

# Sistem Klasifikasi Variabel Tingkat Penerimaan Konsumen Terhadap Mobil Menggunakan Metode *Random Forest*

Yusuf Sulisty Nugroho<sup>1</sup> dan Nova Emiliyawati<sup>2</sup>

*Fakultas Komunikasi dan Informatika, Universitas Muhammadiyah Surakarta  
Jl. A. Yani Tromol Pos I Pabelan, Kartasura, Surakarta, Indonesia  
yusuf.nugroho@ums.ac.id<sup>1</sup>*

**Abstrak**— Saat ini tidak dapat dipungkiri bahwa persaingan pasar industri mobil di berbagai belahan dunia sangat ketat tak terkecuali di Indonesia. Indonesia memiliki industri manufaktur mobil terbesar kedua di Asia Tenggara setelah Thailand. Namun hasil statistik menunjukkan bahwa tingkat penjualan mobil di Indonesia masih belum stabil akibat tingkat penerimaan konsumen yang rendah. Ada banyak faktor yang mempengaruhi tingkat penerimaan konsumen terhadap produk-produk mobil yang baru diluncurkan. Jika hal tersebut terus terjadi, akan berimbas buruk pada profitabilitas produsen mobil di negara ini. Untuk membantu mengurangi masalah tersebut, maka dibuat sebuah sistem untuk mengklasifikasi faktor yang mempengaruhi permasalahan tingkat penerimaan mobil oleh pihak konsumen. Faktor-faktor yang digunakan dalam sistem sebagai variabel yang mempengaruhi tingkat penerimaan konsumen terdiri dari harga pembelian, biaya perawatan, jumlah pintu, kapasitas penumpang, ukuran bagasi, dan taksiran keselamatan penumpang. Sistem pengklasifikasi variabel ini dibangun dengan menerapkan metode *Random Forest* (RF). Pada metode ini, data maupun atribut diambil secara acak sehingga dimungkinkan akan menghasilkan berbagai model pohon keputusan. Adapun variabel yang menempati sebagai *root node* dalam pohon keputusan merupakan variabel yang paling signifikan dalam mempengaruhi tingkat penerimaan mobil pada konsumen. Hasil dari sistem klasifikasi yang dibangun dapat dijadikan pertimbangan bagi produsen mobil di masa mendatang, sehingga produksi mobil menjadi lebih efektif, dapat meminimalisir kerugian, dan meningkatkan profitabilitas produsen.

**Kata kunci**— *data mining*, pohon keputusan, *Random Forest*, sistem klasifikasi, tingkat penerimaan konsumen

## I. PENDAHULUAN

Mobil merupakan kendaraan darat yang bergerak dengan tenaga mesin, memiliki roda empat atau lebih (selalu genap), biasanya untuk menghidupkan mesinnya menggunakan bahan bakar minyak (bensin atau solar). Mobil pribadi adalah mobil milik perserorangan yang biasanya digunakan untuk keperluan pribadi. Mobil inilah yang saat ini sangat diminati oleh masyarakat Indonesia, didorong dengan penjualan mobil baru di enam negara utama Asia Tenggara yang tumbuh 8 persen, dengan total pencapaian 276.000 unit. Indonesia menjadi yang terbesar di Asia Tenggara, dengan penjualan naik 6 persen menjadi 96.000 unit [1].

Berdasarkan Tabel I dapat dilihat bahwa minat konsumen mobil pada tahun 2016 bisa dinyatakan meningkat meskipun tidak stabil. Hal tersebut perlu diimbangi oleh para produsen mobil di Indonesia untuk selalu memperbaiki kinerja

produknya agar produk barunya dapat diterima oleh konsumen pada persaingan pasar yang ketat [2]. Terdapat banyak faktor yang mempengaruhi tingkat penerimaan pada konsumen, begitu pula pada perusahaan mobil. Tingkat akurasi prediksi terhadap diterima atau tidaknya suatu produk tentu saja sangat diperhitungkan oleh perusahaan, karena akan berpengaruh besar pada profitabilitas perusahaan tersebut [3].

Penelitian ini dilakukan dalam rangka untuk mengklasifikasi faktor-faktor yang mempengaruhi tingkat penerimaan konsumen terhadap mobil menggunakan metode *Random Forest* (RF). Metode yang digunakan untuk mengklasifikasi data ini berkaitan erat dengan *data mining*. *Data mining* merupakan teknik yang digunakan untuk menganalisis *dataset* dan memprediksi pola yang terkandung di dalam data, sedangkan klasifikasi merupakan teknik untuk pengumpulan datanya [4].

TABEL I. STATISTIK PENJUALAN MOBIL DI INDONESIA 2014 – 2016

Tahun	Bulan								
	Januari	Februari	Maret	April	Mei	Juni	Juli	Agustus	Total
2014	85.003	88.208	94.093	84.771	88.578	91.492	62.603	96.294	691.042
2015	94.194	88.740	99.740	81.600	79.375	82.172	55.615	90.537	671.973
2016	103.609	111.824	113.067	106.124	96.872	110.614	91.334	96.652	830.096

Sumber : *wholesales* Gabungan Industri Kendaraan Bermotor Indonesia (Gaikindo)

Banyak penelitian yang menggunakan metode *Random Forest*, yaitu salah satunya yang dilakukan oleh [5]. Penelitian yang dilakukannya menggunakan metode *Random Forest* dan *Multivariate Adaptive Regression Spline (MARS) binary response* untuk mengklasifikasi penderita HIV/AIDS di Surabaya. Berdasarkan penelitian yang dilakukan, dapat diketahui variabel yang memiliki pengaruh paling tinggi untuk status HIV/AIDS yaitu usia kemudian diikuti oleh jenis pekerjaan, pernah ditahan karena kasus NAPZA, status nikah, dan selalu pakai jarum steril. Akurasi klasifikasi status HIV/AIDS di Surabaya menggunakan metode MARS sebesar 80,28%. Pada metode RF diperoleh klasifikasi terbaik sebesar 97,80%. Pada RF MARS diperoleh akurasi klasifikasi terbaik sebesar 91,00%.

Referensi [3] membandingkan *Original Random Forest* dengan *Paralel Random Forest* pada ketidakseimbangan data-data untuk memprediksi ketidakseimbangan peminjam dalam mekanisme perbankan. Referensi [3] membandingkan dengan cara menjalankannya pada 2 PC yang berbeda dengan menggunakan R sebagai tools. Penelitian tersebut menyatakan bahwa *Paralel Random Forest* seharusnya dipertimbangkan sebagai praktek standar dalam penanganan *database* yang besar.

Berdasarkan latar belakang tersebut, maka dalam penelitian ini akan dibangun sebuah sistem klasifikasi berbasis *web* dengan menggunakan metode *Random Forest* dengan menerapkan algoritma *gain ratio* sebagai kriteria pemilihan atribut yang tersedia. Sistem ini dibuat untuk membantu perusahaan dalam menentukan variabel-variabel yang

mempengaruhi tingkat penerimaan konsumen terhadap produk yang akan diproduksi. Sehingga proses produksi menjadi semakin efektif, tepat sasaran, dan mengurangi resiko produktifitas mobil.

## II. METODE

### A. Pengumpulan Data

Dalam penelitian ini, *dataset* tingkat penerimaan mobil (*Car Acceptability*) diambil dari situs <http://archive.ics.uci.edu/ml/> yang memiliki jumlah total data sebanyak 1728 data, terdiri dari 7 variabel. Setiap variabel memiliki kelas masing-masing, dimana setiap kelas memiliki batasan tersendiri [6]. Tabel II merupakan potongan *dataset* tentang tingkat penerimaan mobil yang telah dikumpulkan.

### B. Penentuan Atribut dan Kelas

Penerapan metode klasifikasi menggunakan *Random Forest* diperlukan penentuan atribut yang dibedakan menjadi dua jenis yaitu variabel X dan variabel Y. Variabel terikat didefinisikan dalam variabel X dan variabel bebas didefinisikan dalam variabel Y [7]. Variabel X yang digunakan pada penelitian yaitu Harga Pembelian (*Buying Price*), Biaya Perawatan (*Maintenance Price*), Jumlah Pintu (*Number of Doors*), Kapasitas Penumpang (*Capacity*), Ukuran Bagasi (*Size of Luggage Boot*) dan Taksiran Keselamatan (*Estimated Safety*), sedangkan variabel Y yang digunakan berupa Tingkat Penerimaan Konsumen (*Car Acceptability*) seperti yang ditunjukkan dalam Tabel III.

TABEL II. POTONGAN DATASET TINGKAT PENERIMAAN MOBIL

Buying_price	Maintenance_price	Number_of_Doors	Capacity	Size_of_Luggage_boot	Estimated_safety	Car_acceptability
vhigh	vhigh	2	2	small	low	unacc
vhigh	vhigh	2	2	small	med	unacc
vhigh	vhigh	2	2	small	high	unacc
vhigh	low	5 more	4	small	high	acc
vhigh	low	5 more	4	med	low	unacc
vhigh	low	5 more	4	med	med	acc
med	low	5 more	more	big	high	vgood
low	vhigh	2	2	small	low	unacc

TABEL III. JENIS ATRIBUT DAN KELAS DALAM PENELITIAN

Notasi	Atribut	Kelas
X1	Harga Pembelian ( <i>Buying Price</i> )	<i>vhigh</i> (harga > 500 juta), <i>high</i> (250 < harga ≤ 500 juta), <i>med</i> (100 < harga ≤ 250 juta), <i>low</i> (harga ≤ 100 juta)
X2	Biaya Perawatan ( <i>Maintenance Price</i> )	<i>vhigh</i> (biaya > 1 juta/bulan), <i>high</i> (800 < biaya ≤ 1 juta/bulan), <i>med</i> (600 ribu/bulan < biaya ≤ 800 ribu/bulan), <i>low</i> (biaya ≤ 600 ribu/bulan)
X3	Jumlah Pintu ( <i>Number of Doors</i> )	2 (pintu = 2), 3 (pintu = 3), 4 (pintu = 4), 5 more (pintu ≥ 5)
X4	Kapasitas Penumpang ( <i>Capacity</i> )	2 (penumpang = 2), 4 (penumpang = 4), more (penumpang > 4)
X5	Ukuran Bagasi ( <i>Size of Luggage Boot</i> )	<i>small</i> (bagasi ≤ 1 m <sup>3</sup> ), <i>med</i> (1 m <sup>3</sup> < bagasi ≤ 2 m <sup>3</sup> ), <i>big</i> (bagasi > 2 m <sup>3</sup> )
X6	Taksiran Keselamatan ( <i>Estimated Safety</i> )	<i>low</i> (safety ≥ 67%), <i>med</i> (33% ≤ safety < 67%), <i>high</i> (safety < 33%)
Y	Tingkat Penerimaan Mobil ( <i>Car Acceptability</i> )	<i>vgood</i> (persentase = 76-100%), <i>good</i> (persentase = 51-75%), <i>acc</i> (persentase = 26-50%), <i>unacc</i> (persentase = 0-25%)

### C. Penggunaan Metode *Random Forest*

Metode *Random Forest* (RF) merupakan metode yang dapat meningkatkan hasil akurasi, karena dalam membangkitkan simpul anak untuk setiap *node* dilakukan secara acak. Metode ini digunakan untuk membangun pohon keputusan yang terdiri dari *root node*, *internal node*, dan *leaf node* dengan mengambil atribut dan data secara acak sesuai ketentuan yang diberlakukan. *Root node* merupakan simpul yang terletak paling atas, atau biasa disebut sebagai akar dari pohon keputusan. *Internal node* adalah simpul percabangan, dimana *node* ini mempunyai output minimal dua dan hanya ada satu input. Sedangkan *leaf node* atau *terminal node* merupakan simpul terakhir yang hanya memiliki satu input dan tidak mempunyai output. Pohon keputusan dimulai dengan cara menghitung nilai *entropy* sebagai penentu tingkat ketidakmurnian atribut dan nilai *information gain*. Untuk menghitung nilai *entropy* digunakan rumus seperti pada persamaan 1, sedangkan nilai *information gain* menggunakan persamaan 2 [8].

$$Entropy(Y) = -\sum_i p(c|Y) \log_2 p(c|Y) \quad (1)$$

Dimana  $Y$  adalah himpunan kasus dan  $p(c|Y)$  merupakan proporsi nilai  $Y$  terhadap kelas  $c$ .

*Information Gain* ( $Y, a$ )

$$= Entropy(Y) - \sum_{v \in Values(a)} \frac{|Y_v|}{|Y_a|} Entropy(Y_v) \quad (2)$$

Dimana  $Values(a)$  merupakan semua nilai yang mungkin dalam himpunan kasus  $a$ .  $Y_v$  adalah subkelas dari  $Y$  dengan kelas  $v$  yang berhubungan dengan kelas  $a$ .  $Y_a$  adalah semua nilai yang sesuai dengan  $a$ .

### D. Penerapan Algoritma *Gain Ratio*

Pemilihan atribut sebagai simpul, baik akar (*root*) atau simpul internal didasarkan pada nilai *information gain* tertinggi dari atribut-atribut yang ada [9]. Nilai *gain ratio* diperoleh dari hasil perhitungan *information gain* yang dibagi dengan *split information*. Nilai *split information* dapat dilihat pada persamaan 3 [10]. Sedangkan nilai *gain ratio* seperti pada persamaan 4 [11].

$$Split\ Information(S, A) = \sum_{i=1}^c \left( \frac{|S_i|}{|S|} \right) \log_2 \left( \frac{|S_i|}{|S|} \right) \quad (3)$$

Dimana *split information* ( $S, A$ ) adalah nilai estimasi entropi dari variabel input  $S$  yang memiliki kelas  $c$  dan  $|S_i|/|S|$  merupakan probabilitas kelas  $i$  dalam atribut.

$$Gain\ Ratio(S, A) = \frac{Information\ Gain(S, A)}{Split\ Information(S, A)} \quad (4)$$

## III. HASIL DAN PEMBAHASAN

### A. Hasil Pengembangan Sistem

Sistem pengklasifikasi faktor yang mempengaruhi tingkat penerimaan mobil terhadap konsumen telah selesai dibangun yang terdiri dari beberapa halaman. Halaman utama adalah halaman yang pertama kali muncul saat *user* masuk ke dalam sistem. Halaman utama berisi deskripsi singkat tentang sistem. Pada halaman utama, terdapat beberapa menu utama yaitu Home, Data Training, Main Process, dan Login. Tampilan halaman utama ditunjukkan pada Gambar 1.

Gambar 2 menunjukkan tampilan pada menu Data Training. Menu ini untuk menampilkan seluruh data tingkat penerimaan mobil yang akan digunakan dalam proses klasifikasi. Pada menu ini juga terdapat *fitur search* untuk memudahkan *user* dalam mencari data.



Gambar 1. Tampilan halaman utama

DATA PENERIMAAN KONSUMEN MOBIL BERDASARKAN FAKTOR-FAKTOR YANG MEMPENGARUHI								
Search by...		Kata Kunci		Cari				
No	ID Data	Buying Price	Maintenance Price	Number of Doors	Capacity	Size of Luggage Boot	Estimated Safety	Car Acceptability
2	data 2	vhigh	vhigh	2	2	small	med	unacc
3	data 3	vhigh	vhigh	2	2	small	high	unacc
4	data 4	vhigh	vhigh	2	2	med	low	unacc
5	data 5	vhigh	vhigh	2	2	med	med	unacc
6	data 6	vhigh	vhigh	2	2	med	high	unacc
7	data 7	vhigh	vhigh	2	2	big	low	unacc
8	data 8	vhigh	vhigh	2	2	big	med	unacc
9	data 9	vhigh	vhigh	2	2	big	high	unacc
10	data 10	vhigh	vhigh	2	4	small	low	unacc
11	data 11	vhigh	vhigh	2	4	small	med	unacc
12	data 12	vhigh	vhigh	2	4	small	high	unacc
13	data 13	vhigh	vhigh	2	4	med	low	unacc
14	data 14	vhigh	vhigh	2	4	med	med	unacc
15	data 15	vhigh	vhigh	2	4	med	high	unacc
16	data 16	vhigh	vhigh	2	4	big	low	unacc

Gambar 2. Tampilan Data Training

Proses utama berlangsung pada menu Main Process. Di dalam menu ini terdapat *form* untuk mengisikan jumlah data dan jumlah atribut sesuai kebutuhan. Jumlah data dan atribut yang sudah di submit oleh *user* akan diproses menggunakan metode *Random Forest*. Gambar 3 adalah tampilan awal pada menu Main Process.

Proses *mining* menggunakan *Random Forest* dijalankan setelah *user* melakukan *submit* data pada menu Main Process. Proses utama dilanjutkan untuk melakukan perhitungan menggunakan algoritma *gain ratio*. Di dalam proses ini terdapat dua tombol, yaitu tombol Lihat Hasil Pohon Keputusan, dan tombol Cetak Hasil Pohon Keputusan.

Tombol Lihat Hasil Pohon Keputusan digunakan untuk melihat Hasil Pohon Keputusan dari proses yang dilakukan. Sedangkan tombol Cetak Hasil Pohon Keputusan untuk mencetak hasil Pohon Keputusan. Gambar 4 menunjukkan tampilan contoh hasil perhitungan pada Main Process,

sedangkan contoh Hasil Pohon Keputusan ditunjukkan pada Gambar 5.

**Random Forest Method**

**Jumlah Data**

Masukkan jumlah data yang anda inginkan

---

**Jumlah Atribut**

Masukkan jumlah atribut yang anda inginkan (1-6)

---

Gambar 3. Tampilan Awal menu Main Process

Hasil Proses Klasifikasi Random Forest pada Faktor Pengaruh Penerimaan Mobil oleh Konsumen											
			Lihat Hasil Pohon Keputusan		Cetak Hasil Pohon Keputusan						
No	Atribut	Nilai_atribut	Jumlah Kasus Total	Jumlah Kasus Acc	Jumlah Kasus Unacc	Jumlah Kasus Good	Jumlah Kasus Vgood	Entropy	Information Gain	Split Info	Gain Ratio
1	Total	Total	30	4	24	2	0	0.9056			0
2	maintenance_price	vhigh	11	1	10	0	0	0.4395	0.2324	1.8094	0.1284
3	maintenance_price	high	7	1	6	0	0	0.5917	0.2324	1.8094	0.1284
4	maintenance_price	med	10	1	8	1	0	0.9219	0.2324	1.8094	0.1284
5	maintenance_price	low	2	1	0	1	0	1	0.2324	1.8094	0.1284
6	estimated_safety	low	15	0	15	0	0	0	0.2845	1.4984	0.1899
7	estimated_safety	med	7	3	3	1	0	1.4488	0.2845	1.4984	0.1899
8	estimated_safety	high	8	1	6	1	0	1.0613	0.2845	1.4984	0.1899
9	number_of_doors	2	8	1	6	1	0	1.0613	0.209	1.9839	0.1053
10	number_of_doors	3	6	0	6	0	0	0	0.209	1.9839	0.1053
11	number_of_doors	4	7	0	6	1	0	0.5917	0.209	1.9839	0.1053
12	number_of_doors	5 more	9	3	6	0	0	0.9183	0.209	1.9839	0.1053
13	capacity	2	6	0	6	0	0	0	0.0817	1.4566	0.0561
14	capacity	4	16	3	12	1	0	1.0141	0.0817	1.4566	0.0561

Gambar 4. Contoh Hasil Perhitungan dalam Main Process

**HASIL KLASIFIKASI FAKTOR-FAKTOR YANG MEMPENGARUHI TINGKAT PENERIMAAN MOBIL PADA KONSUMEN MENGGUNAKAN METODE RANDOM FOREST**

---

**POHON KEPUTUSAN:**

estimated\_safety = low (acc = 0, unacc = 15, good = 0, vgood = 0) : **unacc**  
 estimated\_safety = med (acc = 3, unacc = 3, good = 0, vgood = 0) : **acc - unacc**  
 estimated\_safety = high (acc = 2, unacc = 5, good = 0, vgood = 2) : **?**  
 | capacity = 2 (acc = 0, unacc = 5, good = 0, vgood = 0) : **unacc**  
 | capacity = 4 (acc = 2, unacc = 0, good = 0, vgood = 1) : **?**  
 | | size\_of\_luggage\_boot = small (acc = 1, unacc = 0, good = 0, vgood = 0) : **acc**  
 | | size\_of\_luggage\_boot = med (acc = 1, unacc = 0, good = 0, vgood = 0) : **acc**  
 | | size\_of\_luggage\_boot = big (acc = 0, unacc = 0, good = 0, vgood = 1) : **vgood**  
 | capacity = more (acc = 0, unacc = 0, good = 0, vgood = 1) : **vgood**

**RULE :**

1. if (estimated\_safety == low) then unacc (id = 1)
2. if (estimated\_safety == med) then acc - unacc (id = 2)
3. if (estimated\_safety == high AND capacity == 2) then unacc (id = 4)
4. if (estimated\_safety == high AND capacity == 4 AND size\_of\_luggage\_boot == small) then acc (id = 6)
5. if (estimated\_safety == high AND capacity == 4 AND size\_of\_luggage\_boot == med) then acc (id = 7)
6. if (estimated\_safety == high AND capacity == 4 AND size\_of\_luggage\_boot == big) then vgood (id = 8)
7. if (estimated\_safety == high AND capacity == more) then vgood (id = 9)

Gambar 5. Contoh Tampilan Hasil Pohon Keputusan

Adapun menu Login digunakan bagi admin untuk menuju halaman Admin. Menu ini memiliki *form login* dimana admin dapat mengisi *username* dan *password*. Gambar 6 merupakan tampilan *form Login*.

Gambar 6. Tampilan form Login

Sementara itu, halaman Admin berisi menu Home, Data Training, Main Process, Pohon Keputusan, dan Logout. Pada menu Home berisi pengantar singkat bagi admin, sedangkan Data Training berisi *dataset* total yang digunakan pada proses. Pada halaman ini terdapat beberapa pilihan bagi admin yaitu menambah data, *upload* data menggunakan file .csv, edit data, dan hapus data. Terdapat pula *fitur search* untuk memudahkan admin dalam mencari data. Tampilan menu Data Training pada halaman admin dapat dilihat pada Gambar 7.

Hi, nova. Sistem Pengklasifikasi

---

**DATA PENERIMAAN MOBIL BERDASARKAN BEBERAPA FAKTOR YANG MEMPENGARUHI**

No	ID Data	Buying Price	Maintenance Price	Number of Doors	Capacity	Size of Luggage Boot	Estimated Safety	Car Acceptability	Option
2	data 2	vhigh	vhigh	2	2	small	med	unacc	
3	data 3	vhigh	vhigh	2	2	small	high	unacc	
4	data 4	vhigh	vhigh	2	2	med	low	unacc	
5	data 5	vhigh	vhigh	2	2	med	med	unacc	
6	data 6	vhigh	vhigh	2	2	med	high	unacc	
7	data 7	vhigh	vhigh	2	2	big	low	unacc	
8	data 8	vhigh	vhigh	2	2	big	med	unacc	
9	data 9	vhigh	vhigh	2	2	big	high	unacc	
10	data 10	vhigh	vhigh	2	4	small	low	unacc	
11	data 11	vhigh	vhigh	2	4	small	med	unacc	
12	data 12	vhigh	vhigh	2	4	small	high	unacc	
13	data 13	vhigh	vhigh	2	4	med	low	unacc	
14	data 14	vhigh	vhigh	2	4	med	med	unacc	

Gambar 7. Tampilan menu Data Training pada Halaman Admin

Adapun menu Main Process di halaman admin memiliki tampilan yang sama dengan menu Main Process pada halaman utama. Pada menu tersebut, admin juga dapat melakukan proses *mining* menggunakan metode *Random Forest*. Setelah proses dilakukan, hasil pohon keputusan dapat dilihat pada menu Pohon Keputusan. Sedangkan menu Logout merupakan menu untuk keluar dari halaman admin.

## B. Pembahasan

Penerapan metode *Random Forest* untuk melakukan klasifikasi variabel yang mempengaruhi tingkat penerimaan konsumen terhadap mobil dapat dihitung secara langsung. Contoh perhitungan algoritma *gain ratio* terhadap 1 atribut dengan 10 data sampel adalah sebagai berikut :

1. Perhitungan *entropy* total Tingkat Penerimaan Mobil (*Car Acceptability*) untuk kelas *acc*, *unacc*, *good*, *vgood* menggunakan persamaan 1.

$$= (-2/10) \log_2 (2/10) + (-7/10) \log_2 (7/10) + (-1/10) \log_2 (1/10) + (0)$$
$$= 1,1568$$

2. Perhitungan *entropy* pada setiap kelas untuk atribut yang telah diambil secara acak oleh sistem yaitu Biaya Perawatan (*Maintenance Price*), dengan menggunakan rumus yang sama dengan perhitungan *entropy* total pada persamaan 1.

a) Kelas *vhigh*

$$= (0) + (-4/4) \log_2 (-4/4) + (0) + (0) = 0$$

b) Kelas *high*

$$= (0) + (-1/1) \log_2 (-1/1) + (0) + (0) = 0$$

c) Kelas *med*

$$= (-2/3) \log_2 (-2/3) + (-1/3) \log_2 (-1/3) + (0) + (0)$$
$$= 0,9183$$

d) Kelas *low*

$$= (0) + (-1/2) \log_2 (-1/2) + (-1/2) \log_2 (-1/2) + (0) = 1$$

3. Perhitungan *Information gain* menggunakan persamaan 2.

$$= 1,1568 - (((4/10) \times 0) + ((1/10) \times 0) + ((3/10) \times 0,9183) + ((2/10) \times 1))$$
$$= 0,6813$$

4. Perhitungan *Split Information* menggunakan persamaan 3.

$$= (4/10) \log_2 (4/10) + (1/10) \log_2 (1/10) + (3/10) \log_2 (3/10) + (2/10) \log_2 (2/10)$$
$$= 0,5287 + 0,3321 + 0,5210 + 0,4643$$
$$= 1,8461$$

5. Perhitungan *gain ratio* dengan menggunakan persamaan 4.

$$= 0,6813/1,8461$$
$$= 0,3690$$

*Root node* yang didapat dari hasil perhitungan merupakan faktor yang paling mempengaruhi pada tingkat penerimaan konsumen mobil. Variabel yang menempati sebagai *root node* merupakan variabel yang memiliki nilai *gain ratio* tertinggi.

## C. Interpretasi Hasil

Pada penelitian ini, sistem pengklasifikasi menggunakan metode *Random Forest* telah diterapkan pada data tingkat penerimaan konsumen terhadap mobil dan dapat menghasilkan sebuah pohon keputusan. Pohon Keputusan menunjukkan bahwa atribut *Estimated Safety* menempati sebagai *root node* sehingga dapat diketahui bahwa atribut ini

merupakan atribut paling signifikan bagi konsumen dalam menentukan jenis mobil yang akan dipilih kemudian diikuti oleh atribut-atribut lainnya yang menempati sebagai *internal node*. Atribut yang menjadi *root node* dapat dijadikan sebagai pertimbangan produsen mobil dalam proses produksi mobil di masa mendatang.

## IV. PENUTUP

Kesimpulan yang dapat diambil berdasarkan penelitian yang telah dilakukan sebagai berikut:

1. Sistem klasifikasi variabel yang mempengaruhi tingkat penerimaan konsumen terhadap mobil telah selesai dibangun sehingga produsen dapat menjadikan hasil klasifikasi dari sistem untuk proses produksi mobil.
2. Klasifikasi variabel yang mempengaruhi tingkat penerimaan konsumen terhadap mobil dapat dilakukan dengan metode *Random Forest* dengan algoritma *gain ratio*.
3. Pohon keputusan yang dihasilkan dapat berbeda-beda akibat penentuan atribut serta pengambilan data dilakukan secara acak.
4. Variabel yang menempati sebagai *root node* dalam setiap pohon keputusan yang dihasilkan memiliki pengaruh yang paling signifikan pada tingkat penerimaan konsumen terhadap mobil.

## REFERENSI

- [1] Nayazri, G.M. Januari-Agustus, Penjualan Mobil 2016 Lampau 2015. 2016 [cited 2016 20 September]; Available from: <http://otomotif.kompas.com/read/2016/09/14/170346815/januari-agustus.penjualan.mobil.2016.lampau.2015>.
- [2] Dewi, N.K., U.D. Syafitri, and S.Y. Mulyadi, Penerapan Metode Random Forest dalam Driver Analysis. Indonesian Journal of Statistics, 2011. 16(1): p. 35-43.
- [3] Zhou, L. and H. Wang, Loan Default Prediction on Large Imbalanced Data Using Random Forests. TELKOMNIKA Indonesian Journal of Electrical Engineering, 2012. 10(6): p. 1519-1525.
- [4] Purno, P. and P. Praneetpolgrang, Comparisons of Classifier Algorithms: Bayesian Network, C4.5, Decision Forest and NBTree for Course Registration Planning Model of Undergraduate Students, in 2008 IEEE International Conference on Systems, Man and Cybernetics 2008, IEEE. p. 3647-3651.
- [5] Nidhomuddin and B.W. Otok, Random Forest dan Multivariate Adaptive Regression Spline (MARS) Binary Response untuk Klasifikasi Penderita HIV / AIDS di Surabaya. Statistika, 2015. 3(1).
- [6] Litchman, M. Machine Learning Repository. 2013; Available from: <http://archive.ics.uci.edu/ml>.
- [7] Nugroho, Y.S. and D. Gunawan, Decision Tree Induction for Classifying the Cholesterol Levels, in The 2nd International Conference on Science, Technology, and Humanity 2016, Universitas Muhammadiyah Surakarta: Surakarta. p. 231-240.
- [8] Schouten, K., F. Frasincar, and R. Dekker, An Information gain-Driven Feature Study for Aspect-Based Sentiment Analysis. Natural Language Processing and Information Systems, 2016: p. 48-59.
- [9] Alpha P, A., & Oslan, Y. (2015). Program Bantu Pemilihan Pakaian dan Bahan Batik Bagi Konsumen dengan Pendekatan Decision Tree Studi Kasus : Toko InBATIK. Jurnal EKSIS, 8(1), 37-46.
- [10] Barros, R. C., de Carvalho, A. C. P. L. F., & Freitas, A. A. (2015). Decision-Tree Induction. In Automatic Design of Decision-Tree Induction Algorithms (pp. 7-45). Springer International Publishing. <https://doi.org/10.1007/978-3-319-14231-9>.
- [11] Kotsiantis, S. B. (2013). Decision Trees: A Recent Overview. Artificial Intelligence Review, 39(4), 261-283. <https://doi.org/10.1007/s10462-011-9272-4>.