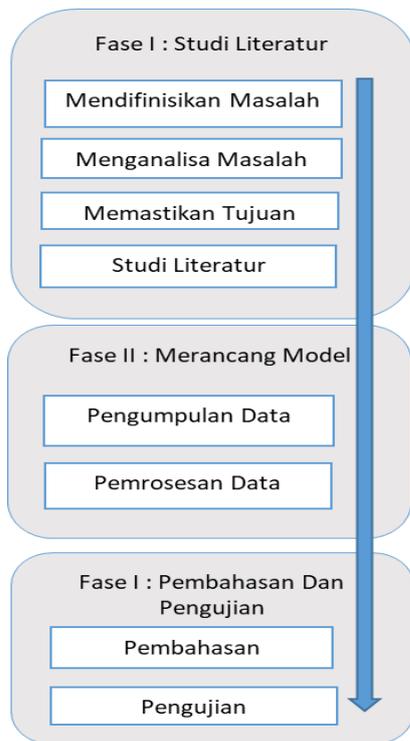
4. Menghitung nilai gain

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

Setelah nilai *gain* ditemukan maka akan dilanjutkan dalam proses pohon keputusan (*decision tree*).

II. METODOLOGI

Kerangka kerja ini adalah langkah-langkah yang akan dilakukan agar pecahkan masalah yang akan dibahas. Kerangka kerja ini penelitian dapat dilihat pada Gambar 1



Gbr. 1 Kerangka Penelitian

A. Fase Studi Literatur

1) Mendefinisikan Masalah

Pada tahap ini akan dipelajari lebih dalam masalah yang ada. Tahap ini adalah langkah pertama untuk menentukan rumusan masalah penelitian, yang mana masalahnya adalah adanya kesulitan bagi calon pembeli mobil untuk memilih mobil yang tepat, dikarenakan semakin banyaknya variasi mobil saat ini.

2) Menganalisis Masalah

Masalah yang ditemukan kemudian akan dianalisis. Langkah dalam proses analisis masalah adalah langkah untuk memahami masalah yang telah ditentukan. Setelah dianalisis masalah yang ditemukan adalah di saat varian mobil semakin banyak dengan berbagai fitur yang beragam, hal itu membuat calon pembeli tidak tahu bagaimana cara memilih kendaraan yang tepat untuk dibeli karena kendaraan tidak bias diklasifikasikan dengan mudah.

3) Menetapkan Tujuan

Berdasarkan pemahaman masalah yang dimiliki telah dianalisis, langkah selanjutnya adalah menentukan tujuan yang ingin dicapai dalam penelitian ini. Dalam tujuan ini target yang ingin

dicapai adalah bagaimana cara mengklasifikasikan kendaraan sesuai dengan variable yang ada sehingga bias membantu pembeli untuk memilih kendaraan (mobil) yang tepat.

4) Studi literatur

Penelitian ini dilakukan untuk melengkapi perbendaharaan aturan, konsep, teori yang mendukung dalam pemecahan masalah dalam penelitian ini. Penelitian juga dilakukan melalui jurnal yang memiliki hubungan dengan penelitian dan referensi lainnya. Penelitian ini bertujuan untuk mengumpulkan data, baik data pokok maupun data pendukung, di mana semua data dibutuhkan dalam penelitian.

B. Merancang Model

1) Mengumpulkan data

Kumpulkan data untuk pelatihan dan pengujian *Decision Tree* dengan algoritma C4.5. Semakin banyak data yang didapatkan, semakin baik menyelesaikan masalah. Teknik pengumpulan data menggunakan data sekunder yaitu data yang diambil dari sebuah *web repository machine learning* [7]. Data tersebut berasal dari model keputusan hirarkis sederhana yang awalnya dikembangkan untuk demonstrasi DEX, M. Bohanec, V. Rajkovic: Sistem pakar untuk pengambilan keputusan. [8] Data termasuk informasi 1728 mobil yang digunakan untuk penelitian. Setiap catatan memiliki 7 atribut, yang dijelaskan pada Tabel 1. [9]

TABEL I. NAMA ATRIBUT DAN NILAINYA

Nama Atribut	Nilai Atribut
Buying	v-high (very high), high, med, low
Maint	v-high (very high), high, med, low
Doors	2, 3, 4, 5-more (5 or more)
Persons	2, 4, more (4 or more)
Lug_Boot	small, med, big
Safety	low, med, high

Data set disajikan pada Tabel 2.

TABEL II. DATA SET

No	buying	maint	doors	persons	lug boot	safety	quality
1	vhigh	vhigh	2	2	small	low	unacc
35	vhigh	vhigh	3	2	big	med	unacc
36	vhigh	vhigh	3	2	big	high	unacc
37	vhigh	vhigh	3	4	small	low	unacc
38	vhigh	vhigh	3	4	small	med	unacc
39	vhigh	vhigh	3	4	small	high	unacc
...
1728	low	low	5more	more	big	high	vgood

2) Memproses Data

Data yang diperoleh akan digunakan untuk menganalisis data sebagai bahan yang akan diproses untuk megklasifikasi data dengan menggunakan metode klasifikasi C4.5. Data diolah dan diklasifikasikan berdasarkan 6 [1, 1] kriteria yang untuk dihitung nilai Entropy dan gain nya. Kriteria pada Tabel 2.

TABEL III. KRITERIA

Kriteria	Deskripsi	Label
buying	Harga Beli	vhigh,high,med,low.
maint	Biaya Perawatan	vhigh,high,med,low.
doors	Jumlah Pintu	2,3,4,5more.
persons	Daya Tampung Untuk Memuat Orang	2,4,more.
lug_boot	Luas Bagasi	small,med,big.
safety	Estimasi Tingkat Keamanan	low,med,high.

Adapun kelas label terdiri dari 4 label yaitu unacc,acc,good dan vgood dengan distribusi sebagai berikut

TABEL IV. DISTRIBUSI CLASS

Class	N	N[%]
unacc	1210	70,02%
acc	384	22,22%
good	69	3,99%
v-good	65	3,76%

III. HASIL DAN PEMBAHASAN

Permodelan *Decision Tree* menggunakan algoritma C4.5 melalui beberapa tahapan yaitu

menghitung nilai *entropy* dan nilai *gain* dari data training. Setelah didapat nilai *gain* tertinggi maka akan dikonversi ke dalam pohon keputusan. Perhitungan nilai tersebut di representasikan pada Node 1 Tabel 4.

TABEL V. PERHITUNGAN NODE 1

Kriteria	Sub kriteria	Jumlah Kasus	unacc	acc	good	v-good	Entropy	Gain
	Total	1728	1210	384	69	65	1,206	
buying	high	432	108		324		0,811	0,09649
	Low	432	89	46	258	39	1,571	
	Med	432	115	23	268	26	1,405	
	vhigh	432	72		360		0,65	
maint	High	432	105		314	13	0,983	0,0737
	Low	432	92	46	268	26	1,491	
	Med	432	115	23	268	26	1,405	
	vhigh	432	72		360		0,65	
doors	2	432	81	15	326	10	1,053	0,00449
	3	432	99	18	300	15	1,212	
	4	432	102	18	292	20	1,27	
	5more	432	102	18	292	20	1,27	
persons	2	576			576		0	0,219663
	4	576	198	36	312	30	1,481	
	More	576	186	33	322	35	1,478	
lug_boot	big	576	144	24	368	40	1,371	0,030008
	med	576	135	24	392	25	1,256	
	small	576	105	21	450		0,9	
safety	Low	576	204	30	277	65	1,616	0,262184
	med	576			576		0	
	height	576	180	39	357		1,215	

Perhitungan dapat direpresentasikan seluruh kriteria dengan perhitungan:

$$\text{Entropy total} = (\text{Jumlah UNACC} / \text{Jumlah Kasus}) * \text{IMLOG2} (\text{Jumlah UNACC} / \text{Jumlah Kasus}) + (-\text{ACC} / \text{Jumlah Kasus}) * \text{IMLOG2} (\text{Jumlah ACC} / \text{Jumlah Kasus}) + (-\text{GOOD} / \text{Jumlah Kasus}) * \text{IMLOG2} (\text{Jumlah GOOD} / \text{Jumlah Kasus}) + (-\text{VGOOD} / \text{Jumlah Kasus}) * \text{IMLOG2} (\text{Jumlah VGOOD} / \text{Jumlah Kasus})$$

$$(-1210 / 1728) * IMLOG2 (1210 / 1728) + (-384 / 1728) * IMLOG2 (384 / 1728) + (-69 / 1728) * IMLOG2 (69 / 1728) + (-65 / 1728) * IMLOG2 (65 / 1728)$$

= **1.20574097**

Gain kriteria nilai BUYING:

$$(Entropy Total) - (Jumlah Kasus Kriteria high / Jumlah Kasus) * (Entropy high) - ((Jumlah kasus Low/Jumlah kasus) * Entropy low) - ((Jumlah kasus med/Jumlah kasus) * Entropy med) - ((Jumlah kasus vhigh/Jumlah kasus) * Entropy vhigh)$$

$$= (1.20574097001217) - ((432 / 1728) * 0.811278124459133) - ((432 / 1728) * 1.5709721230675) - ((432 / 1728) * 1.40489533419526) - ((432 / 1728) * 0.650022421648355)$$

= **0.096448969**

Gain kriteria nilai MAINT:

$$(Entropy Total) - (Jumlah Kasus Kriteria high / Jumlah Kasus) * (Entropy high) - ((Jumlah kasus Low/Jumlah kasus) * Entropy low) - ((Jumlah kasus med/Jumlah kasus) * Entropy med) - ((Jumlah kasus vhigh/Jumlah kasus) * Entropy vhigh)$$

$$= (1.20574097001217) - ((432 / 1728) * 0.98263655113859) - ((432 / 1728) * 1.49059378538056) - ((432 / 1728) * 1.40489533419526) - ((432 / 1728) * 0.650022421648355)$$

= **0.073703947**

Gain kriteria nilai DOORS:

$$(Entropy Total) - (Jumlah Kasus Kriteria 2/ Jumlah Kasus) * (Entropy 2) - ((Jumlah kasus 3 / Jumlah kasus) * Entropy nilai 3) - (Jumlah Kasus Kriteria 4 / Jumlah Kasus) * (Entropy 4) - (Jumlah Kasus Kriteria 5more / Jumlah Kasus) * (Entropy 5more)$$

$$= (1.20574097001217) - ((432 / 1728) * 1.05341558458016) - ((432 / 1728) * 1.21179972092103) - ((432 / 1728) * 1.26990285402049) - ((432 / 1728) * 1.26990285402049)$$

= **0.004485717**

$$1.26990285402049) - ((432 / 1728) * 1.26990285402049)$$

= **0.004485717**

Gain kriteria nilai PERSONS:

$$(Entropy Total) - (Jumlah Kasus Kriteria 2/ Jumlah Kasus) * (Entropy 2) - ((Jumlah kasus 4/Jumlah kasus) * Entropy 4) - (Jumlah Kasus Kriteria More / Jumlah Kasus) * (Entropy More)$$

$$= (1.20574097001217) - ((576 / 1728) * 0) - ((576 / 1728) * 1.48071993029481) - ((576 / 1728) * 1.47751408972199)$$

= **0.219662963**

Gain kriteria nilai LUG_BOOT:

$$(Entropy Total) - (Jumlah Kasus Kriteria big / Jumlah Kasus) * (Entropy big) - ((Jumlah kasus med / Jumlah kasus) * Entropy med) - (Jumlah Kasus Kriteria small / Jumlah Kasus) * (Entropy small)$$

$$= (1.20574097001217) - ((576 / 1728) * 1.37121627948142) - ((576 / 1728) * 1.25591133953832) - ((576 / 1728) * 0.900070867273973)$$

= **0.030008141**

Gain kriteria nilai SAFETY:

$$(Entropy Total) - (Jumlah Kasus Kriteria big/ Jumlah Kasus) * (Entropy big) - ((Jumlah kasus med/Jumlah kasus) * Entropy med) - (Jumlah Kasus Kriteria small/ Jumlah Kasus) * (Entropy small)$$

$$= (1.20574097001217) - ((576 / 1728) * 1.61551184957961) - ((576 / 1728) * 0) - ((576 / 1728) * 1.21515799079412)$$

= **0.262184357**

Dapat diketahui bahwa nilai Gain dari 6 Atribut tersebut adalah :

TABEL VI. NILAI GAIN DALAM KRITERIA

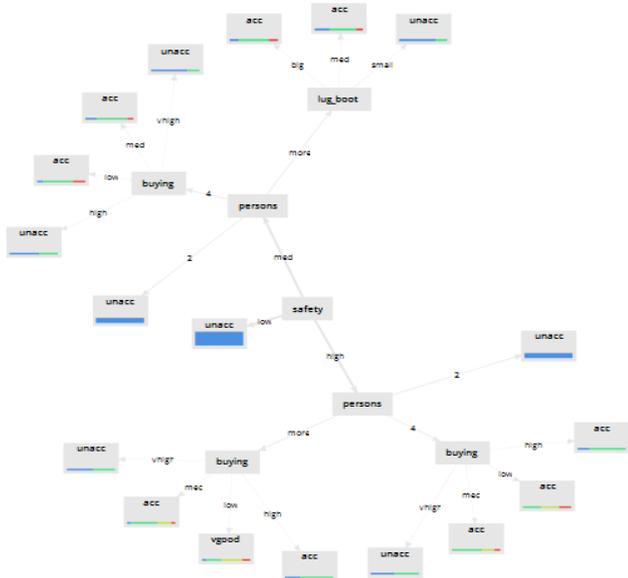
No.	Kriteria	Gain
1	buying	0.096448969
2	maint	0.073703947

No.	Kriteria	Gain
3	doors	0.004485717
4	persons	0.219662963
5	lug_boot	0.030008141
6	safety	0.262184357

Dan nilai *gain* tertinggi adalah kriteria **SAFETY** sebesar **0.262184356554263** yang akan direpresentasikan dalam bentuk pohon keputusan (*decision tree*).

A. Pohon Keputusan (Decision Tree)

Pohon keputusan yang terbentuk dari node 1 direpresentasikan pada Gambar 2 dengan menggunakan software *rapidminer*.



Gbr 1. Pohon Keputusan Hasil Perhitungan Node 1

Selain melalui grafik, pohon keputusan (*Decision Tree*) juga dapat dideskripsikan sebagai berikut:

- safety = high
- | persons = 2: unacc {unacc=192, acc=0, vgood=0, good=0}
- | persons = 4
- | | buying = high: acc {unacc=12, acc=36, vgood=0, good=0}
- | | buying = low: acc {unacc=0, acc=18, vgood=18, good=12}
- | | buying = med: acc {unacc=0, acc=30, vgood=12, good=6}
- | | buying = vhigh: unacc {unacc=24, acc=24, vgood=0, good=0}
- | persons = more

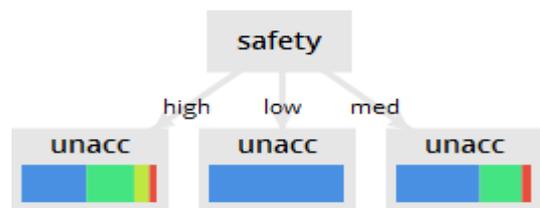
- | | buying = high: acc {unacc=15, acc=33, vgood=0, good=0}
- | | buying = low: vgood {unacc=4, acc=15, vgood=21, good=8}
- | | buying = med: acc {unacc=4, acc=26, vgood=14, good=4}
- | | buying = vhigh: unacc {unacc=26, acc=22, vgood=0, good=0}
- safety = low: unacc {unacc=576, acc=0, vgood=0, good=0}
- safety = med
- | persons = 2: unacc {unacc=192, acc=0, vgood=0, good=0}
- | persons = 4
- | | buying = high: unacc {unacc=30, acc=18, vgood=0, good=0}
- | | buying = low: acc {unacc=6, acc=30, vgood=0, good=12}
- | | buying = med: acc {unacc=12, acc=30, vgood=0, good=6}
- | | buying = vhigh: unacc {unacc=36, acc=12, vgood=0, good=0}
- | persons = more
- | | lug_boot = big: acc {unacc=12, acc=40, vgood=0, good=12}
- | | lug_boot = med: acc {unacc=20, acc=35, vgood=0, good=9}
- | | lug_boot = small: unacc {unacc=49, acc=15, vgood=0, good=0}

Pada kriteria **safety** menghasilkan nilai *gain* tertinggi dengan hasil 0,262184357 yang dihitung dengan pada Tabel 6 node 1.2.

TABEL VII. PERHITUNGAN NODE 1.2

Kriteria	Sub Kriteria	Jumlah Kasus	unacc	acc	good	v-good	Entropy	Gain
safety	Low	576	204	30	277	65	1,616	0,262184
	med	576			576		0	
	height	576	180	39	357		1,215	

Pohon keputusan yang terbentuk dari node 1.2 direpresentasikan pada Gambar 3 menggunakan software *RapidMiner*.



Gbr 2. Pohon Keputusan Hasil Perhitungan Node 1.2

Selain melalui grafik, pohon keputusan (*Decision Tree*) juga dapat dideskripsikan sebagai berikut:

- TREE
- safety = high: unacc {unacc=277, acc=204, vgood=65, good=30}
- safety = low: unacc {unacc=576, acc=0, vgood=0, good=0}
- safety = med: unacc {unacc=357, acc=180, vgood=0, good=39}

B. Pengujian Confusion Matrix

Pengujian data set digunakan 1728 data dengan metode *confusion matrix* yang diuji menggunakan *software* WEKA. Didapatkan matrix dari dataset

Tabel VII. Matrix Dataset

Class	unacc	acc	vgood	Good
unacc	1164	43	0	3
acc	33	333	7	11
vgood	0	3	58	4
good	0	17	11	41

Untuk menghitung nilai akurasi nya. Perhitungan *confusion matrix* pada Tabel 7

TABEL VIII. KLASIFIKASI CONFUSION MATRIX

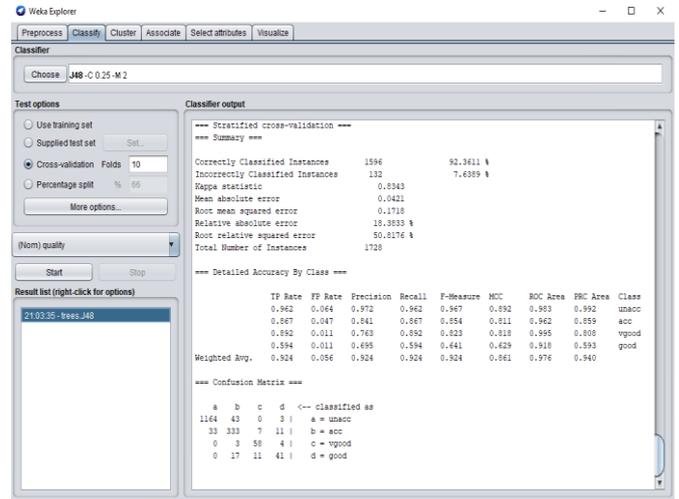
Actual Value		Prediksi Pengujian	
		Positif benar	Negatif berbeda
Actual Value	Positif benar	1596 (TP)	531 (FP)
	Negatif berbeda	0 (FP)	0 (TN)

$$\begin{aligned}
 \text{Akurasi} &= \frac{tp + tn}{tp + tn + fp + fn} \times 100\% \\
 &= \frac{1596 + 0}{1596 + 531 + 0 + 0} \times 100\% = 92,4\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Presisi} &= \frac{tp}{tp + fp} \times 100\% \\
 &= \frac{1596}{1596 + 531} \times 100\% = 92,4\%
 \end{aligned}$$

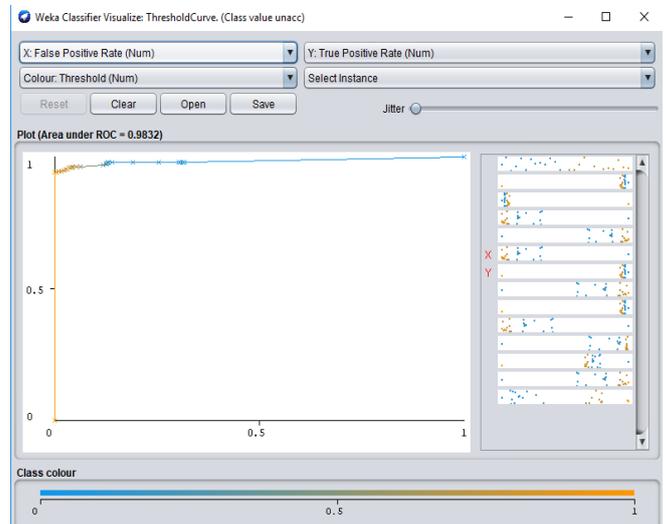
$$\begin{aligned}
 \text{Recall} &= \frac{tp + tn}{fn + tp} \times 100\% \\
 &= \frac{1596}{1596 + 531} \times 100\% = 92,4\%
 \end{aligned}$$

Hasil pengujian *confusion matrix* menggunakan *software* WEKA dengan 10 fold validation dapat dilihat pada Gambar 4.

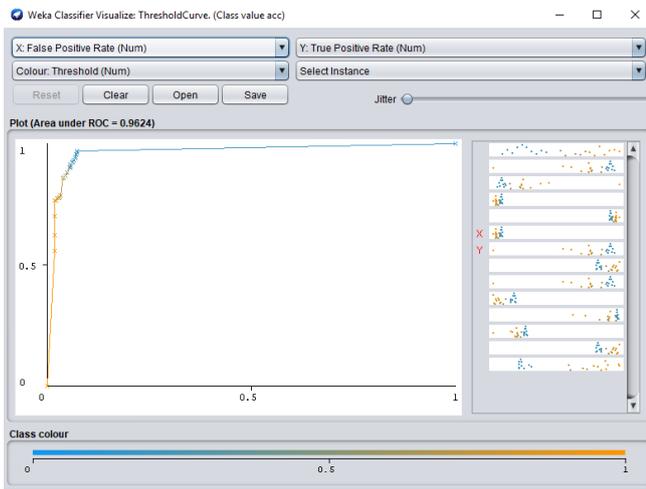


Gbr 3. Hasil Pengujian WEKA

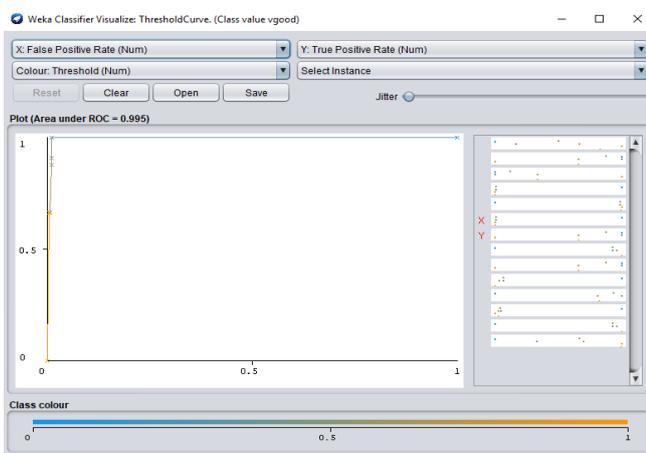
Hasil pengujian juga direpresentasikan dalam bentuk *Area Under Curve* (AUC) pada berikut :



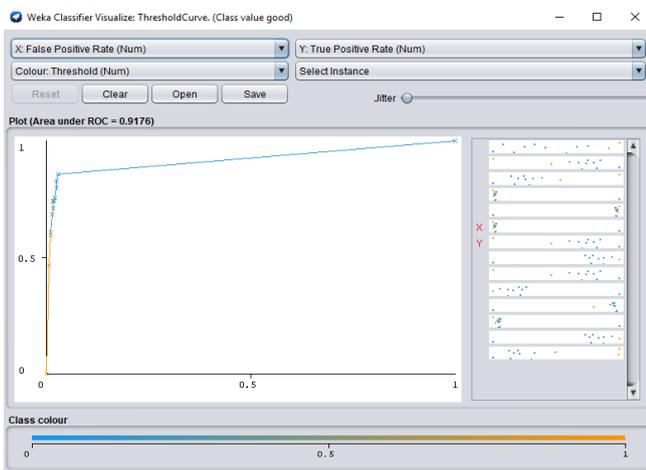
Gbr 4. AUC Class Value UNACC



Gbr 5.AUC Class Value ACC



Gbr 6.AUC Class Value GOOD



Gbr 7.AUC Class Value V-GOOD

IV. KESIMPULLAN

Berdasarkan uraian, penjelasan dan pengujian yang telah dilakukan, didapat beberapa kesimpulan sebagai berikut :

1. Metode algoritma C4.5 dapat memberi kemudahan untuk mengelompokkan mobil berdasarkan kualitasnya dengan metode *decision tree* menggunakan atribut-atribut yang digunakan antara lain *safety, persons, buying, maint, lug_boot, doors*
2. Pemilihan kendaraan mobil yang tepat sesuai kebutuhan dapat menggunakan aplikasi *data mining* RapidMiner menggunakan metode *decision tree* serta algoritma C4.5 dengan tingkat keakuratan dari perhitungan dengan metode *confusion matrix* yang sudah dilakukan dengan menggunakan *software* WEKA dengan tingkat akurasi 92.4% , Precision 92.4% dan Recall 92.4%.

V. UCAPAN TERIMA KASIH

Kami mengucapkan terima kasih yang sebesar-besarnya kepada pihak yang telah membantu sehingga tulisan ini dapat diselesaikan. Terutama kami mengucapkan banyak terima kasih kepada dosen mata kuliah Machine Learning yaitu Bapak Dr.Yan Rianto M.Eng yang telah banyak memberikan ilmu kepada kami terkait tentang Data Mining dan Machine Learning.

REFERENSI

- [1] J. Larkin, *M&A activity to drive fundamental changes in global automotive industry in 2010*, vol. 190, no. 5. 2010.
- [2] E. Elisa, “Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti,” *J. Online Inform.*, vol. 2, no. 1, p. 36, 2017.
- [3] R. Wajhillah, S. Nusa, and M. Sukabumi, “Optimasi Algoritma Klasifikasi C4 . 5 Berbasis Particle Swarm Optimization,” no. May, 2016.
- [4] G. Dimitoglou, J. a Adams, and C. M. Jim, “Comparison of the C4.5 and a Naive Bayes Classifier for the Prediction of Lung Cancer Survivability,” *J. Neural Comput.*, vol. 4, no. 8, pp. 1–9, 2012.
- [5] M. Berliana, M. Gito, and M. Agus, “Perbandingan Algoritma Naive Bayes dan C4.5 Dalam Klasifikasi Data Mining,” vol. 5, no. 4, pp. 455–464, 2018.
- [6] Y. S. Luvia, A. P. Windarto, S. Solikhun, and D. Hartama, “Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Keberhasilan Mahasiswa Di Amik Tunas Bangsa,” *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 1, no. 1, p. 75, 2018.

- [7] <https://archive.ics.uci.edu/ml/datasets/car+evaluation>,
“Document.pdf.”
- [8] B. Zupan, M. Bohanec, I. Bratko, and J. Demsar,
“Machine learning by function decomposition,” *Icml*,
pp. 421–429, 1997.
- [9] M. Bohanec and V. Rajkovic, “Knowledge acquisition
and explanation for multi-attribute decision making.,”
vol. 1, no. 1, pp. 1–19, 1988.