

# KLASIFIKASI HARGA MOBIL MENGUNAKAN METODE DECISION TREE ALGORITMA C4.5

Yonathan Anggraiwan<sup>1</sup>, Bakti Siregar<sup>2</sup>

<sup>1</sup> Program Studi Statistika, Fakultas Science, Technology and Mathematics, Universitas Matana, Jl. CBD Barat Kav, RW.1, Curug Sangereng, Kelapa Dua, Tangerang Regency, Banten 15810  
E-mail: [1yonathanggraiwan@gmail.com](mailto:1yonathanggraiwan@gmail.com), [2siregar.bakti@matanauniversity.ac.id](mailto:2siregar.bakti@matanauniversity.ac.id)

## Abstrak

Untuk mempermudah kegiatan sehari-hari, banyak orang yang menggunakan kendaraan bermotor, yaitu motor atau mobil. Seperti yang sudah kita ketahui, kendaraan tersebut memiliki berbagai bagian atau partisi yang menentukan harga dari suatu kendaraan bermotor tersebut. Seperti mesin, jumlah silinder, hemat borosnya penggunaan bahan bakar, dan sebagainya. Pada kesempatan kali ini, peneliti melakukan klasifikasi harga mobil, menggunakan metode decision tree algoritma C4.5 dengan aplikasi RapidMiner. Tujuan dari penelitian ini adalah untuk mengetahui hasil klasifikasi yang tepat terhadap data harga mobil, dimana harga dari mobil tersebut dipengaruhi oleh banyak bagian seperti besar tenaga kuda pada mesin mobil, jenis transmisi, jumlah silinder, dan masih banyak lagi. Dimana setelah diteliti, klasifikasi harga mobil menggunakan metode decision tree algoritma C4.5 dengan aplikasi RapidMiner ini memiliki tingkat akurasi sebesar 82.1%.

**Kata kunci**— Kendaraan Bermotor, Bagian atau Partisi, Klasifikasi, Decision Tree Algoritma C4.5, Aplikasi RapidMiner,

## Abstract

To ease daily activities, many people use vehicles, like motorbike or car. As we know, the vehicles have various parts that impact the price of a vehicles. Like the machine, the number of cylinders, efficient or wasteful use of fuel, and many more. This time, I do classification car price, using the decision tree method with the C4.5 algorithm in RapidMiner application. The purpose of this research is to find out the exact result of classification car price, where is the price of the car is impacted by many parts of this car, like the horsepower of the car engine, the type of transmission, the number of cylinders, and many more. Where after research, the classification of car prices using the decision tree algorithm C4.5 method that I do in RapidMiner application, has an accuracy rate of 82.1%.

**Keywords**— Vehicles, Parts, Classification, Decision Tree Algorithm C4.5, RapidMiner Application.

## 1. PENDAHULUAN

Kendaraan bermotor merupakan alat yang penting bagi manusia, karena kendaraan bermotor sangat membantu aktivitas manusia dalam melakukan kegiatan sehari-hari. Seperti pergi ke suatu tempat, kendaraan bermotor sangat mengefisiensi waktu bagi para penggunanya. Kendaraan bermotor dibagi menjadi berbagai jenis, ada yang dapat digunakan untuk kegiatan sehari-hari sampai yang digunakan untuk mengangkut barang-barang dalam jumlah yang besar.

Karena kendaraan bermotor merupakan suatu alat yang penting, maka pertumbuhan jumlah kendaraan bermotor yang diperlukan setiap tahunnya juga semakin meningkat [1].

Perlu kita ketahui bahwa harga dari kendaraan bermotor tidaklah murah, sehingga perlu persiapan lebih untuk dapat membeli kendaraan bermotor. Harga dari suatu kendaraan bermotor, ditentukan pada bagian-bagian kendaraan atau *parts* yang ada pada kendaraan tersebut. Seperti yang sedang peneliti teliti, terdapat 15 bagian atau *parts* yang mempengaruhi harga dari suatu mobil, penjelasan 15 bagian mobil tersebut dapat dilihat pada tabel 1 dan tabel 2.

Penelitian ini merupakan proses *data mining* atau penggalian data. *Data mining* adalah suatu proses yang terdiri dari pengumpulan, dan pemakaian data untuk diolah agar dapat menemukan pola atau suatu hubungan dalam suatu *dataset* [2]. *Data mining* merupakan bagian dari tahapan proses *Knowledge Discovery in Database* (KDD). Dengan *data mining*, kita dapat melakukan klasifikasi data, prediksi data, dan mendapatkan informasi lain yang bermanfaat dari kumpulan data atau *dataset* dalam jumlah yang besar [3].

## 2. METODE PENELITIAN

Jenis penelitian ini merupakan penelitian kuantitatif yang menggunakan angka dan diklasifikasikan menggunakan teknik statistika. Metode yang peneliti gunakan adalah metode *decision tree* algoritma C4.5 yang diolah dengan menggunakan aplikasi RapidMiner. Algoritma C4.5 merupakan suatu algoritma yang dikembangkan dari model algoritma ID3, dimana algoritma C4.5 ini dapat mengatasi adanya *missing value*, dapat mengatasi data yang berkelanjutan atau data kontinyu, dapat melakukan *pruning* data, serta menggunakan nilai dari *gain ratio* sebagai kunci utama pemecahan masalah yang ada. Peneliti memilih menggunakan metode *decision tree* algoritma C4.5 karena metode ini telah dianggap sebagai algoritma yang sangat banyak digunakan untuk melakukan klasifikasi data atau objek, dimana hasil klasifikasinya dapat diperoleh dengan jelas. Tahapan dalam klasifikasi data menggunakan algoritma C4.5 adalah 1) Melakukan pengumpulan data, 2) Melakukan seleksi data, 3) Melakukan pengolahan data, 4) Melakukan pengujian data, dan 5) Membuat kesimpulan penelitian [4]. Proses klasifikasi tersebut akan menghasilkan gambaran untuk mendapatkan informasi yang baru [3]. Algoritma C4.5 merupakan salah satu algoritma yang dapat menghasilkan satu pohon keputusan atau yang biasanya kita sebut sebagai *decision tree* [5]. Dalam penelitian sebelumnya yang dilakukan oleh Ade Yuliana dan Duwi Bayu Pratomo, mereka belakukan penelitian dengan menggunakan metode *decision tree* algoritma C4.5 karena metode tersebut merupakan metode yang diusulkan untuk memprediksi kepuasan mahasiswa terhadap kinerja dosen prodi Teknik Informatika, dimana dalam penelitian tersebut, mereka menyatakan bahwa berdasarkan keseluruhan hasil pengujian, metode *decision tree* algoritma C4.5 lebih akurat dibandingkan dengan metode *decision stump* [6].

Pengolahan data penelitian ini dimulai dengan melakukan pembersihan data atau *cleaning* data atau *preprocessing* data. *Preprocessing* data merupakan langkah pertama untuk mengubah data yang mentah menjadi data yang siap untuk di analisis [2]. Langkah ini peneliti lakukan dengan menggunakan metode regresi linear berganda. Setelah itu, peneliti membagi data tersebut menjadi data *training* dan data *testing* dengan rasio 80:20. Pembagian data tersebut diperlukan untuk pembuatan model *decision tree* dengan algoritma C4.5 Setelah model tersebut diuji, maka langkah selanjutnya adalah perhitungan akurasi dan penelitian ini terselesaikan.

Permasalahan dalam penelitian ini adalah adanya faktor-faktor yang mempengaruhi harga dari kendaraan bermotor. Faktor-faktor ini peneliti dapatkan dengan menggunakan metode regresi linear berganda, dan etelah diketahui faktor-faktor apa saja yang berpengaruh, faktor-faktor yang tidak berpengaruh akan dibuang.

Batasan masalah yang ada adalah menggunakan metode *decision tree* untuk mengklasifikasi harga dari mobil dengan menggunakan dataset yang penulis dapatkan dari situs Kaggle.com, dengan tujuan penelitian untuk mengetahui bagaimana mengklasifikasi harga mobil, menghitung nilai akurasi serta nilai *gain ratio* untuk menentukan *node* dahan dan *node* daun didalam grafik pohon dalam metode *decision tree* algoritma C4.5. Selain itu, tujuan penelitian ini adalah untuk memberikan kepada masyarakat mengenai gambaran klasifikasi harga mobil dengan menggunakan metode *decision tree* dalam aplikasi RapidMiner.

Data yang diambil dan digunakan oleh peneliti merupakan *dataset* berbentuk excel yang didapatkan dari situs Kaggle, dengan *dataset* yang penulis teliti berjudul "Predict Car Price" Data ini diolah dengan menggunakan aplikasi RapidMiner, dimana data ini merupakan data yang berisikan harga-harga dari suatu mobil, yang menggunakan mata uang *dollar*. Penulis menggunakan metode *decision tree* algoritma C4.5 untuk mengklasifikasi harga mobil dalam *dataset* ini. Dataset dapat diakses melalui URL berikut: <https://www.kaggle.com/c/car-price-prediction-competition/overview>.

Menurut V. Roshan Joseph pada jurnalnya yang berjudul *Optimal Ratio for Data Splitting*, rasio pembagian data yang paling sering digunakan adalah 80:20 atau 80% untuk data *training* dan 20% untuk data *testing* [ 7]. Sampai saat ini, belum terdapat aturan atau penelitian yang jelas tentang rasio pembagian data yang optimal, sehingga peneliti juga mengikuti prinsip Pareto yang menggunakan pemisahan data dengan rasio 80:20. Jika akurasi yang dimiliki bagus, maka pemodelan klasifikasi harga mobil menggunakan metode *decision tree* algoritma C4.5 dapat dilakukan dan penelitian ini dapat diselesaikan. Tetapi jika akurasi yang dimiliki kurang bagus, maka peneliti harus menggunakan metode yang lain untuk mengklasifikasi harga mobil.

Dataset berbentuk excel ini memiliki variabel-variabel *independent* (variabel tidak terikat, biasanya dinotasikan “*x*”) serta 1 (satu) variabel *dependent* (variabel terikat, biasanya dinotasikan “*y*”) sebagai berikut:

**Tabel 1** Tabel Keterangan Variabel *Dependent* “*y*”

<i>Price</i> atau Harga	Variabel ini merupakan keterangan harga dari suatu mobil disaat pertama kali dibeli oleh konsumen, dimana menggunakan mata uang <i>dollar</i> .
-------------------------	---

**Tabel 2** Tabel Keterangan Variabel *Independent* “*x*”

Variabel <i>Independent</i> “ <i>x</i> ”	
<i>Make</i>	Variabel <i>make</i> merupakan variabel dari suatu merk mobil. Terdapat 36 merk mobil pada data mobil ini. Merk-merk mobil tersebut adalah Acura, Aston Martin, Bentley, BMW, Buick, Cadillac, Chevrolet, Chrysler, Dodge, Ferrari, FIAT, GMC, Honda, Hyundai, Infiniti, KIA, Lamborghini, Land Rover, Lexuc, Lincoln, Maserati, Mazda, Mercedes-Benz, Mitsubishi, Nissan, Plymouth, Pontiac, Porsche, Saab, Scion, Subaru, Suzuki, Toyota, Volkswagen dan Volvo.
Model	Variabel model merupakan variabel dari jenis atau tipe mobil tersebut. Terdapat ratusan jenis atau tipe mobil dalam data mobil ini. Contohnya seperti mobil dengan merk Aston Martin memiliki 11 tipe, yaitu DB7, DB9, DB9GTS, DBS, Rapide S, V12 Vanquish, V12 Vantage, V12 Vantage S, V8 Vantage, Vanquish dan Virage.
<i>Year</i>	Variabel <i>year</i> merupakan variabel yang menginformasikan kepada kita tahun pembuatan mobil tersebut. Tahun pembuatan mobil pada data mobil ini berjarak dari tahun 1990 hingga tahun 2017.

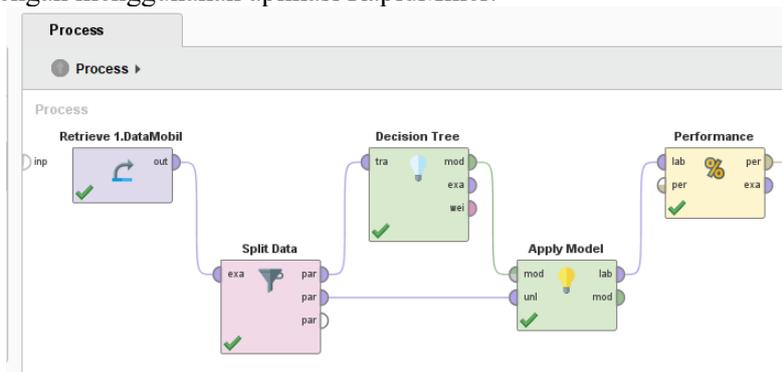
<i>Engine Fuel Type</i>	Variabel <i>engine fuel type</i> merupakan variabel yang menjelaskan tipe mesin yang terdapat didalam mobil tersebut, yang dihubungkan dengan jenis bahan bakar yang dapat digunakan, yaitu Flex-Fuel, Premium Unleaded, Regular Unleaded, dan Other.
<i>Engine HP</i>	Variabel <i>engine HP (horse power)</i> merupakan variabel yang menjelaskan seberapa besar tenaga yang dihasilkan dari mesin mobil tersebut, dimana <i>horse power</i> dalam data mobil ini berjarak dari 73 hingga 750 HP.
<i>Engine Cylinders</i>	Variabel ini menjelaskan berapa banyak silinder yang ada di mobil tersebut, yang berjarak dari 3 hingga 12 silinder mobil.
<i>Transmission Type</i>	Variabel ini menjelaskan tipe transmisi dari mobil tersebut. Transmisi mobil dari data mobil ini ada 3, yaitu Automatic, Manual dan Automated Manual.
<i>Driven Wheels</i>	Variabel <i>driven wheels</i> merupakan variabel yang menjelaskan apakah mobil tersebut digerakkan oleh roda belakang ( <i>rear wheels</i> ), roda depan ( <i>front wheels</i> ), keempat roda tersebut ( <i>all wheels</i> ) atau dapat diubah menjadi 2 roda depan-belakang atau keempat roda tersebut yang digunakan ( <i>four wheels</i> ).
<i>Number of Doors</i>	Variabel ini menjelaskan jumlah dari pintu yang ada didalam suatu mobil, yang berjumlah dari 2 hingga 4 pintu.
<i>Market Category</i>	Variabel ini menjelaskan kategori pasar penjualan mobil tersebut. Kategori pasar ini terdiri dari 11 kategori, yaitu Crossover, Diesel, Exotic, Factory Tuner, Flex Fuel, Hatchback, High-Performance, Hybrid, Luxury, Performance dan NODATA jika tidak ada kategori pasarnya.
<i>Vehicle Size</i>	Variabel ini menjelaskan tentang ukuran dari suatu mobil, seperti ukuran <i>compact, mid size</i> , dan besar atau <i>large</i> .
<i>Vehicle Style</i>	Variabel ini menjelaskan gaya apa yang dimiliki oleh mobil tersebut. Terdapat 15 gaya mobil dalam data mobil ini, yaitu 2dr Hatchback, 2dr SUV, 4dr Hatchback, 4dr SUV, Cargo Van, Convertible, Convertible SUV, Coupe, Crew Cab Pickup, Extended Cab Pickup, Passenger Minivan, Passenger Van, Regular Cab Pickup, Sedan dan Wagon.
<i>Highway MPG</i>	Variabel <i>highway MPG</i> merupakan variabel yang mencantumkan angka rata-rata mil per galon (MPG) yang dihasilkan jika mobil tersebut dibawa berkendara ke jalan yang mulus dan panjang seperti jalan tol. <i>Highway MPG</i> dalam data mobil ini memiliki jarak dari 17 hingga 44 MPG.
<i>City MPG</i>	Variabel <i>city MPG</i> merupakan variabel yang mencantumkan angka rata-rata mil per galon (MPG) yang dihasilkan jika mobil tersebut dibawa berkendara ke jalan yang ada di kota. <i>City MPG</i> dalam data mobil ini memiliki jarak dari 8 hingga 32 MPG.
<i>Popularity</i>	Variabel ini merupakan variabel ketenaran atau kepopuleran dari mobil tersebut. Skor kepopuleran mobil dalam data mobil ini memiliki jarak dari yang ter-rendah adalah 61, dan yang tertinggi adalah 3916.

Data mobil ini berjumlah 6898 data, yang kemudian peneliti bersihkan dan peneliti transformasikan data numerik menjadi kategorik. Transformasi data dalam *preprocessing data*, merupakan salah satu langkah yang diperlukan agar proses *data mining* dapat dilaksanakan [8]. Data-data numerik dalam data mobil ini, peneliti bagi menggunakan 3 kuartil, dan peneliti

transformasikan menjadi 4 kelas, yaitu kelas A, B, C, dan D, sama dengan variabel harga, peneliti transformasikan menjadi 4 kelas, yaitu *Low Class Price*, *Medium Class Price*, *High Class Price*, dan *Premium Class Price*.

### 3. HASIL DAN PEMBAHASAN

Data mobil yang semulanya berjumlah 6898 data, setelah dibersihkan menjadi berjumlah 4758 data mobil. Setelah peneliti melakukan pembersihan dan transformasi data, data tersebut peneliti olah dengan menggunakan aplikasi RapidMiner.



Gambar 1. Gambar Menggunakan Fungsi Operator Decision Tree Pada Aplikasi RapidMiner

Gambar diatas merupakan langkah-langkah untuk menggunakan fungsi operator *decision tree* algoritma C4.5 dalam aplikasi RapidMiner. Pengujian metode dilakukan dengan menggunakan data sebesar 80% untuk pemodelan dan 20%. Setelah semua fungsi operator tersebut ditarik dan diatur seperti langkah diatas, maka perhitungan akurasi dapat dilihat dengan bentuk *confusion matrix*. *Confusion matrix* adalah salah satu metode yang digunakan untuk mengetahui kinerja dari suatu metode klasifikasi data [9].

Tabel 3 Tabel Perhitungan *Confusion Matrix*

	Actual HCP	Actual LCP	Actual PCP	Actual MCP
Pred. HCP	239	6	43	38
Pred. LCP	1	255	0	35
Pred. PCP	21	0	253	3
Pred. MCP	43	25	1	230

$$Accuracy = \frac{T.HCP + T.LCP + T.PCP + T.MCP}{T.HCP + F.HCP + T.LCP + F.LCP + T.PCP + F.PCP + T.MCP + F.MCP} \quad (1)$$

Keterangan:

T.HCP = *True High Class Price* (Nilai aktual bernilai *high class price*, dengan hasil prediksi bernilai *high class price*).

T.LCP = *True Low Class Price* (Nilai aktual bernilai *low class price*, dengan hasil prediksi bernilai *low class price*).

T.PCP = *True Premium Class Price* (Nilai aktual bernilai *premium class price*, dengan hasil prediksi bernilai *premium class price*).

T.MCP = *True Medium Class Price* (Nilai aktual bernilai *medium class price*, dengan hasil prediksi bernilai *medium class price*).

F.HCP = *False High Class Price* (Nilai aktual bernilai *high class price*, dengan hasil prediksi selain *higher class price*)

F.LCP = *False Low Class Price* (Nilai aktual bernilai *low class price*, dengan hasil prediksi selain *low class price*)

F.PCP = *False High Price* (Nilai aktual bernilai *premium class price*, dengan hasil prediksi selain *premium class price*)

F.MCP = *False Lower Intermediate Price* (Nilai aktual bernilai *medium class*, dengan hasil prediksi selain *medium class price*)

$$= \frac{\text{Accuracy}}{239 + (1 + 21 + 43) + 255 + (6 + 0 + 25) + 253 + (43 + 0 + 1) + 230 + (38 + 35 + 0)} \quad (2)$$

$$\text{Accuracy} = \frac{977}{1190} = 0,821 \text{ atau } 82.1\%$$

Pada rumus diatas, tertera bahwa besar akurasi dari model *decision tree* algoritma C4.5 yang peneliti gunakan untuk mengklasifikasi data harga mobil adalah sebesar 82.1%. Menurut peneliti, akurasi sebesar 82.1% sudah cukup baik untuk penelitian ini. Tetapi, perlu kita ketahui juga bahwa nilai akurasi tersebut dapat ditingkatkan. Dalam penelitiannya Aldi Nurzahputra dan Much Aziz Muslim, nilai akurasi dapat ditingkatkan dengan menerapkan *adaboost* pada proses klasifikasi data [10]. Informasi ini dapat berguna kepada para pembaca yang ingin meningkatkan angka akurasi dari suatu klasifikasi.

Karena *tree* yang dihasilkan menggunakan data mobil ini terlalu besar, maka peneliti melakukan perhitungan *node decision tree* dengan menghitung nilai dari *entropy*, *gain*, *split information* dan nilai *gain ratio* dari setiap variabel yang ada pada penelitian ini. Nilai dari *gain ratio* merupakan acuan kita semua untuk membuat grafik *decision tree* algoritma C4.5 secara manual. Tabel 4 dibawah ini merupakan tabel yang menampilkan angka dari perhitungan *entropy*, *gain*, *split information* dan nilai *gain ratio* dari setiap variabel, yang dapat digunakan sebagai acuan untuk membuat grafik *decision tree*.

**Tabel 4** Tabel Perhitungan Nilai Entropy, Gain, Split Information dan Gain Ratio Dalam Perhitungan Root Node

Atribut		Jumlah Data	Low Class Price	Medium Class Price	High Class Price	Premium Class Price	Entropy	Gain	Split Info	Gain Ratio
Jumlah Data		4758	1144	1211	1215	1188	2,00			
Year	A	3103	286	810	950	1057	1,87	0,31	1,46	0,21
	B	772	172	283	212	105	1,92			
	C	499	303	118	52	26	1,49			
	D	384	383	0	1	0	0,03			
Engine Fuel Type	Regular	2846	1104	902	585	255	1,84	0,29	1,32	0,22
	Premium	1512	7	237	458	810	1,46			
	FlexFuel	339	29	57	154	99	1,77			
	Other	61	4	15	18	24	1,80			
Engine HP Score	A	1318	19	48	308	943	1,10	0,69	2,00	0,35
	B	1095	36	242	609	208	1,57			
	C	1156	250	601	269	36	1,61			
	D	1189	839	320	29	1	1,00			

Cylinder Score	A	105	5	0	0	100	0,21	0,11	1,7	0,06
	B	1783	393	491	472	427	1,99			
	C	1548	582	377	197	392	1,91			
	D	1322	164	343	546	269	1,87			
Transmission Type	Automatic	3312	471	902	1011	928	1,95	0,14	1,10	0,13
	Manual	1175	668	231	130	146	1,65			
	Automated	271	5	78	74	114	1,66			
Driven Wheels	Front	1852	689	698	387	78	1,73	0,20	1,89	0,11
	Rear	1313	298	217	309	489	1,94			
	All	975	38	164	314	459	1,65			
	Four	618	119	132	205	162	1,97			
Door Score	A	1274	454	230	206	384	1,92	0,06	1,02	0,05
	B	141	84	34	23	0	1,37			
	C	3343	606	947	986	804	1,98			
Vehicle Size	Compact	1935	799	608	283	245	1,83	0,15	1,54	0,10
	Midsize	1798	217	485	619	477	1,92			
	Large	1025	128	118	313	466	1,77			
Highway MPG	A	265	128	101	25	11	1,55	0,36	1,68	0,21
	B	1630	432	569	378	251	1,94			
	C	2167	407	414	644	702	1,96			
	D	696	177	127	168	224	1,97			
City MPG	A	589	172	235	114	68	0,52	0,29	1,80	0,15
	B	2141	518	332	673	618	1,95			
	C	685	218	25	154	288	1,71			
	D	1343	236	619	274	214	1,85			
Popularity	A	1085	201	230	358	296	1,96	0,03	2,00	0,02
	B	1285	397	339	247	302	1,98			
	C	1286	268	445	344	229	1,95			
	D	1102	278	197	266	361	1,97			
Make	Chevrolet	487	151	140	119	77	1,96	0,52	5,16	0,10
	Dodge	422	161	124	95	42	1,87			
	Volkswagen	369	58	204	96	11	1,55			
	Toyota	307	84	73	117	33	1,88			
	Nissan	254	63	74	89	28	1,90			
	GMC	251	59	53	93	46	1,94			
	Cadillac	189	8	0	20	161	0,73			
	Mazda	185	70	97	17	1	1,38			
	Honda	179	35	79	61	4	1,63			
	BMW	161	6	1	36	118	1,03			
	Suzuki	159	102	55	2	0	1,02			
	Mercedes	150	32	0	20	98	1,26			
	Infiniti	148	9	2	65	72	1,36			
	Hyundai	143	53	58	27	5	1,68			
	Audi	137	5	0	43	89	1,10			
	Acura	112	19	16	43	34	1,89			
	Subaru	110	18	51	41	0	1,47			
Volvo	96	8	5	43	40	1,57				
Kia	92	41	30	16	5	1,71				

**Yonathan Anggraiwan: Klasifikasi Harga Mobil Menggunakan Metode Decision Tree Algoritma C4.5**

	Mitsubishi	88	25	52	11	0	1,34			
	Buick	86	16	24	36	10	1,85			
	Lexus	82	10	0	27	45	1,37			
	Pontiac	72	35	29	8	0	1,39			
	Land Rover	66	1	3	16	46	1,15			
	Lincoln	63	4	0	29	30	1,28			
	Chrysler	58	18	17	20	3	1,79			
	Aston Martin	46	0	0	0	46	0,00			
	Saab	41	0	6	23	12	1,39			
	Plymouth	31	31	0	0	0	0,00			
	Bentley	30	0	0	0	30	0,00			
	Maserati	29	0	0	0	29	0,00			
	Porsche	28	1	0	0	27	0,22			
	Scion	27	18	8	1	0	1,09			
	Lamborghini	26	0	0	0	26	0,00			
Ferrari	19	0	0	0	19	0,00				
FIAT	15	5	10	0	0	0,92				
Market Category	NO DATA	1484	698	492	236	58	1,64	0,45	2,82	0,16
	Crossover	865	32	280	337	216	1,73			
	Luxury	792	90	17	238	447	1,46			
	Hatchback	429	226	159	34	10	1,43			
	Flex Fuel	390	45	88	169	88	1,85			
	Performance	278	33	137	97	11	1,58			
	Factory	191	16	17	25	133	1,36			
	Exotic	174	0	0	0	174	0,00			
	HighPerformance	84	0	5	52	27	1,20			
	Diesel	56	4	14	23	15	1,81			
	Hybrid	15	0	2	4	9	1,34			
Vehicle Style	Sedan	1188	305	299	263	321	2,00	0,25	3,18	0,08
	4dr SUV	1120	61	316	376	367	1,80			
	Couple	505	134	65	91	215	1,86			
	Convertible	320	26	60	68	166	1,71			
	Extended	268	94	86	83	5	1,69			
	4dr Hatch	261	131	88	31	11	1,59			
	Crew Cab	254	0	71	122	61	1,52			
	2dr Hatch	208	105	89	14	0	1,28			
	Wagon	208	77	43	54	34	1,51			
	Passenger Minivan	154	42	52	55	5	1,73			
	Regular Cab	146	86	28	32	0	1,39			
	2dr SUV	48	42	3	2	1	0,73			

	Cargo Van	36	24	8	4	0	1,22			
	Passenger Van	29	7	3	19	0	1,23			
	Convertible SUV	13	10	0	1	2	0,99			

Pada tabel tersebut, nilai *gain ratio* tertinggi terdapat pada variabel *Engine HP Score*, dimana variabel ini memiliki 4 kelas, yaitu kelas A, kelas B, kelas C, dan kelas D. Sehingga, peneliti mengambil 4 kesimpulan, yaitu:

- a. Pada kelas A, didominasi oleh data harga mobil berkelas premium, yaitu sebesar 71.5% dan data harga mobil berkelas tinggi, yaitu sebesar 23.3%.
- b. Pada kelas B, terdapat *leaf node* yang didominasi oleh data harga mobil yang tinggi sebesar 55%, data harga mobil menengah atau *medium price* sebesar 22%, dan data harga mobil premium sebesar 19%.
- c. Pada kelas C, terdapat *leaf node* didominasi oleh harga mobil menengah, yaitu sebesar 52%, data harga mobil medium sebesar 23%, dan data harga mobil rendah sebesar 21.6%.
- d. Pada kelas D, data ini didominasi oleh *leaf node* data harga mobil yang rendah atau *low class price*, yaitu sebesar 70%, dan data harga mobil medium sebesar 27%.

#### 4. KESIMPULAN DAN SARAN

Penelitian ini memiliki 5 kesimpulan, yaitu:

- a. Pada kelas A, didominasi oleh data harga mobil berkelas premium, yaitu sebesar 71.5% dan data harga mobil berkelas tinggi, yaitu sebesar 23.3%.
- b. Pada kelas B, terdapat *leaf node* yang didominasi oleh data harga mobil yang tinggi sebesar 55%, data harga mobil menengah atau *medium price* sebesar 22%, dan data harga mobil premium sebesar 19%.
- c. Pada kelas C, terdapat *leaf node* didominasi oleh harga mobil menengah, yaitu sebesar 52%, data harga mobil medium sebesar 23%, dan data harga mobil rendah sebesar 21.6%.
- d. Pada kelas D, data ini didominasi oleh *leaf node* data harga mobil yang rendah atau *low class price*, yaitu sebesar 70%, dan data harga mobil medium sebesar 27%.
- e. *Decision tree* dengan algoritma C4.5 merupakan metode untuk melakukan klasifikasi data yang memberikan informasi tersembunyi dari suatu data mentah. Metode ini sering dipakai oleh banyak peneliti untuk melakukan klasifikasi data. Dengan metode ini, peneliti melakukan klasifikasi harga mobil dan memiliki nilai akurasi sebesar 82.1%.

Saran dari peneliti untuk peneliti berikutnya yang melakukan klasifikasi menggunakan metode *decision tree* algoritma C4.5 adalah peneliti berikutnya dapat menambahkan variabel-variabel yang tidak ada pada variabel tersebut. Contohnya pada data prediksi harga mobil yang peneliti pakai, peneliti berikutnya dapat menambahkan 1 variabel baru, yaitu negara asal mobil tersebut, seperti Jepang, Eropa, Amerika, dan lain-lain.

#### DAFTAR PUSTAKA

- [1] Yulanto, D. M., & Iskandar, H., "STUDI ANALISIS PERKEMBANGAN TEKNOLOGI KENDARAAN LISTRIK HIBRIDA," *Journal of Automotive Technology Vocational Education*, 31-44, 2021.
- [2] N. Iriadi, "Implementasi Data Mining Pada Klasifikasi Ketidakhadiran Pegawai Menggunakan Metode C4.5," *Computer Science (CO-SCIENCE)*, vol. I, pp. 53-61, 2021.

- [3] Mardi, Y., “ Data Mining : Klasifikasi Menggunakan Algoritma C4.5,” Jurnal Edik Informatika, 213-219, 2017.
- [4] Asroni, dkk, “Penerapan Algoritma C4.5 untuk Klasifikasi Jenis Pekerjaan Alumni di Universitas Muhammadiyah Yogyakarta,” Jurnal Semesta Teknik, 158-165, 2018.
- [5] Budiarto, S. P., “Klasifikasi Data Mahasiswa Menggunakan Metode Decision Tree Algoritma C4.5 divisualisasikan dalam GIS (Studi Kasus: Universitas PGRI Banyuwangi),” *Jurnal CITISEE*, 45-52, 2018.
- [6] Yuliana, A. & Pratomo, D. B., “ALGORITMA DECISION TREE (C4.5) UNTUK MEMPREDIKSI KEPUASAN MAHASISWA TERHADAP KINERJA DOSEN POLITEKNIK TEDC BANDUNG,” Seminar Nasional Inovasi Teknologi UN PGRI Kediri, 377-384, 2017.
- [7] V. R. Joseph, "Optimal Ratio for Data Splitting," *ARVIV*, 2022.
- [8] Firdaus, M. & C. Triawan, A., “IMPLEMENTASI ALGORITMA DECISION TREE UNTUK KLASIFIKASI POLA SERANGAN PADA LOG FILE,” Repository Universitas Muhammadiyah Jember, 2019
- [9] Karsito & Susanti, S., “KLASIFIKASI KELAYAKAN PESERTA PENGAJUAN KREDIT RUMAH DENGAN ALGORITMA NAÏVE BAYES DI PERUMAHAN AZZURA RESIDENCIA,” Jurnal Teknologi Pelita Bangsa – SIGMA, 43-48, 2019.
- [10] Nurzahputra, A. & Muslim, M. A., “PENINGKATAN AKURASI PADA ALGORITMA C4.5 MENGGUNAKAN ADABOOST UNTUK MEMINIMALKAN RESIKO KREDIT,” Prosiding SNATIF, 243-247, 2017.