

Naskah Publikasi-Rieke Reza Andarista-18.01.55.0018- 09122021 *by Tete Hayati*

Submission date: 09-Dec-2021 07:46AM (UTC+0700)

Submission ID: 1724932702

File name: NaskahPublikasi_Rieke_Reza_Andarista_1801550018.pdf (1.07M)

Word count: 5382

Character count: 30883

Penerapan Data Mining Algoritma C4.5 Untuk Klasifikasi Prediksi Hasil Pengujian Kendaraan Bermotor Pada Dinas Perhubungan Kabupaten Rembang

Rieke Reza Andarista^{1,*}, Rieke Reza Andarista¹, Arief Jananto²

^{1,2}Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Stikubank Semarang

Jl. Tri Lomba Juang No 1 Semarang 50241, Kota Semarang, Jawa Tengah

Email : ^{1,*}riekereza222@gmail.com, ²ajananto09@edu.unisbank.ac.id

^{*)} Email Penulis Utama

Abstrak - Pengujian Kendaraan Bermotor Dinas Perhubungan merupakan bagian dari bidang sarana dan prasarana yang mempunyai tugas pokok dan fungsi memberikan pelayanan pada masyarakat yaitu dalam melakukan pengujian kendaraan bermotor yang meliputi proses pendaftaran, proses pembayaran atau retribusi, proses pemeriksaan teknis. Namun pada kenyataannya masih terdapat kendaraan yang belum laik jalan dan tidak lulus uji. Hal ini disebabkan kurangnya perawatan kendaraan secara rutin sehingga menyebabkan kendaraan rusak, selain itu usia kendaraan yang diatas sepuluh tahun juga rentan mengalami kerusakan jika tidak dilakukan perawatan rutin terhadap kendaraan. kerusakan tersebut meliputi tebal tidaknya asap pada gas buang kendaraan, ban kendaraan jarang dirawat sehingga ban tersebut gundul dan tipis, tingkat kebisingan klakson yang tidak memenuhi standart pabrik akan membahayakan pengguna jalan, selain itu lampu kendaraan harus berfungsi dengan baik dan berfungsinya rem pada kendaraan. Kondisi ini dapat mempengaruhi pada saat melakukan pengujian kendaraan bermotor pada proses pemeriksaan teknis yang dapat menyebabkan kendaraan tidak lulus uji. Kendaraan yang tidak lulus uji disebabkan karena beberapa faktor. Untuk mengelompokan faktor yang mempengaruhi kendaraan bermotor tidak lulus uji menggunakan metode klasifikasi dengan algoritma C4.5 yang diharapkan dapat membantu mengetahui prediksi hasil pengujian kendaraan bermotor yang dilihat dari faktor yang mempengaruhinya. Penelitian ini menggunakan 428 record. Hasil dari penelitian ini dengan menggunakan perhitungan manual dan tools Rapidminer. Hasil perhitungan manual dengan menggunakan 100 record yang akan dibagi menjadi 0,8 untuk data training dan 0,2 untuk data testing menghasilkan pohon keputusan Kedalaman Alur Ban sebagai root node dengan nilai gain sebesar 0,325 dan tingkat keakurasian sebesar 95,00% menghasilkan 6 rule/aturan. Dari perhitungan manual dan menggunakan tools Rapidminer menghasilkan nilai akurasi yang sama dan pohon keputusan yang sama. Hasil implementasi algoritma C4.5 dengan jumlah 424 record yang akan dibagi menjadi tiga kali percobaan dengan rasio 70:30 menghasilkan nilai akurasi 96,06% dan 80:20 menghasilkan nilai akurasi 94,12% sedangkan 90:10 menghasilkan nilai akurasi 95,24%. Yang akan digunakan yaitu 70:30 dengan nilai akurasi tertinggi 96,06% menghasilkan pohon keputusan dengan efisiensi rem sebagai root node dan menghasilkan 14 rule/aturan.

Kata Kunci: Pengujian Kendaraan Bermotor, Data Mining, Klasifikasi, Rapidminer, Algoritma C4.5.

Abstract - Testing of Motorized Vehicles The Department of Transportation is part of the field of facilities and infrastructure which has the main task and function of providing services to the community, namely in conducting motor vehicle testing which includes the registration process, payment or retribution processes, technical inspection processes. But in reality there are still vehicles that are not roadworthy and do not pass the test. This is due to the lack of routine vehicle maintenance, which causes the vehicle to be damaged, in addition to the age of the vehicle above ten years, it is also susceptible to damage if not carried out routine maintenance of the vehicle. The damage includes thick or not smoke in vehicle exhaust, vehicle tires are rarely cared for so that the tires are bald and thin, horn noise levels that do not meet factory standards will endanger road users, besides vehicle lights must function properly and the brakes on the vehicle function properly. This condition can affect when testing motorized vehicles on the technical inspection process which can cause the vehicle to not pass the test. Vehicles that do not pass the test are caused by several factors. To classify the factors that affect motor vehicles that do not pass the test, use the classification method with the C4.5 algorithm which is expected to help determine the predictions of motorized vehicle test results seen from the factors that influence it. This study uses 428 records. The results of this study using manual calculations and tools Rapidminer. The results of manual calculations using 100 records which will be divided into 0.8 for training data and 0.2 for testing data produce a Tire Groove Depth decision tree as the root node with a gain value of 0.325 and an accuracy level of 95.00% resulting in 6 rules/ rule. From manual calculations and using Rapidminer tools, it produces the same accuracy value and the same decision tree. The results of the implementation of the C4.5 algorithm with a total of 424 records that have been prepared which will be divided into three trials with a ratio of 70:30 producing an accuracy value of 96.06% and 80:20 producing an accuracy value of 94.12% while 90:10 producing an accuracy value 95.24%. What will be used is 70:30 with the highest accuracy value of 96.06% producing a decision tree with brake efficiency as the root node and producing 14 rules.

Keywords: Motor Vehicle Testing, Data Mining, Classification, Rapidminer, C4.5 Algorithm.

1. PENDAHULUAN

1.1 Latar Belakang Masalah

Dinas Perhubungan Kabupaten Rembang merupakan unsur pelaksana urusan Pemerintahan di bidang Perhubungan Kabupaten Rembang. Dinas Perhubungan mempunyai tugas membantu Bupati dalam melaksanakan urusan pemerintahan bidang Perhubungan yang menjadi kewenangan daerah dan tugas pembantuan yang ditugaskan kepada daerah [1]. Pada Dinas Perhubungan terdapat layanan berupa layanan pengujian kendaraan bermotor. Pengujian Kendaraan Bermotor merupakan pemeriksaan kondisi kendaraan apakah kendaraan tersebut memenuhi persyaratan teknis dan laik jalan atau tidak. Namun pada kenyataannya masih terdapat kendaraan yang tidak lulus uji dan tidak laik jalan. Hal ini disebabkan kurangnya perawatan kendaraan secara rutin sehingga menyebabkan kendaraan rusak, selain itu usia kendaraan yang diatas sepuluh tahun juga rentan mengalami kerusakan jika tidak dilakukan perawatan rutin terhadap kendaraan. kerusakan tersebut meliputi tebal tidaknya asap pada gas buang kendaraan, ban kendaraan jarang dirawat sehingga ban tersebut gundul dan tipis, tingkat kebisingan klakson yang tidak memenuhi standart pabrik akan membahayakan pengguna jalan, selain itu lampu kendaraan harus berfungsi dengan baik dan berfungsinya rem pada kendaraan. Kondisi ini dapat mempengaruhi pada saat melakukan pengujian kendaraan bermotor pada proses pemeriksaan teknis yang dapat menyebabkan kendaraan tidak lulus uji karena beberapa faktor. Untuk mengelompokkan faktor yang mempengaruhi kendaraan bermotor tidak lulus uji menggunakan metode klasifikasi dengan algoritma C4.5 yang diharapkan dapat membantu mengetahui prediksi hasil pengujian kendaraan bermotor berdasarkan faktor yang mempengaruhinya. Adapun untuk tools yang digunakan agar lebih mudah dalam mengimplementasikan dan memberikan gambaran yang jelas dari hasil penelitian, peneliti menggunakan Rapidminer.

Penelitian tentang Penerapan Data Mining dengan Algoritma Iterative Dichotomiser Three (ID3) Untuk Klasifikasi Hasil Uji Kelayakan Kendaraan Bermotor menggunakan metode klasifikasi dengan algoritma ID3 dengan data berjumlah 15 menghasilkan rule dan sistem yang lebih efektif dan efisien. [2]. Penelitian yang dilakukan oleh I. Budiman and R. Ramadina yaitu penerapan data mining klasifikasi dengan algoritma C4.5 menghasilkan informasi memprediksi masa studi tepat waktu mahasiswa di program studi Ilmu Komputer FMIPA UNLAM, dengan tingkat akurasi yang dihasilkan sebesar 73,33% dengan jumlah 30 data testing dari 140 data set. [3]. Penelitian tentang implementasi Klasifikasi Uji Berkala Kelayakan Kendaraan Bermotor Menggunakan Metode Decision Tree Dengan Algoritma (ID3) menggunakan JAVA untuk membangun sistem menghasilkan akurasi dari 2 skenario uji coba didapatkan akurasi terbaik sebesar 100% pada skenario pengujian 1 dan 84% pada skenario pengujian 2. [4]. Penelitian yang dilakukan oleh A. Purwanto, A. Primajaya, and A. Voutama penerapan algoritma C4.5 menggunakan weka menghasilkan label atribut yang menyatakan tingkatan kasus penumonia Rendah, Sedang menghasilkan pohon keputusan yang terdiri dari 6 rule. [5]

1.2 Tujuan Dan Manfaat Penelitian

Adapun tujuan dari penelitian ini yaitu untuk menganalisis dan menerapkan algoritma C4.5 untuk klasifikasi prediksi hasil pengujian kendaraan bermotor dan mengetahui tingkat keakurasiannya. Manfaat yang didapat dari penelitian ini adalah dapat mengetahui rule atau aturan dan tingkat akurasi klasifikasi prediksi hasil pengujian kendaraan bermotor sebagai bahan pertimbangan dalam pengambilan kebijakan untuk tiap pengujian kendaraan bermotor di Dinas Perhubungan Kabupaten Rembang.

1.3 Landasan Teori

a. Data Mining

Data mining dalam istilah sederhana adalah penemuan pola yang berguna dalam pengolahan data, data mining juga disebut sebagai ilmu pengetahuan, machine learning, dan analisis prediksi [6].

b. Klasifikasi

Klasifikasi merupakan model (fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui [7].

c. Algoritma C4.5.

Algoritma C4.5 menghasilkan beberapa rule dan pohon keputusan dengan tujuan untuk meningkatkan keakuratan dari prediksi yang sedang dilakukan, disamping itu algoritma C4.5 merupakan algoritma yang mudah dimengerti. [8]. Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut [9]

a. Pemilihan atribut sebagai akar

b. Buat cabang untuk tiap-tiap nilai

c. Bagi kasus dalam cabang

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

Dengan atribut yang nilai information gain yang paling tertinggi dipilih sebagai atribut test untuk simul dengan rumus:

$$SplitInfo(S,A) = \sum_{i=1}^c \frac{S_i}{S} \log_2 \frac{S_i}{S} \dots\dots\dots(1)$$

Dimana:

S = Ruang (data) sample yang digunakan untuk training.

A = Atribut.

Si = Jumlah sample untuk atribut i.

Entropy berperan sebagai parameter untuk mengukur varian dari data sampel dengan rumus :

$$Entropy(S) = \sum_{i=1}^n - P_i * \log_2 P_i \dots\dots\dots(2)$$

Keterangan :

S : himpunan kasus

n : jumlah partisi S

Pi : proporsi dari Si terhadap S

Notasi information gain adalah Gain (S.A) yang berarti dalam data atribut A relative terhadap output

$$Gain(S,A) = Entropy(S) - \sum_i^n \frac{|S_i|}{|S|} * Entropy(S_i) \dots\dots\dots(3)$$

Keterangan :

S : himpunan kasus

A : atribut

N : jumlah partisi atribut A

ISi : jumlah kasus pada partisi ke-i

IS : jumlah kasus dalam S

d. Confusion Matrix

Rumus untuk menghitung tingkat akurasi, precision dan recall pada matriks adalah [10] :

$$accuracy = \frac{tp+tn}{tp+tn+fp+fn} * 100\% \dots\dots\dots(4)$$

keterangan

tp : banyaknya data dengan nilai sebenarnya positif dan nilai prediksi positif (True Positif).

tn : banyaknya data dengan nilai sebenarnya negatif dan nilai prediksi negatif (True Negatif).

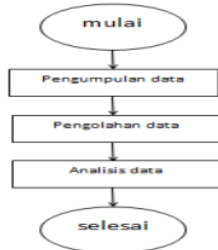
fp : banyaknya data dengan nilai sebenarnya negatif dan nilai prediksi positif (False Positif).

fn : banyaknya data dengan nilai sebenarnya positif dan nilai prediksi negatif (False Negatif).

2. METODE PENELITIAN

1.4 Tahapan Penelitian

Tahapan dari metode penelitian yang dilakukan oleh peneliti akan disajikan pada gambar 1 [11].



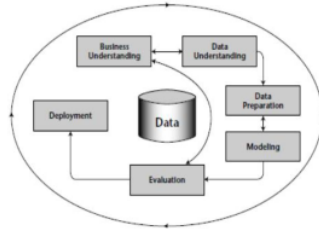
Gambar 1. Tahapan Penelitian

Berikut penjelasan gambar 1

1. pengumpulan data, dengan teknik penyebaran kuesioner terhadap 428 pemohon yang akan melakukan pengujian kendaraan bermotor pada Dinas Perhubungan Kabupaten Rembang menggunakan *google form*.
2. Pengolahan data, data yang didapatkan masih berupa data mentah dan masih terdapat beberapa data yang tidak dibutuhkan oleh karena itu harus dilakukan proses preparation agar data siap untuk diminingkan.
3. Analisa data, data yang telah diolah lalu dilakukan analisa dengan menggunakan metode CRISP-DM.

1.5 Metode Analisa Data

Metode analisis data mining pada penelitian ini menggunakan model CRISP – DM[12]



Gambar 2. Tahapan CRISP-DM.

Tahap – tahap CRISP – DM sebagai berikut:

1. Fase Pemahaman Bisnis

Tujuan dari penelitian ini adalah untuk menerapkan metode data mining menggunakan Algoritma C4.5 Decision Tree untuk mengetahui klasifikasi prediksi hasil pengujian kendaraan bermotor dengan seakurat mungkin Sehingga dapat diketahui suatu pola – pola atau aturan – aturan dari pohon keputusan yang dihasilkan dapat digunakan untuk pengambilan kebijakan pada masa yang akan datang.

2. Fase Pemahaman Data

Pada Fase pemahaman data ini dilakukan pengumpulan data menggunakan kuesioner untuk mengumpulkan informasi dari responden yang berjumlah 428 record.

3. Fase Persiapan Data

Berikut tahap– tahap dalam persiapan data :

a. Tahap Pemilihan Atribut

Pada tahap ini akan dilakukan pemilihan atribut karena pada dataset awal tidak semua atribut akan digunakan dalam proses data mining. Berikut adalah tabel dataset yang sudah diseleksi dapat dilihat pada Tabel 1.

Tabel 1. Atribut Hasil Seleksi.

No	Atribut
1.	Tahun pembuatan
2.	Service berkala
3.	Kepemilikan kendaraan
4.	Jenis pelayanan
5.	Emisi gas buang
6.	Speedometer
7.	Efisiensi rem
8.	Tingkat kebisingan klakson
9.	Kedalaman alur ban
10.	Kemampuan pancar lampu utama
11.	Kaca film kendaraan
12.	Hasil

b. tahap pembersihan data pada penelitian ini dilakukan pembersihan data pada data hasil kuesioner yang semula berjumlah 428 record menjadi 424 record.

c. Mendiskritisasi Atribut yang akan di discretize adalah tahun pembuatan kendaraan menjadi <2010 TAHUN dan >2010 TAHUN dapat dilihat pada tabel 2.

Tabel 2. atribut hasil discretize

Tahun	Keterangan
1980-2010	<10 Tahun
2010-2021	>10 Tahun

d. Menentukan Atribut Target yang akan menjadi atribut target adalah Hasil.

e. Split data

Pada perhitungan manual menggunakan 100 record dengan rasio 0,8 untuk data training dan 0,2 untuk data testing kemudian diimplementasikan pada Rapidminer dan pada implementasi

pada Rapidminer dengan jumlah 424 record dilakukan percobaan tiga kali yaitu 70:30, 80:20, 90:10 yang akan digunakan yaitu nilai akurasi tertinggi .

4. Fase Pemodelan

Dalam tahap pemodelan ini menggunakan teknik data mining dengan metode klasifikasi dan menggunakan algoritma C4.5 Decision Tree dengan menggunakan tools Rapidminer.

a. Flowchart analisa data mining menggunakan algoritma C4.5 dapat dilihat pada Gambar 3.



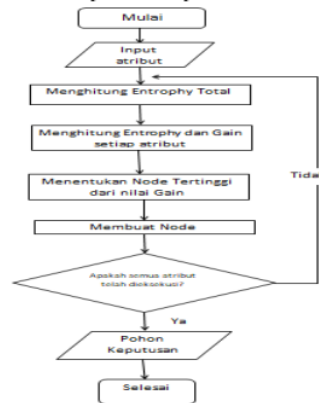
Gambar 3. Flowchart Analisa Data Mining

Berikut Penjelasan *Flowchart* Analisa Data Mining

1. Input data, Menginput data Hasil Kuesioner Pengujian Kendaraan Bermotor Pada Dinas Perhubungan Kabupaten Rembang.
2. Data preprosesing, Proses Mempersiapkan data , terdiri dari beberapa langkah seperti menentukan type data, pemilihan atribut, pembersihan data, mendiskritisasi atribut, set role, split data.
3. Data training, Hasil dari proses preprosesing menjadi data training untuk data latih.
4. Modeling, Proses modeling analisis data yang telah diimplementasikan ke algoritma C4.5.
5. Data testing, Hasil dari proses preprosesing menjadi data testing untuk data uji.
6. Validasi, Proses pengujian menggunakan data uji menggunakan apply model dan performance.
7. Hasil, Proses pengujian akurasi dan confusion matrix
8. Kesimpulan, Output ini akan menghasilkan sebuah kesimpulan berupa rule atau aturan dengan mengetahui keakuratan hasil.

b. Flowchart Algoritma C4.5

Berikut adalah flowchart algoritma C4.5 dapat dilihat pada Gambar 4. [13]



Gambar 4. Flowchart Algoritma C4.5.

Berikut Penjelasan *Flowchart* Algoritma C4.5

1. Input atribut, Menginput atribut yang sudah dipreprosesing yang siap untuk dimining.
2. Menghitung entropy total dengan rumus persamaan (1) dan (2).
3. Menghitung entropy dan gain setiap atribut dengan rumus persamaan (3).

	KEDALAMAN ALUR BAN	ALUR BAN > 1 MM	3	9	12	0.81		
		ALUR BAN < 1 MM	67	1	68	0.11		
							0.215	0.325

	KEMAMPUAN PANCAR LAMPU UTAMA	BAIK	61	0	61	0.00		
		BURUK	9	10	19	1.00		
							0.2375	0.30

Pada Tabel 3 Entropy total atribut hasil adalah 0,54 dengan jumlah kasus 80 record yaitu 70 “LULUS” dan 10 “TIDAK LULUS”. Dari hasil perhitungan atribut dengan nilai information gain sebagai nilai terbesar adalah 0,325 yaitu atribut kedalaman alur ban. Selanjutnya atribut kedalaman alur ban dijadikan sebagai root node (akar). Berikut bentuk pohon keputusan root node dapat dilihat pada Gambar 5.



Gambar 5. Pohon Keputusan Root Node

Pada Gambar 5. merupakan pohon keputusan bagian node akar. Atribut yang gainnya terbesar yang menjadi root adalah kedalaman alur ban yang menghasilkan dua node yaitu pada node pertama menghasilkan alur ban > 1 mm dengan jumlah kasus 68 record terdiri dari 67 “LULUS” dan 1 “TIDAK LULUS) sedangkan node kedua menghasilkan node alur ban < 1 mm dengan jumlah kasus 12 terdiri dari 3 “LULUS” dan 9 “TIDAK LULUS”.

2. Perhitungan Node 1

Pada perhitungan node pertama yaitu alur ban > 1 mm menghasilkan 67 “LULUS” dan 1 “TIDAK LULUS”. Atribut yang akan dihitung adalah tahun, service berkala kendaraan, kepemilikan kendaraan, jenis pelayanan, emisi gas buang, speedometer, efisiensi rem, tingkat kebisingan klakson, kemampuan pancar lampu utama, dan kaca film kendaraan. Berikut adalah contoh perhitungan manual yang diambil yaitu atribut speedometer. Hasil perhitungan dapat dilihat pada Tabel 4.

SPEEDOMETER	SESUAI	LULUS	65	ENTROPY(SPEEDOMETER)=(65+0)/68*INFO(65,0)+(2+1)/68*INFO(2,1)	
		TIDAK LULUS	0	ENTROPY(SPEEDOMETER)=65/68*0+3/68*0.92	
	TIDAK SESUAI	LULUS	2	ENTROPY(SPEEDOMETER)	0.040588
		TIDAK LULUS	1	GAIN(SPEEDOMETER)	0.069412
			68	SPLIT INFO(65,0)	0
				SPLIT INFO(2,1)	0.92

Tabel 4. Hasil Perhitungan Node 1

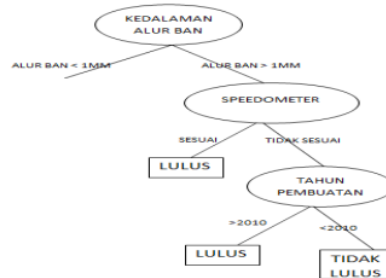
NODE	ATRIBUT	VALUE	LABEL		JUMLAH KASUS	SPLIT INFO	ENTROPY	GAIN
			LULUS	TIDAK LULUS				
NODE 1	KEDALAMAN ALUR BAN = ALUR BAN > 1 MM		67	1	68	0.11		
	TAHUN	>2010	55	0	55	0.00		
		<2010	12	1	12	0.30		
							0.074559	0.035441

	SPEEDOMETER	SESUAI	65	0	65	0.00		
		TIDAK SESUAI	2	1	3	0.92		
							0.040588	0.069412

	TINGKAT KEBISINGAN KLAKSON	STANDART PABRIK	59	1	60	0.12		

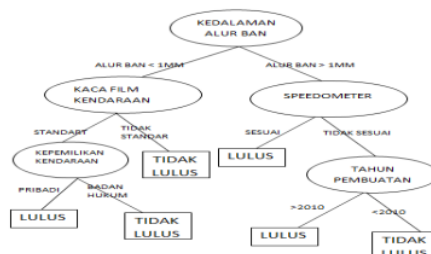
KEMAMPUAN PANCAR LAMPU UTAMA	BAIK	1	0	1	0		
	BURUK	1	1	2	1.00		

Pada Tabel 5. perhitungan pohon keputusan node 1.2 menghasilkan gain terbesar yaitu atribut tahun sebesar 0.92 yang mempunyai dua node yaitu node <2010 dan >2010. Node <2010 dengan jumlah kasus 0 "LULUS" Dan 1 "TIDAK LULUS" maka node <2010 menghasilkan keputusan "TIDAK LULUS" sedangkan node >2010 dengan jumlah kasus 2 "LULUS" dan 0 "TIDAK LULUS" maka node >2010 menghasilkan keputusan "LULUS". Bentuk pohon keputusan dapat dilihat pada Gambar 7.



Gambar 7. Pohon Keputusan Node 1.2

Pada Gambar 7. perhitungan root node kedalaman alur ban > 1 mm menghasilkan node speedometer yang memiliki dua node yaitu sesuai dan tidak sesuai. Node sesuai menghasilkan keputusan "LULUS" dan Node tidak sesuai menghasilkan node tahun pembuatan memiliki dua node yaitu node >2010 dan node <2010. Node >2010 menghasilkan keputusan "LULUS" dan node <2010 menghasilkan keputusan "TIDAK LULUS". Pada perhitungan node 2 yaitu alur ban < 1 mm jumlah kasus 12 terdiri dari 3 "LULUS" dan 9 "TIDAK LULUS". Atribut yang akan dihitung adalah tahun, service berkala kendaraan, kepemilikan kendaraan, jenis pelayanan, emisi gas buang, speedometer, efisiensi rem, tingkat kebisingan klakson, kemampuan pancar lampu utama, dan kaca film kendaraan. Menghasilkan gain terbesar adalah atribut kaca film kendaraan sebesar 0.54 yang menghasilkan dua node, node pertama yaitu tidak standart dengan jumlah kasus 0 "LULUS" dan 9 "TIDAK LULUS" maka node tidak standart menghasilkan keputusan "TIDAK LULUS" sedangkan node kedua yaitu standart dengan jumlah kasus 3 "LULUS" dan 1 "TIDAK LULUS" maka node standart menghasilkan node 2.1. Selanjutnya akan melakukan perhitungan pada node 2.1 yaitu standart dengan jumlah kasus 4 terdiri dari 3 "LULUS" dan 1 "TIDAK LULUS". Atribut yang akan dihitung adalah tahun, service berkala kendaraan, kepemilikan kendaraan, jenis pelayanan, emisi gas buang, efisiensi rem, tingkat kebisingan klakson, kepemilikan pribadi. Menghasilkan gain terbesar yaitu atribut kepemilikan kendaraan sebesar 0.81 yang mempunyai dua node yaitu node badan hukum dan pribadi. Node badan hukum dengan jumlah kasus 0 "LULUS" dan 1 "TIDAK LULUS" maka node badan hukum menghasilkan keputusan "TIDAK LULUS" sedangkan node pribadi dengan jumlah kasus 3 "LULUS" dan 0 "TIDAK LULUS" maka node pribadi menghasilkan keputusan "LULUS". Bentuk pohon keputusan dapat dilihat pada Gambar 8.



Gambar 8. Pohon Keputusan Node 2.2

Berdasarkan hasil pohon keputusan pada gambar 8 ,maka dapat disimpulkan beberapa kondisi atau rule yang terjadi, yaitu

- Jika kedalaman alur ban > 1mm, speedometer sesuai maka lulus.
- Jika kedalaman alur ban > 1mm, speedometer tidak sesuai dan tahun pembuatan kendaraan >2010 maka lulus.
- Jika kedalaman alur ban > 1mm, speedometer tidak sesuai dan tahun pembuatan kendaraan <2010 maka tidak lulus.

- d. Jika kedalaman alur ban < 1mm, kaca film kendaraan tidak standart maka tidak lulus.
 - e. Jika kedalaman alur ban < 1mm, kaca film kendaraan standart, kepemilikan kendaraan badan hukum maka tidak lulus.
 - f. Jika kedalaman alur ban < 1mm, kaca film kendaraan standart, dan kepemilikan kendaraan pribadi maka lulus.
3. Confusion Matrix

Untuk menghitung akurasi menggunakan rumus persamaan (4)

Tabel 6. Confusion Matrix

	true TIDAK LULUS	true LULUS
PRED.TIDAK LULUS	2	1
PRED.LULUS	0	17

Pada Tabel 6 menunjukkan hasil uji coba yang dilakukan dengan data testing dengan jumlah 20 record menghasilkan jumlah kelas tidak lulus yang di prediksi benar sebanyak 2 record, yang diprediksi salah sebanyak 1 record. Kelas lulus yang diprediksi benar 17 record, yang diprediksi salah sebanyak 0 record menghasilkan nilai akurasi sebesar 95,00%. Berikut hasil perhitungan dengan menggunakan rumus akurasi

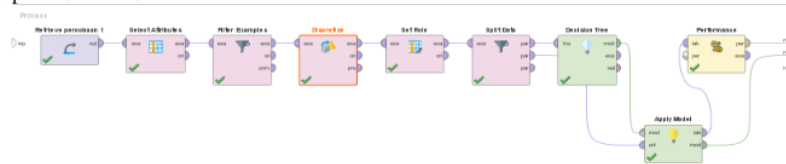
$$\begin{aligned}
 \text{accuracy} &= \frac{tp+tn}{tp+tn+fp+fn} \times 100\% \\
 &= \frac{17+2}{17+2+0+1} \times 100\% \\
 &= \frac{19}{20} \times 100\% \\
 &= 95,00\%
 \end{aligned}$$

3.2 Implementasi Algoritma C4.5 pada Rapidminer

Implementasi algoritma C4.5 pada Rapidminer memiliki dua tahap percobaan. Pertama implementasi perhitungan manual pada Rapidminer menggunakan data berjumlah 100 record yang kemudian dibagi menjadi 80 untuk data training dan 20 untuk data testing. Dan kedua implementasi algoritma C4.5 menggunakan 424 record yang kemudian dibagi menjadi 80 untuk data training dan 20 untuk data testing.

1. Implementasi Algoritma C4.5 Perhitungan Manual Pada Rapidminer.

Percobaan pertama menggunakan 100 record yang telah dipreparation untuk proses pengolahan data menggunakan algoritma C4.5 untuk mendapatkan pohon keputusan dan rule pada rapidminer dapat dilihat pada Gambar 9.



Gambar 9. Pengujian Algoritma C4.5 pada Rapidminer.

Menghasilkan pohon keputusan dan hasil akurasi pengujian pada Gambar 8. menghasilkan sebagai berikut.

a. Pohon Keputusan.

Dari hasil pengujian pada rapidminer data training yang dihubungkan pada operator modeling decision tree diperoleh pohon keputusan seperti gambar 10.



Gambar 10. Pohon Keputusan Percobaan Pertama
 Berikut keterangan pohon keputusan pada rapidminer dapat dilihat pada gambar 11.

Tree

```

KEDALAMAN ALUR BAN = ALUR BAN < 1 MM
|
| KACA FILM KENDARAAN = STANDART
| |
| | KEPEMILIKAN KENDARAAN = BADAN HUKUM: TIDAK LULUS (TIDAK LULUS=1, LULUS=0)
| | KEPEMILIKAN KENDARAAN = PRIBADI: LULUS (TIDAK LULUS=0, LULUS=3)
| KACA FILM KENDARAAN = TIDAK STANDART: TIDAK LULUS (TIDAK LULUS=0, LULUS=0)
KEDALAMAN ALUR BAN = ALUR BAN > 1 MM
|
| SPEEDOMETER = SESUAI: LULUS (TIDAK LULUS=0, LULUS=65)
| SPEEDOMETER = TIDAK SESUAI
| |
| | TAHUN PEMBUATAN KENDARAAN = [-∞ - 2010.0]: TIDAK LULUS (TIDAK LULUS=1, LULUS=0)
| | TAHUN PEMBUATAN KENDARAAN = [2010.0 - ∞]: LULUS (TIDAK LULUS=0, LULUS=2)
    
```

Gambar 11. Keterangan Pohon Keputusan

- b. Hasil Pengukuran Akurasi
 Hasil akurasi menggunakan tools Rapidminer dapat dilihat pada gambar 12.

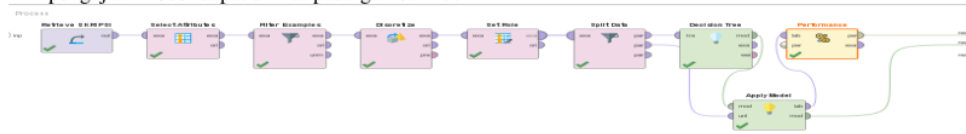
accuracy: 95.00%

	true TIDAK LULUS	true LULUS	class precision
pred. TIDAK LULUS	2	1	66.67%
pred. LULUS	0	17	100.00%
class recall	100.00%	94.44%	

Gambar 12. Hasil Accuracy

Pada gambar 12 Accuracy data testing dengan jumlah 20 record menghasilkan jumlah kelas tidak lulus yang di prediksi benar sebanyak 2 record, yang diprediksi salah sebanyak 1 record. Kelas lulus yang diprediksi benar 17 record, yang diprediksi salah sebanyak 0 record menghasilkan nilai akurasi sebesar 95.00%.

- 2. Implementasi Algoritma C4.5 Pada Rapidminer.
 Pengujian ini digunakan 424 data untuk mengetahui pohon keputusan, nilai accuracy. Untuk pengujian model dapat dilihat pada gambar 13.



Gambar 13. Proses Pengujian Pada Rapidminer.

- a. Pohon Keputusan.
 Dari hasil pengujian pada rapidminer diperoleh pohon keputusan seperti gambar dibawah ini dapat dilihat pada gambar 14.



Gambar 14. Hasil Pohon Keputusan.

Berdasarkan hasil pohon keputusan diatas ,maka dapat disimpulkan beberapa kondisi atau rule yang terjadi, yaitu

1. Jika efisiensi rem lebih dari 50% dan kedalaman alur ban >1 mm maka lulus.
 2. Jika efisiensi rem lebih dari 50%, kedalaman alur ban <1 mm dan tahun pembuatan >2010 maka lulus.
 3. Jika efisiensi rem lebih dari 50%, kedalaman alur ban <1 mm dan tahun pembuatan <2010 dan kemampuan pancar lampu utama buruk maka tidak lulus.
 4. Jika efisiensi rem lebih dari 50%, kedalaman alur ban < 1mm dan tahun pembuatan <2010 dan kemampuan pancar lampu utama baik maka lulus.
 5. Jika efisiensi rem kurang dari 50%, kaca film kendaraan standart dan kepemilikan kendaraan pribadi maka lulus.
-
12. Jika efisiensi rem kurang dari 50%, kaca film kendaraan tidak standart, speedometer tidak sesuai dan kemampuan pancar lampu utama buruk maka tidak lulus.
 13. Jika efisiensi rem kurang dari 50%, kaca film kendaraan tidak standart, speedometer tidak sesuai, kemampuan pancar lampu utama baik dan service berkala rutin maka tidak lulus.
 14. Jika efisiensi rem kurang dari 50%, kaca film kendaraan tidak standart, speedometer tidak sesuai, kemampuan pancar lampu utama baik, service berkala tidak rutin maka tidak lulus.

Berikut keterangan pohon keputusan pada Rapidminer dapat dilihat pada gambar 15.

Tree

```

EFISIENSI REM = KURANG DARI 50%
| KACA FILM KENDARAAN = STANDART
| | KEPEMILIKAN KENDARAAN = BADAN HUKUM
| | | SERVICE BERKALA KENDARAAN = RUTIN: LULUS (TIDAK LULUS=0, LULUS=2)
| | | SERVICE BERKALA KENDARAAN = TIDAK RUTIN: TIDAK LULUS (TIDAK LULUS=1, LULUS=0)
| | | KEPEMILIKAN KENDARAAN = PRIBADI: LULUS (TIDAK LULUS=0, LULUS=14)
| | KACA FILM KENDARAAN = TIDAK STANDART
| | | SPEEDOMETER = SESUAI
| | | | KEDALAMAN ALUR BAN = ALUR BAN < 1 MM
| | | | | TAHUN PEMBUATAN KENDARAAN = [-∞ - 2010.0]: TIDAK LULUS (TIDAK LULUS=3, LULUS=0)
| | | | | TAHUN PEMBUATAN KENDARAAN = [2010.0 - ∞]
| | | | | SERVICE BERKALA KENDARAAN = RUTIN: TIDAK LULUS (TIDAK LULUS=1, LULUS=0)
| | | | | SERVICE BERKALA KENDARAAN = TIDAK RUTIN: LULUS (TIDAK LULUS=0, LULUS=1)
| | | | | KEDALAMAN ALUR BAN = ALUR BAN > 1 MM: LULUS (TIDAK LULUS=0, LULUS=5)
| | | | | SPEEDOMETER = TIDAK SESUAI
| | | | | KEMAMPUAN PANCAR LAMPU UTAMA = BAIK
| | | | | SERVICE BERKALA KENDARAAN = RUTIN: TIDAK LULUS (TIDAK LULUS=1, LULUS=0)
| | | | | SERVICE BERKALA KENDARAAN = TIDAK RUTIN: TIDAK LULUS (TIDAK LULUS=2, LULUS=1)
| | | | | KEMAMPUAN PANCAR LAMPU UTAMA = BURUK: TIDAK LULUS (TIDAK LULUS=18, LULUS=0)
| | | | | KEMAMPUAN PANCAR LAMPU UTAMA = BAIK: LULUS (TIDAK LULUS=0, LULUS=2)
| | | | | KEMAMPUAN PANCAR LAMPU UTAMA = BURUK: TIDAK LULUS (TIDAK LULUS=1, LULUS=0)
| | | | | TAHUN PEMBUATAN KENDARAAN = [2010.0 - ∞]: LULUS (TIDAK LULUS=0, LULUS=4)
| | | | | KEDALAMAN ALUR BAN = ALUR BAN > 1 MM: LULUS (TIDAK LULUS=0, LULUS=241)

```

Gambar 15. Keterangan Pohon Keputusan.

b. Hasil Pengukuran Akurasi

Dari hasil pengujian dengan data testing yang diperoleh dari perhitungan menggunakan RapidMiner dengan confusion matrix dapat dilihat pada gambar 16.

accuracy: 96,06%

	true TIDAK LULUS	true LULUS	class precision
pred. TIDAK LULUS	9	2	81,82%
pred. LULUS	3	113	97,41%
class recall	75,00%	98,26%	

Gambar 16. Hasil Accuracy.

Pada gambar 16. hasil accuracy data testing dengan jumlah 127 record menghasilkan jumlah kelas tidak lulus yang di prediksi benar sebanyak 9 record, yang diprediksi salah sebanyak 2 record. Kelas lulus yang diprediksi benar 113 record, yang diprediksi salah sebanyak 3 record menghasilkan nilai akurasi sebesar 96,06%.

4. KESIMPULAN

Berdasarkan hasil pengujian data hasil kuesioner pengujian kendaraan bermotor menggunakan algoritma C4.5 dapat disimpulkan bahwa :

1. Penerapan data mining telah mendapatkan model klasifikasi menggunakan algoritma C4.5 untuk klasifikasi prediksi hasil pengujian kendaraan bermotor pada Dinas Perhubungan Kabupaten Rembang. Atribut Kedalaman Alur Ban merupakan indikator yang memiliki pengaruh paling besar dalam penelitian ini, hal ini terlihat dari hasil akhir pohon keputusan yang memunculkan Kedalaman Alur Ban node tertinggi sebagai root node.
2. Hasil dari implementasi algoritma C4.5 pada perhitungan manual dan Rapidminer menggunakan 100 record yang dibagi 0.8 untuk data training dan 0.2 untuk data testing dihitung secara manual dan menggunakan tools Rapidminer menghasilkan pohon keputusan Kedalaman Alur Ban yang memiliki node tertinggi sebagai root node dan tingkat keakurasian sebesar 95,00%, Dari perhitungan manual dan menggunakan tools Rapidminer menghasilkan nilai akurasi yang sama dan pohon keputusan yang sama menghasilkan 6 rule/aturan.
3. Hasil dari implementasi algoritma C4.5 Pada Rapidminer dengan menggunakan keseluruhan data yang telah dipreprosesing berjumlah 424 record dengan tiga kali percobaan dengan rasio masing-masing 70:30 menghasilkan akurasi 96,06% dan 80:20 menghasilkan akurasi 94,12% sedangkan 90:10 menghasilkan akurasi 95,24% yang akan digunakan yaitu rasio 70:30 dengan tingkat keakurasiannya tertinggi akan diproses menggunakan tools Rapidminer menghasilkan pohon keputusan Efisiensi rem memiliki node tertinggi sebagai root node dan tingkat keakurasian sebesar 96,06% menghasilkan 14 rule/aturan.

UCAPAN TERIMAKASIH

Terima kasih kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini. Terima kasih kepada Allah SWT yang telah memberikan kelancaran, rahmat dan hidayah-Nya, terima kasih kepada kedua orangtua yang senantiasa memberikan doa dan dukungan, terima kasih kepada dosen pembimbing yang telah membimbing untuk menyelesaikan penelitian ini, terima kasih kepada pihak Dinas Perhubungan Kabupaten Rembang yang telah membantu dan memberikan kesempatan untuk melakukan penelitian. Semoga dengan adanya penelitian ini dapat memberikan wawasan baru kepada penulis dan pembaca.

REFERENCES

- [1] R. Strategis and P. K. Rembang, "Rencana strategis (renstra)," 2021.
- [2] K. W. Saputro, "Penerapan Data Mining dengan Algoritma Iterative Dichotomiser Three (ID3) Untuk Klasifikasi Hasil Uji Kelayakan Kendaraan Bermotor (Studi Kasus : Dinas Perhubungan Kabupaten Nganjuk)," *Simki-techsain*, vol. 1, no. 3, pp. 1-9, 2017.
- [3] I. Budiman and R. Ramadina, "Penerapan Fungsi Data Mining Klasifikasi untuk Prediksi Masa Studi Mahasiswa Tepat Waktu pada Sistem Informasi Akademik Perguruan Tinggi," *Ijccs*, vol. x, No.x, no. 1, pp. 1-5, 2015.
- [4] R. K. D. Mirza Krista Dewayana, Rizal Setya Perdana, "JURNAL_mirza_new.pdf."

- [5] A. Purwanto, A. Primajaya, and A. Voutama, "Penerapan Algoritma C4 . 5 dalam Prediksi Potensi Tingkat Kasus Pneumonia di Kabupaten Karawang," *J. Sist. dan Teknol. Inf.*, vol. 08, no. 4, pp. 390–396, 2020, doi: 10.26418/justin.v8i4.41959.
- [6] I. Amirulloh and Taufiqurrochman, "Komparasi Model Klasifikasi Algoritma Keterlambatan Siswa Masuk Sekolah," *Semin. Nas. Sains dan Teknol. Fak. Tek. Univ. Muhammadiyah Jakarta*, no. November, pp. 1–4, 2017.
- [7] B. Suherman, "Implementasi data mining untuk memprediksi pemasaran produk helm dengan algoritma c4.5 pada pt. indosafety manufacture," 2018.
- [8] M. A. Puspa, "Implementasi Data Mining Klasifikasi Algoritma C4.5 Dalam Perekrutan Perangkat Desa," *Build. Informatics, Technol. Sci.*, vol. 1, no. 2, pp. 92–97, 2019.
- [9] S. Teknik, I. Disusun, and N. Nim, "Skripsi penerapan algoritma c4.5 dalam memprediksi data penjualan bisnis gerai busana muslim," 2019.
- [10] V. M. Magfirah, "Penerapan data mining untuk klasifikasi kepuasan pelanggan transportasi online (ojek online) menggunakan algoritma c4.5," 2018.
- [11] D. Exasanti and A. Jananto, "Analisa Hasil Pengelompokan Wilayah Kejadian Non- Kebakaran Menggunakan Agglomerative Hierarchical Clustering pada Dinas Pemadam Kebakaran Kota Semarang," *J. tekno Kompak*, vol. 15, no. 2, pp. 63–75.
- [12] F. I. Komputer and U. D. Nuswantoro, "PENYAKIT STROKE DENGAN KLASIFIKASI DATA MINING PADA," 2011.
- [13] T. Novika, H. Okprana, A. P. Windarto, and H. Siahaan, "Penerapan Data Mining Klasifikasi Tingkat Pemahaman Siswa Pada Pelajaran Matematika," vol. 5, pp. 9–17, 2021, doi: 10.30865/mib.v5i1.2498.

Naskah Publikasi-Rieke Reza Andarista-18.01.55.0018-09122021

ORIGINALITY REPORT

6%

SIMILARITY INDEX

6%

INTERNET SOURCES

5%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

1

ejurnal.teknokrat.ac.id

Internet Source

6%

Exclude quotes On

Exclude matches < 2%

Exclude bibliography On