

Identifikasi Faktor-Faktor yang Memengaruhi Waktu Pembelian Kendaraan Kedua Menggunakan Metode CART*

Eka Setiawaty¹, Farit Mochamad Afendi^{2†}, Cici Suhaeni³

¹²³Department of Statistics, IPB University, Indonesia
†corresponding author: fmafendi@apps.ipb.ac.id

Copyright © 2021 Eka Setiawaty, Farit Mochamad Afendi, and Cici Suhaeni. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Increased competition between personal vehicle dealers make them need strategies to hold their customers and increase their sales. One of the strategies they could apply is prospecting their customers at the right time. We could predict the right time by identifying the relationship between the length of their purchase time and its factors based on the transaction data of Z Company from 2002 to 2015 using Classification and Regression Trees (CART). Data analysis is separated between groups of customers who made the second purchase maximum of 10 years after the first purchase (group A) and more than 10 years after the first purchase (group B). Group A's regression tree produces 8 terminal nodes with MAD value 1.84 years. The independent variables that plays a role are tenor, job, age, and brand. Group B's regression tree produces 4 terminal nodes. Authorized service and job come out as independent variables which affect the splitting process. MAD value for Group B's regression tree is 0.56 years.

Keywords: CART, regression tree, vehicle purchase distance.

1. Pendahuluan

1.1 Latar Belakang

Dewasa ini, penggunaan kendaraan roda empat sebagai kendaraan pribadi tidak lagi menjadi hal yang asing dalam kehidupan sehari-hari. Peningkatan jumlah kendaraan roda empat yang beredar di jalan berbanding lurus dengan peningkatan penjualannya dari tahun ke tahun. Meningkatnya penjualan dari tahun ke tahun tersebut merupakan hal yang positif bagi perusahaan otomotif dalam mengembangkan bisnisnya. Namun, hal tersebut berdampak pada meningkatnya persaingan antar perusahaan otomotif yang bergerak di Indonesia. Meningkatnya persaingan antar perusahaan yang terjadi membuat setiap perusahaan perlu merencanakan strategi pemasaran yang efektif.

* Received: Jan 2019; Reviewed: May 2021; Published: May 2021

Salah satu strategi yang dapat diterapkan untuk meningkatkan penjualan, yaitu dengan mempelajari waktu pembelian pelanggan. Perusahaan dapat memprospek pelanggan di waktu yang tepat dengan mempelajari waktu pembelian pelanggan. Penentuan waktu yang tepat dapat didekati dengan mempelajari hubungan antara peubah-peubah penjelas yang berpengaruh terhadap jarak waktu pelanggan dalam melakukan pembelian kendaraan roda empat pertama dan keduanya.

Classification and Regression Trees atau CART adalah metode statistik nonparametrik yang dapat digunakan untuk melihat hubungan antara peubah respon dan peubah penjelas yang ditampilkan dalam hasil akhir berupa struktur pohon. Pohon yang dihasilkan bergantung dari jenis data pada peubah respon. CART menghasilkan pohon klasifikasi ketika peubah respon kategorik dan menghasilkan pohon regresi ketika peubah respon numerik.

Venables dan Ripley (1994) dalam Kurniawan T (2002) menyatakan bahwa penggunaan metode berstruktur pohon relatif jarang digunakan oleh statistikawan meskipun metode ini memudahkan bagi peneliti bidang lainnya karena dapat memberikan lebih banyak informasi dalam pengambilan keputusan. Selain itu, Komalasari (2007) juga menyebutkan bahwa CART dapat melakukan eksplorasi data untuk penyusunan model regresi yang melibatkan banyak peubah penjelas yang kompleks, seperti memiliki dimensi yang besar ataupun terdapat jenis peubah dengan skala campuran, dan dapat menjadi model alternatif ketika terdapat asumsi-asumsi metode regresi yang tidak terpenuhi. Oleh karena itu, pada penelitian ini akan dilakukan analisis untuk mendapatkan faktor-faktor yang berpengaruh terhadap jarak waktu pembelian kendaraan roda empat pertama dan kedua, serta prediksi jarak waktu pembelian menggunakan CART.

1.2 Tujuan Penelitian

Penelitian ini bertujuan untuk mengidentifikasi faktor-faktor yang berpengaruh terhadap jarak waktu pembelian kendaraan roda empat pertama dan kedua, serta memprediksi jarak waktu antara kedua pembelian.

2. Tinjauan Pustaka

2.1 *Classification and Regression Tree*

Classification and Regression Trees atau yang lebih dikenal dengan CART adalah metode statistik nonparametrik yang dapat digunakan untuk melihat hubungan antara peubah respon dan peubah penjelas secara eksploratif. Lewis (2000) menyatakan, karena CART merupakan metode nonparametric, maka CART tidak ada asumsi yang diperlukan terkait dengan peubah penjelasnya. CART dapat menangani peubah penjelas numerik yang menjulur kanan ataupun kiri serta data multimodal se baik menangani peubah penjelas kategorik nominal maupun ordinal. Hal tersebut merupakan suatu keunggulan karena dapat mempermudah peneliti dalam praproses data.

CART merupakan metode penyekatan rekursif biner yang dalam proses analisisnya setiap simpul selalu disekat menjadi dua sekatan yang disebut sebagai

simpul anak (Kurniawan T 2002). CART menampilkan hasil akhir berupa struktur pohon dengan penyekat simpul peubah respon yang berjenis kontinu maupun kategorik. Analisis dengan peubah respon kategorik menghasilkan pohon klasifikasi, sedangkan analisis dengan peubah respon numerik menghasilkan pohon regresi.

Loh (2014) menyatakan bahwa dalam perkembangannya metode pohon regresi dapat menyesuaikan diri dengan hampir seluruh model klasik statistika, seperti metode kuadrat terkecil, kuantil, regresi logistik, Poisson, dan model hazard proporsional. Metode pohon regresi mengenal beberapa istilah, yaitu simpul akar, simpul dalam, dan simpul akhir. Simpul akar, dikenal juga sebagai simpul awal atau simpul induk, merupakan simpul yang mengandung seluruh data. Simpul dalam merupakan anak atau hasil sekatan dari simpul akar yang bukan merupakan simpul akhir. Simpul akhir atau terminal adalah simpul anak yang tidak dapat disekat lagi.

Breiman L (1993) menyatakan bahwa terdapat tiga tahapan penting dalam membentuk pohon regresi, yaitu:

1. Tahapan pertama adalah langkah penyekatan di setiap simpul. Langkah penyekatan simpul perlu memperhatikan beberapa aspek, yaitu:
 - a. Menentukan segala kemungkinan penyekatan berdasarkan setiap peubah penjelas. Ada beberapa aturan dalam menentukan kemungkinan penyekatan, yaitu:
 - i. Setiap penyekatan bergantung pada nilai dari satu peubah penjelas.
 - ii. Ketika X merupakan peubah penjelas kontinu, penyekatan yang diperbolehkan adalah $X \leq c$ dan $X > c$, dengan c adalah nilai rata-rata antara dua nilai amatan peubah X yang berbeda secara berurutan. Jika X memiliki n nilai yang berbeda, maka terdapat $n - 1$ penyekatan.
 - iii. Ketika X merupakan peubah penjelas kategorik, penyekatan berasal dari seluruh kemungkinan penyekatan berdasarkan terbentuknya dua anak gugus yang saling lepas. Jika X merupakan peubah kategorik nominal dengan n kategori berbeda, maka akan terdapat $2^{n-1} - 1$ penyekatan. Namun, jika X merupakan peubah kategorik ordinal yang memiliki n kategori berbeda, maka akan terdapat $n - 1$ penyekatan.
 - b. Memilih penyekatan terbaik dari masing-masing peubah penjelas. Pemilihan penyekat terbaik dari kumpulan penyekatan terbaik yang telah terbentuk sebelumnya. Penyekatan terbaik merupakan penyekatan yang memaksimalkan ukuran kehomogenan di dalam masing-masing simpul anak relatif terhadap simpul induknya dan memaksimalkan fungsi penyekatan antara dua simpul anak. Jumlah kuadrat sisaan pada tiap simpul dapat dijadikan sebagai kriteria kehomogenan pada tiap simpulnya. Misalkan sebuah data terdiri atas $(x_1, y_1), (x_2, y_2), \dots, (x_a, y_a)$. Jika peubah penjelas X memiliki n kategori berbeda atau n nilai berbeda, maka:

$$JKS(t) = \sum_{x_i \in t} [y_{ij} - \underline{y}(t)]^2$$

dengan:

t : simpul atau node, merupakan kumpulan;

x_i : kategori atau nilai dari peubah $X, i = 1, 2, \dots, n$;

y_{ij} : nilai peubah Y saat $X = x_i, j = 1, 2, \dots, a$;

$JKS(t)$: jumlah kuadrat sisaan pada simpul t ;

y_{ij} : nilai y ; dan

$\bar{y}(t)$: rata-rata nilai y pada simpul t . Misalkan s adalah penyekat yang membagi t menjadi simpul anak kanan t_R dan kiri t_L . Fungsi penyekatan yang digunakan adalah:

$$\varphi(s, t) = JKS(t) - \{JKS(t_L) + JKS(t_R)\}$$

Dengan:

$\varphi(s, t)$: fungsi penyekatan simpul t oleh penyekat s ;

$JKS(t)$: jumlah kuadrat sisaan pada simpul t ;

$JKS(t_L)$: jumlah kuadrat sisaan pada simpul t_L ; dan

$JKS(t_R)$: jumlah kuadrat sisaan pada simpul t_R .

Jika s^* adalah penyekat terbaik dari seluruh penyekat s , maka penyekatan terbaik oleh s^* adalah sebagai berikut:

$$\varphi(s^*, t) = \varphi(s, t)$$

dengan:

$\varphi(s^*, t)$: fungsi penyekatan simpul t oleh penyekat s^* ;

$\varphi(s, t)$: fungsi penyekatan simpul t oleh penyekat s^* ; dan

Ω : himpunan seluruh kemungkinan penyekatan.

- c. Mengulangi tahapan penyekatan setelah penyekatan untuk simpul induk telah terbentuk. Tahapan penyekatan diulangi untuk simpul-simpul berikutnya sampai didapatkan simpul terminal.
2. Tahapan kedua yaitu aturan yang menentukan sebuah simpul menjadi simpul terminal. Sebuah simpul ditetapkan sebagai simpul terminal apabila memenuhi kriteria penghentian penyekatan (*stopping rule*). Proses penyekatan simpul dihentikan apabila jumlah amatan dalam simpul berjumlah kurang dari nilai yang ditentukan atau berdasarkan pada ambang batas dari nilai fungsi penyekatan. Pemilihan aturan penghentian berdampak pada pohon yang dihasilkan.
3. Tahapan ketiga yaitu aturan untuk menentukan nilai prediksi ($\hat{y}(t)$) bagi masing-masing simpul terminal. Nilai $\hat{y}(t)$ untuk setiap simpul terminal adalah rata-rata dari nilai $y(t)$ pada setiap simpul terminal tersebut.

3. Metodologi

3.1 Data

Data yang digunakan dalam penelitian ini bersumber dari data penjualan dan servis mobil pada Perusahaan Z sejak tahun 2002 sampai tahun 2015 di area DKI Jakarta. Perusahaan Z memiliki pelanggan sebanyak 299,572 orang selama kurun waktu tersebut. Data yang digunakan merupakan data pelanggan yang melakukan pembelian kendaraan lebih dari satu kali di perusahaan tersebut, yaitu sebanyak 63,890 pelanggan. Peubah respon pada penelitian ini adalah selisih antara waktu pembelian kendaraan pertama dan kedua. Peubah yang digunakan sebagai peubah penjelas terdapat pada Tabel 1.

Tabel 1 Daftar peubah Penjelas

Peubah	Skala	Kategori
Jenis kelamin	Nominal	1 Laki-laki 2 Perempuan
Usia	Kontinu	-
Status	Nominal	1 Belum menikah 2 Menikah 3 Cerai
Pekerjaan	Nominal	1 Wirausaha 2 PNS 3 IRT/Pensiunan 4 Pegawai Swasta 5 Lainnya
Tenor	Kontinu	-
Merek	Nominal	1 A 2 B 3 C 4 D 5 E
Servis resmi	Nominal	0 Tidak pernah 1 Pernah

3.2 Prosedur Analisis Data

Prosedur analisis yang dilakukan dalam penelitian ini yaitu:

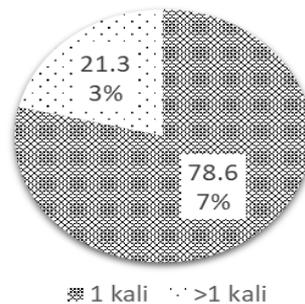
- 1) Menyiapkan data dengan langkah sebagai berikut:
 - a. Menyeleksi transaksi pembeli, hanya transaksi pertama dan kedua yang akan digunakan.
 - b. Membentuk peubah usia (usia pelanggan ketika pertama kali melakukan transaksi pembelian) dari selisih antara tanggal lahir pelanggan dan waktu transaksi pembelian pertama.
 - c. Menggabungkan data penjualan dan data servis untuk membentuk peubah servis.
 - d. Membentuk peubah servis. Peubah servis merupakan peubah biner, bernilai 1 ketika pelanggan pernah melakukan servis di bengkel resmi dan 0 jika tidak.
 - e. Menggabungkan data pembelian pertama dan kedua untuk masing-masing pelanggan.
 - f. Membentuk peubah jarak pembelian dari selisih antara tanggal pembelian pertama dan tanggal pembelian kedua.
 - g. Melengkapi peubah tenor yang kosong dengan menghitung selisih antara tanggal transaksi pembelian dengan tanggal jatuh tempo kredit.
 - h. Menyeleksi data berdasarkan kelengkapan peubahnya.
 - i. Memilih data pelanggan dengan pembelian lebih dari satu kali.
- 2) Melakukan eksplorasi data.
- 3) Berdasarkan eksplorasi data, data peubah respon mengandung pencilan yang berpengaruh terhadap proses penyekatan. Oleh sebab itu, analisis dilakukan terpisah dengan membagi data menjadi 2 kelompok berdasarkan jarak pembelian,

- yaitu kelompok yang memiliki jarak pembelian ≤ 10 tahun (Kelompok A) dan kelompok yang memiliki jarak pembelian 10 tahun (Kelompok B).
- 4) Membagi data menjadi data training 80% dan data testing 20% untuk kedua kelompok data.
 - 5) Melakukan analisis CART pada data *training*.
 - 6) Mengaplikasikan hasil analisis CART pada data *testing*.
 - 7) Menghitung nilai MAD bagi setiap kelompok data dan interpretasi hasil.

4. Hasil dan Pembahasan

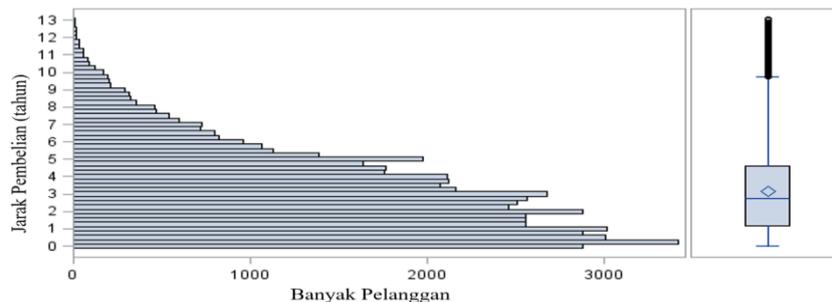
4.1 Gambaran Umum Data

Data yang digunakan pada penelitian ini adalah data penjualan dan data servis Perusahaan Z dari tahun 2002 sampai dengan tahun 2015. Selama kurun waktu tersebut, perusahaan memiliki pelanggan yang berjumlah 299 572 orang. Gambar 1 menampilkan bahwa terdapat 21.3% (63 890) pelanggan yang melakukan pembelian kendaraan lebih dari satu kali di perusahaan tersebut.



Gambar 1. Persentase pelanggan berdasarkan jumlah pembelian

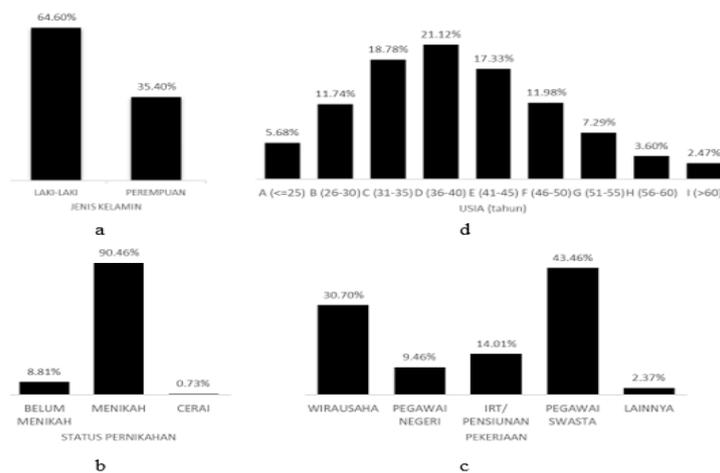
Jarak pembelian merupakan peubah yang dijadikan respon pada penelitian ini. Jarak pembelian memiliki nilai rata-rata 3.15 tahun, median 2.75 tahun, modus 0.003 tahun, dan standar deviasi 2.39 tahun. Gambar 2 menunjukkan bahwa peubah jarak pembelian memiliki banyak pencilon atas dan menjulur ke kanan. Nilai jarak pembelian terkecil yaitu 0.003 tahun. Nilai tersebut sangat kecil mengingat satuan jarak pembelian adalah dalam tahun. Jika dikonversi ke dalam hari, artinya pelanggan tersebut melakukan pembelian kedua satu hari setelah pembelian pertama. Hal tersebut menarik karena di masyarakat, umumnya orang melakukan pembelian kendaraan roda empat kedua dalam jangka waktu beberapa tahun setelah pembelian kendaraan roda empat pertama. Pelanggan yang melakukan pembelian kendaraan roda empat kedua paling lama adalah 13.08 tahun setelah pembelian kendaraan pertama.



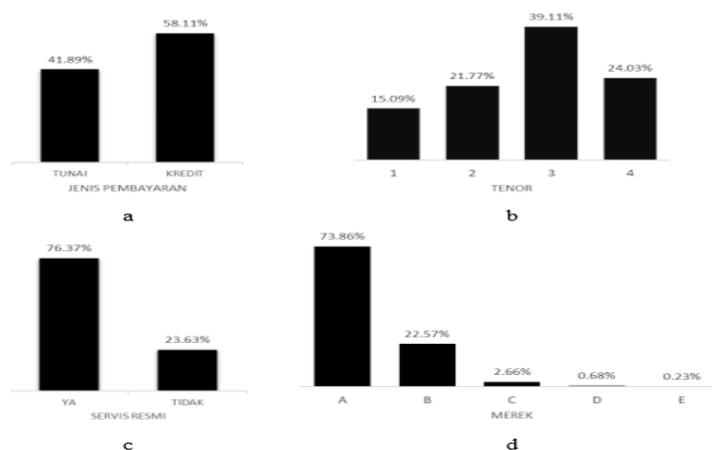
Gambar 2. Diagram batang dan boxplot peubah jarak pembelian

Peubah penjelas yang digunakan dalam analisis adalah karakteristik pelanggan yang diperoleh dari data diri dan data transaksi. Karakteristik pelanggan berdasarkan data diri meliputi jenis kelamin, usia, status pernikahan, dan pekerjaan, sedangkan karakteristik pelanggan berdasarkan data transaksi meliputi merek, jenis pembayaran, dan servis resmi.

Gambar 3 menunjukkan karakteristik pelanggan berdasarkan data dirinya. Gambar (a) menunjukkan karakteristik pelanggan berdasarkan jenis kelaminnya. Pelanggan didominasi oleh pelanggan berjenis kelamin laki-laki. Gambar (b) menunjukkan pelanggan terbanyak adalah mereka yang telah menikah dibandingkan dengan yang belum menikah ataupun yang telah bercerai. Gambar (c) menunjukkan pelanggan terbanyak berasal dari kalangan pegawai swasta dan pada posisi selanjutnya berasal dari pelanggan yang bekerja sebagai wirausaha, ibu rumah tangga (IRT) atau pensiunan, pegawai negeri sipil (PNS), dan pada posisi terakhir dengan jumlah 2.37% ditempati oleh pelanggan dengan jenis pekerjaan lainnya. Gambar (d) menunjukkan data pelanggan berdasarkan rentang usia ketika pertama kali membeli mobil. Gambar (d) menampilkan bahwa pelanggan paling banyak melakukan pembelian pertamanya ketika berusia 31 sampai dengan usia 45 tahun dan paling sedikit ketika berusia lebih dari 60 tahun.



Gambar 3. Karakteristik pelanggan berdasarkan (a) jenis kelamin, (b) status pernikahan, (c) pekerjaan, dan (d) usia



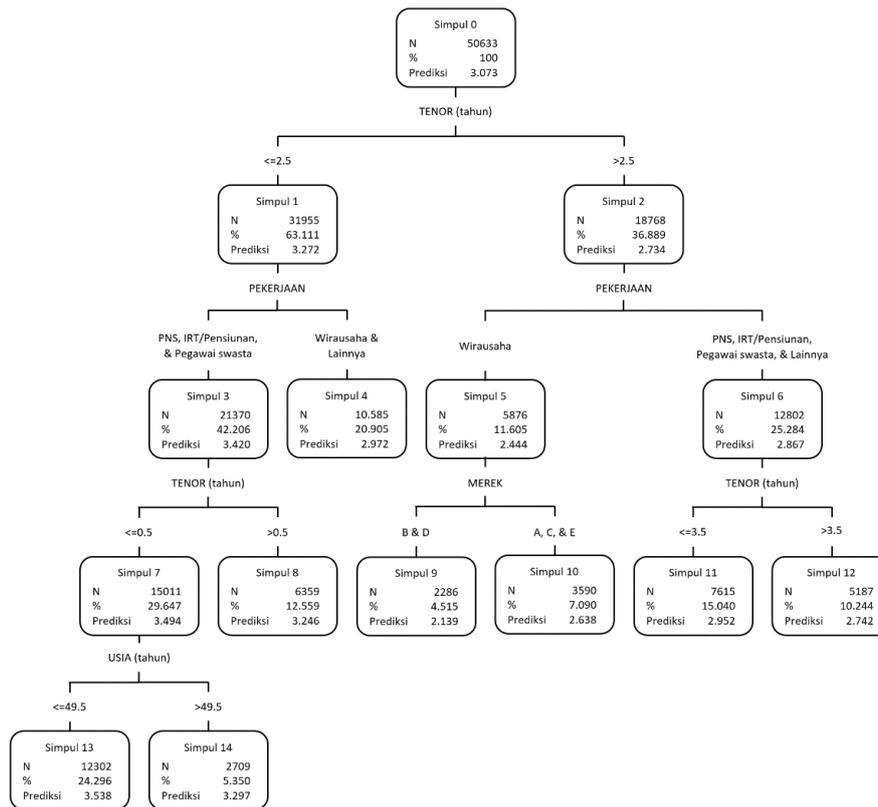
Gambar 4. Karakteristik pelanggan berdasarkan data transaksi: (a) jenis pembayaran, (b) tenor, (c) servis resmi, dan (d) merek

Gambar 4 menunjukkan karakteristik pelanggan berdasarkan data transaksinya. Gambar (a) menunjukkan karakteristik pelanggan berdasarkan jenis pembayaran yang dilakukan ketika membeli mobil pertamanya. Mayoritas pelanggan menggunakan metode pembayaran kredit dibandingkan dengan metode pembayaran tunai. Berdasarkan Gambar (b), dari 37 129 pelanggan yang memilih pembayaran secara kredit, 39.11% pelanggan memilih jangka waktu pembayaran kredit selama tiga tahun. Persentase tersebut merupakan persentase tertinggi di antara pilihan jangka waktu pembayaran kredit lainnya. Gambar (c) menunjukkan karakteristik pelanggan berdasarkan perilaku servisnya. mayoritas pelanggan pernah melakukan servis di bengkel resmi. Gambar (d) menunjukkan pilihan merek yang dilakukan pelanggan saat membeli mobil pertamanya. Berdasarkan gambar tersebut, mayoritas pelanggan (73.86%) memilih mobil merek A sebagai mobil pertamanya. Posisi selanjutnya, ditempati oleh merek B, C, D, dan E secara berurutan.

4.2 Classification and Regression Tree

Classification and Regression Trees atau CART adalah metode nonparametrik yang dapat mempelajari hubungan antara peubah respon dan peubah penjelas. CART dapat menganalisis peubah respon dengan jenis data kategorik maupun numerik. Perbedaannya, CART akan menghasilkan pohon klasifikasi ketika peubah respon kategorik dan pohon regresi ketika peubah respon numerik. Peubah penjelas yang digunakan dalam penelitian ini terdiri atas tujuh peubah, yaitu jenis kelamin, usia, status pernikahan, pekerjaan, merek, tenor, dan servis resmi, sedangkan peubah respon yang digunakan dalam penelitian ini adalah jarak pembelian. Analisis pada penelitian ini terbagi menjadi dua kelompok, yaitu analisis untuk kelompok pelanggan yang melakukan pembelian kedua maksimal sepuluh tahun setelah pembelian pertama (Kelompok A) dan analisis untuk kelompok pelanggan yang melakukan pembelian kedua setelah sepuluh tahun (Kelompok B). Kelompok A terdiri atas 63,267 pelanggan, sedangkan kelompok B terdiri atas 623 pelanggan.

Gambar 5 menunjukkan pohon regresi untuk kelompok A, yaitu kelompok pelanggan yang melakukan pembelian kedua ≤ 10 tahun setelah pembelian pertama. Pohon regresi kelompok A menghasilkan 14 simpul dengan enam simpul dalam dan delapan simpul akhir. Penyekat pertama pohon regresi kelompok A adalah peubah tenor. Peubah tenor merupakan peubah numerik dengan nilai pasti yang terdiri atas: 0, 1, 2, 3, dan 4. Pelanggan dengan tenor = 0 berarti pelanggan yang membayar secara tunai, sedangkan pelanggan dengan tenor > 0 adalah pelanggan yang membayar secara kredit dengan jangka waktu kredit 1, 2, 3, atau 4 tahun. Peubah tenor menyekat simpul induk menjadi simpul 1 dan simpul 2. Kelompok pelanggan yang tenornya > 2.5 tahun (simpul 2) lebih cepat melakukan pembelian kedua dibandingkan kelompok pelanggan yang tenornya ≤ 2.5 tahun (simpul 1) dengan perbandingan jarak 2.734 tahun dan 3.272 tahun. Selanjutnya, masing-masing kelompok pelanggan disekat lagi berdasarkan peubah pekerjaan.



Gambar 5. Pohon regresi pelanggan kelompok A (jarak waktu pembelian ≤ 10 tahun)

Kelompok pelanggan dengan tenor ≤ 2.5 tahun disekat menjadi kelompok pelanggan dengan pekerjaan PNS, IRT/pensiunan, dan pegawai swasta (simpul 3); serta kelompok pelanggan dengan pekerjaan wirausaha dan lainnya (simpul 4). Kelompok pelanggan dengan pekerjaan wirausaha dan lainnya memiliki waktu pembelian kedua yang lebih cepat dibandingkan dengan kelompok pelanggan yang bekerja sebagai PNS, IRT/pensiunan, dan pegawai swasta. Selanjutnya, kelompok pelanggan dengan pekerjaan PNS, IRT/pensiunan, dan pegawai swasta terbagi lagi berdasarkan peubah tenor menjadi kelompok dengan tenor ≤ 0.5 tahun (simpul 7) atau yang membayar secara tunai dan kelompok dengan tenor > 0.5 tahun (simpul 8). Kelompok pelanggan dengan tenor ≤ 0.5 tahun memiliki jarak waktu pembelian yang lebih lama dibandingkan dengan pelanggan yang memiliki jarak waktu tenor > 0.5 tahun. Hal tersebut menunjukkan bahwa semakin lama tenor yang diajukan, maka semakin cepat jarak waktu pembelian kendaraan pelanggan tersebut. Pelanggan-pelanggan dengan tenor ≤ 0.5 tahun disekat kembali berdasarkan peubah usia menjadi kelompok pelanggan yang berusia ≤ 49.5 tahun dengan jarak waktu pembelian 3.538 tahun dan kelompok pelanggan yang berusia > 49.5 tahun dengan jarak waktu pembelian 3.297 tahun.

Kelompok pelanggan dengan tenor > 2.5 tahun terbagi menjadi dua kelompok. Kelompok pertama adalah pelanggan dengan pekerjaan wirausaha (simpul 5) dengan jarak waktu pembelian 2.444 tahun, sedangkan kelompok kedua adalah pelanggan dengan pekerjaan PNS, IRT/pensiunan, pegawai swasta, dan lainnya (simpul 6) dengan jarak waktu pembelian 2.867 tahun. Selanjutnya, kelompok pelanggan dengan pekerjaan wirausaha (simpul 5) terbagi berdasarkan peubah merek menjadi kelompok pelanggan yang membeli merek B dan D, serta kelompok pelanggan yang membeli merek A, C, dan E. Kelompok pelanggan yang membeli merek B dan D

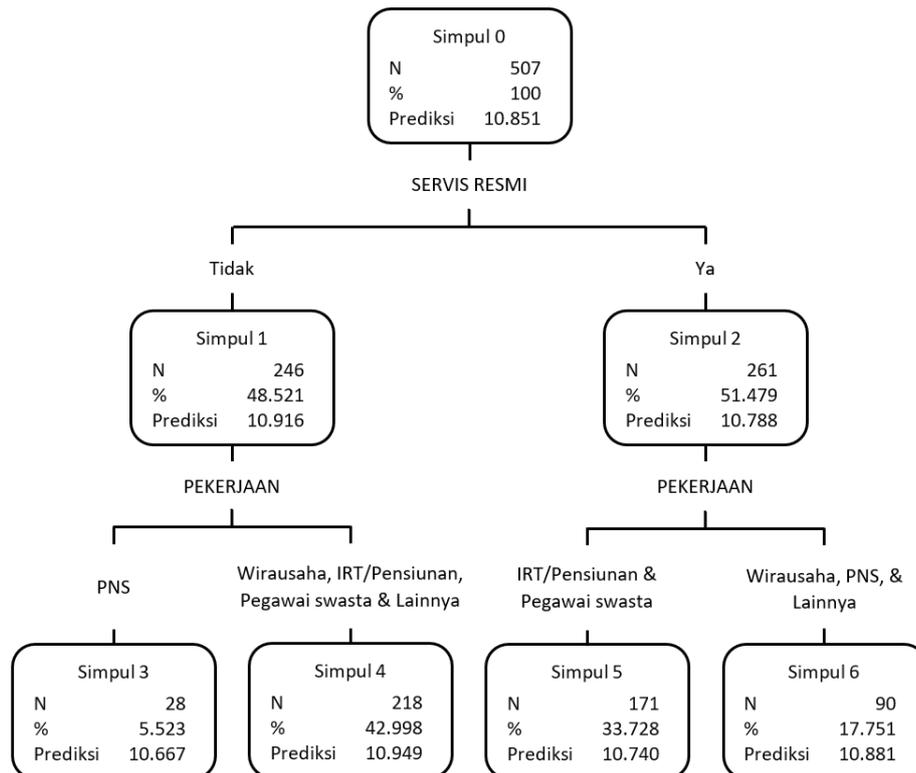
melakukan pembelian kedua lebih cepat dibandingkan dengan kelompok pelanggan yang membeli merek A, C, dan E. Kelompok pelanggan dengan pekerjaan PNS, IRT/pensiunan, pegawai swasta, dan lainnya terbagi lagi berdasarkan peubah tenor menjadi tenor ≤ 3.5 tahun dan tenor > 3.5 tahun.

Ringkasan karakteristik simpul pohon regresi kelompok A berdasarkan peubah pencirinya terdapat pada Tabel II. Kelompok yang paling cepat melakukan pembelian kedua adalah kelompok pelanggan pada simpul ke-9, yaitu kelompok pelanggan dengan karakteristik tenor > 2.5 tahun, bekerja sebagai wirausaha, dan membeli merek B atau D. Kelompok pelanggan tersebut memiliki jarak waktu pembelian 2.139 tahun. Kelompok yang paling lama dalam melakukan pembelian kedua adalah kelompok pelanggan yang termasuk anggota simpul ke-13, yaitu kelompok pelanggan dengan ciri-ciri tenor ≤ 0.5 tahun, bekerja sebagai PNS, IRT/pensiunan, atau pegawai swasta, dan berusia ≤ 49.5 tahun. Berdasarkan hasil prediksi nilai jarak pembelian pada pohon regresi, diperoleh nilai MAD sebesar 1.84. Artinya, nilai rata-rata kesalahan prediksi dari metode CART untuk data ini sebesar 1.84 tahun.

Pohon regresi kelompok B, kelompok pelanggan yang melakukan pembelian kedua > 10 tahun setelah pembelian pertama, ditunjukkan oleh Gambar 6. Pohon regresi tersebut memiliki enam simpul dengan dua simpul dalam dan empat simpul akhir.

Tabel II
KARAKTERISTIK SIMPUL TERMINAL BERDASARKAN HASIL POHON
REGRESI PELANGGAN KELOMPOK A

No simpul	Jumlah pelanggan	Persentase	Prediksi	Peubah penciri
4	10585	20.905	2.972	Tenor ≤ 2.5 tahun, Pekerjaan: wirausaha atau lainnya
8	6359	12.559	3.246	$0.5 < \text{Tenor} \leq 2.5$ tahun, Pekerjaan: PNS, IRT/pensiunan, atau pegawai swasta
9	2286	4.515	2.139	Tenor > 2.5 tahun, Pekerjaan: wirausaha, Merek: B atau D
10	3590	7.09	2.638	Tenor > 2.5 tahun, Pekerjaan: wirausaha, Merek: A, C, atau E
11	7615	15.04	2.952	$2.5 < \text{Tenor} \leq 3.5$ tahun, Pekerjaan: PNS, IRT/pensiunan, pegawai swasta, atau lainnya
12	5187	10.244	2742	Tenor > 3.5 tahun, Pekerjaan: PNS, IRT/pensiunan, pegawai swasta, atau lainnya
13	12302	24.296	3.538	Tenor ≤ 0.5 , Pekerjaan: PNS, IRT/pensiunan, atau pegawai swasta, Usia ≤ 49.5 tahun
14	2709	5.35	3.297	Tenor ≤ 0.5 , Pekerjaan: PNS, pegawai swasta, Usia > 49.5 tahun



Gambar 6. Pohon regresi pelanggan kelompok B (jarak waktu pembelian > 10 tahun)

Pohon regresi kelompok B pertama kali disekat oleh peubah servis resmi. Pelanggan yang tidak pernah melakukan servis resmi (simpul 1) memiliki jarak waktu servis 10.916 tahun, sedangkan pelanggan yang pernah servis di servis resmi memiliki jarak waktu servis 10.788 tahun. Artinya, pelanggan yang pernah melakukan servis di tempat resmi melakukan pembelian kedua lebih cepat dibandingkan dengan yang tidak pernah.

Penyekatan kedua dilakukan berdasarkan peubah pekerjaan. Kelompok pelanggan yang tidak pernah melakukan servis resmi terbagi menjadi kelompok pelanggan yang bekerja sebagai PNS dan kelompok pelanggan yang bekerja sebagai wirausaha, IRT/pensiunan, pegawai swasta, dan lainnya dengan jarak waktu pembelian 10.667 tahun dan 10.949 tahun. Kelompok pelanggan yang pernah melakukan servis resmi terbagi menjadi kelompok pelanggan yang bekerja sebagai IRT/pensiunan dan pegawai swasta, serta kelompok pelanggan yang bekerja sebagai wirausaha, PNS, dan lainnya dengan jarak waktu pembelian 10.740 tahun dan 10.881 tahun.

Ringkasan karakteristik simpul akhir pohon regresi kelompok B terdapat pada Tabel III. Kelompok pelanggan yang paling cepat melakukan pembelian kedua adalah kelompok pelanggan anggota simpul ke-3 dengan jarak waktu pembelian 10.667 tahun, yaitu pelanggan dengan yang tidak pernah melakukan servis resmi dan bekerja sebagai PNS. Kelompok pelanggan yang paling lama melakukan pembelian kedua adalah kelompok pelanggan yang memiliki ciri-ciri tidak pernah servis resmi dan bekerja sebagai wirausaha, IRT/pensiunan, pegawai swasta, dan lainnya dengan jarak waktu pembelian 10.949 tahun. Nilai MAD pohon regresi kelompok B adalah 0.56. Hal tersebut menunjukkan bahwa nilai kesalahan prediksi pohon regresi kelompok B

adalah 0.56 tahun.

Tabel III
KARAKTERISTIK SIMPUL TERMINAL BERDASARKAN HASIL POHON
REGRESI PELANGGAN KELOMPOK B

No simpul	Jumlah pelanggan	Persentase	Prediksi	Peubah penciri
3	28	5.523	10.667	Tidak pernah servis resmi, Pekerjaan: PNS
4	218	42.908	10.949	Tidak pernah servis resmi, Pekerjaan: pegawai swasta, wirausaha, IRT/pensiunan, atau lainnya
5	171	33.728	10.74	Pernah servis resmi, Pekerjaan: IRT/pensiunan atau pegawai swasta
6	90	17.751	10.881	Pernah servis resmi, Pekerjaan: wirausaha, PNS, atau lainnya

5. Simpulan

Pohon regresi kelompok A memiliki delapan simpul akhir dengan peubah penyekat tenor, pekerjaan, usia, dan merek. Kelompok pelanggan yang melakukan pembelian kedua tercepat, 2.139 tahun, adalah kelompok pelanggan dengan karakteristik tenor > 2.5 tahun, bekerja sebagai wirausaha, dan membeli merek B atau D. Kelompok pelanggan yang melakukan pembelian kedua paling lama, 3.538 tahun, adalah kelompok pelanggan dengan ciri-ciri tenor ≤ 0.5 tahun, bekerja sebagai PNS, IRT/pensiunan, atau pegawai swasta, dan berusia ≤ 49.5 tahun. Rata-rata nilai kesalahan pohon regresi kelompok A adalah 1.84 tahun.

Pohon regresi kelompok B memiliki empat simpul akhir dengan peubah penyekat servis resmi dan pekerjaan. Kelompok pelanggan yang melakukan pembelian kedua tercepat, 10.667 tahun, yaitu kelompok pelanggan dengan ciri-ciri tidak pernah servis resmi dan bekerja sebagai PNS. Kelompok pelanggan yang paling lama melakukan pembelian kedua, 10.949 tahun, adalah kelompok pelanggan dengan ciri-ciri tidak pernah servis resmi dan bekerja sebagai wirausaha, IRT/pensiunan, pegawai swasta, dan lainnya. Rata-rata nilai kesalahan pohon regresi kelompok B adalah 0.56 tahun.

Daftar Pustaka

- Breiman L, J.H. Friedman, R. O. d. C. S. (1993). Classification and Regression Trees. New York(US): Chapman and Hall.
- Komalasari, W. (2007). Metode pohon regresi untuk eksploratori data dengan peubah yang banyak dan kompleks. *Informatika Pertanian* 16(1), 967–980.
- Kurniawan T, Aunuddin, W. Y. (2002). Penerapan metode pemangkasan dalam cart (classification and regression tree). *Forum Statistika dan Komputasi* 7(2), 19–25.
- Loh, W. (2014). Fifty years of classification and regression trees. *International Statistical Review* 82(3), 329–348.