

## **KLASIFIKASI KARAKTERISTIK KECELAKAAN LALU LINTAS DI KOTA DENPASAR DENGAN PENDEKATAN *CLASSIFICATION AND REGRESSION TREES (CART)***

I Gede Agus Jiwadiana<sup>§1</sup>, I Komang Gde Sukarsa<sup>2</sup>, I Gusti Ayu Made Srinadi<sup>3</sup>

<sup>1</sup>Jurusan Matematika, Fakultas MIPA - Universitas Udayana [Email: sakjiwa@yahoo.co.id]

<sup>2</sup>Jurusan Matematika, Fakultas MIPA - Universitas Udayana [Email: sukarsakomang@yahoo.co.id]

<sup>3</sup>Jurusan Matematika, Fakultas MIPA - Universitas Udayana [Email: srinadiigustiayumade@yahoo.co.id]

<sup>§</sup>Corresponding Author

### **ABSTRACT**

*The aim of this research is to determine the classification characteristics of traffic accidents in Denpasar city in January-July 2014 by using Classification And Regression Trees (CART). Then, for determine the explanatory variables into the main classifier of CART. The result showed that optimum CART generate three terminal node. First terminal node, there are 12 people were classified as heavy traffic accident characteristics with single accident, and second terminal nodes, there are 68 people were classified as minor traffic accident characteristics by type of traffic accident front-rear, front-front, front-side, pedestrians, side-side and location of traffic accident in district road and sub-district road. For third terminal node, there are 291 people were classified as medium traffic accident characteristics by type of traffic accident front-rear, front-front, front-side, pedestrians, side-side and location of traffic accident in municipality road and explanatory variables into the main splitter to make of CART is type of traffic accident with maximum homogeneity measure of 0.03252.*

*Keywords: CART, traffic accident, terminal node*

### **1. PENDAHULUAN**

Umumnya masalah klasifikasi diselesaikan dengan menggunakan metode regresi logistik dan analisis diskriminan (Lusyanti [1]). Analisis diskriminan memerlukan asumsi multivariate normal dan varians kovarians sama, sedangkan metode regresi logistik tidak memerlukan kedua asumsi tersebut. Metode ini memerlukan data yang lengkap, dan sensitif terhadap outlier.

Salah satu metode alternatif masalah klasifikasi yang lebih longgar atau tidak terikat oleh asumsi-asumsi dan tidak berbentuk probabilitas adalah metode klasifikasi berstruktur pohon yang diperkenalkan oleh Breimen, L., Friedmen, J., Olshen, R., dan Stone, C. pada tahun 1984 yaitu *Classification and Regression Trees (CART)*. Metode CART merupakan alat yang digunakan dalam eksplorasi data nonparametrik yang dapat

digunakan untuk melihat hubungan antara variabel terikat dengan variabel bebas yang berukuran besar dan kompleks (Pratiwi dan Zain [2]). Variabel bersifat kompleks dapat berupa dimensinya yang besar atau jenis variabel bebasnya campuran, misalnya kontinu dan kategorik, baik nominal maupun ordinal. Jika variabel terikatnya berupa variabel kontinu maka akan diperoleh model pohon regresi, jika variabel terikatnya kategorik maka akan diperoleh model pohon klasifikasi.

Model CART dikenal karena kesederhanaan dan efisiensi ketika berhadapan dengan kasus yang jumlah datanya besar. CART diperoleh menggunakan pembagian cepat yang secara rekursif mempartisi data menjadi sub yang lebih kecil.

Penelitian menggunakan metode CART sudah pernah dilakukan oleh Suniantara (2008) serta Pratiwi dan Zain [2]. Suniantara (2008)

menerapkan metode CART dalam bidang kesehatan yaitu memprediksi berat badan bayi lahir. Penelitian ini menghasilkan model pohon regresi karena variabel terikatnya bersifat kontinu. Pratiwi dan Zain [2] meneliti tentang klasifikasi pengangguran terbuka menggunakan CART di Provinsi Sulawesi Utara.

Salah satu kasus yang terkait dengan metode CART adalah karakteristik kecelakaan lalu lintas. Pihak kepolisian telah mengumpulkan data kecelakaan lalu lintas setiap harinya sehingga menghasilkan data dalam jumlah yang besar untuk setiap bulannya. Data tersebut berupa data bertipe kategorik. Penelitian tentang kecelakaan lalu lintas pernah dilakukan oleh Afidah [3]. Masalah yang muncul dalam kecelakaan lalu lintas ini adalah pada karakteristik kecelakaan lalu lintas. Karakteristik kecelakaan lalu lintas ini dibagi menjadi tiga kelompok yaitu ringan, sedang, dan berat. Ketiga kelompok tersebut dipengaruhi oleh empat faktor yaitu faktor manusia, kendaraan, jalan, dan alam (lingkungan) (Afidah [3]). Perlu cara untuk mengetahui faktor-faktor apakah yang dapat memengaruhi karakteristik kecelakaan lalu lintas tersebut, sehingga pihak kepolisian dapat mengambil tindakan yang dapat mengurangi kondisi korban. Penelitian tentang kecelakaan lalu lintas pernah dilakukan oleh Afidah [3], meneliti pola tingkat keparahan korban lalu lintas dengan menggunakan regresi logistik multinomial.

Permasalahan dalam penelitian adalah bagaimana klasifikasi karakteristik dan variabel apa yang menjadi pemilah utama dalam penentuan klasifikasi karakteristik kecelakaan lalu lintas di Kota Denpasar dari bulan Januari sampai Juli 2014?

Tujuan penelitian adalah untuk mengetahui klasifikasi karakteristik dan variabel yang menjadi pemilah utama dalam penentuan klasifikasi karakteristik kecelakaan lalu lintas di Kota Denpasar dari bulan Januari sampai Juli 2014.

## 2. METODE PENELITIAN

Penelitian ini menggunakan data sekunder yang diperoleh dari POLRESTA (Kepolisian Resor Kota) tentang data Kecelakaan Lalu Lintas di Kota Denpasar dari bulan Januari sampai bulan Juli 2014.

Gambaran umum variabel terikat dan variabel bebas dalam penelitian dapat dilihat pada Tabel 1.

Tabel 1. Gambaran Umum Variabel Terikat dan Variabel Bebas

No	Peubah	Jenis	Kategori
1	Karakteristik kecelakaan lalu lintas (variabel terikat)	Diskrit	1. Ringan 2. Sedang 3. Berat
2	Jenis kecelakaan lalu lintas (variabel bebas)	Diskrit	1. Tunggal 2. Samping-samping 3. Depan-samping 4. Depan-depan 5. Depan-belakang 6. Pejalan kaki
3	Waktu (variabel bebas)	Diskrit	1. Lalu lintas padat (antara pukul 06.00 WITA – 08.00 WITA, antara pukul 12.00 WITA – 13.30 WITA, antara pukul 16.00 WITA – 18.00 WITA) 2. Lalu lintas sepi (selain waktu padat)
4	Lokasi Status jalan (variabel bebas)	Diskrit	1. Nasional 2. Provinsi 3. Kabupaten 4. Kodya 5. Kecamatan
5	Jenis kelamin (variabel bebas)	Diskrit	1. Laki-laki 2. Perempuan
6	Umur (variabel bebas)	Kontinu	

Penelitian ini dilaksanakan dengan metode studi kasus, yaitu dilakukan dengan menerapkan teori untuk menganalisis data dengan menggunakan program CART Pro EX V6.0.

Langkah-langkah yang dilakukan dalam penelitian ini adalah:

- 1) Penentuan pemilah dan pemilahan secara rekursif pada simpul dengan menggunakan pemilah Indeks Gini yaitu:

$$i(t) = \sum_{j \neq i} p(j/t)p(i/t)$$

dengan  $i(t)$  adalah nilai indeks gini,  $p(j/t)$  adalah proporsi kelas  $j$  pada simpul  $t$ , dan  $p(i/t)$  adalah proporsi kelas  $i$  pada simpul  $t$  (Timofeev [4]). Penetapan sebuah simpul sebagai simpul terminal, jika banyak data pada simpul tersebut kurang atau sama dengan 5.

- 2) Penandaan label kelas simpul yang ditentukan berdasarkan jumlah kelas terbanyak pada simpul.
- 3) Pemangkasan CART yang dilakukan untuk mendapatkan pohon yang minimum dengan menggunakan persamaan berikut:

$$R_\alpha(T) = R(T) + \alpha |\overline{T}|$$

dengan  $R(T)$  adalah *Resubstitution Estimate* (Proporsi kesalahan pada sub pohon  $T$ ,  $\alpha$  adalah kompleksitas parameter, dan  $|\overline{T}|$  adalah jumlah simpul terminal pada pohon  $T$  (Breimen [5]).

- 4) Pemilihan CART optimal dengan menggunakan *Cross Validation V-Fold Estimate*. Amatan dalam  $L$  dibagi menjadi  $V$  bagian yang saling lepas dengan ukuran kurang lebih sama besar ditiap kelasnya. *Learning sample* ke- $v$  dengan  $v=1,2,\dots,V$  digunakan untuk membentuk pohon  $T_k^{(v)}$ . Misal  $d^{(v)}(x)$  adalah hasil pengklasifikasian, maka penduga sampel uji untuk  $R^{ts}(T_k^{(v)})$  adalah:

$$R^{ts}(T_k^{(v)}) = \frac{1}{N_v} \sum_{(x_n, j_n) \in L_v} X(d^{(v)}(x_n \neq j_n))$$

Dengan menggunakan amatan induk  $L$  untuk membentuk deretan pohon  $\{T_k\}$ , maka penduga validasi silang lipat  $V$  untuk  $T_k^{(v)}$  adalah:

$$R^{cv}(T_k) = \frac{1}{V} \sum_{v=1}^V R^{ts}(T_k^{(v)})$$

Pohon klasifikasi optimum dipilih  $T_k^*$

dengan  $R^{cv}(T_k^*) = \min_k R^{cv}(T_k)$ .

### 3 HASIL DAN PEMBAHASAN

Objek orang pada penelitian ini adalah data kecelakaan lalu lintas di Kota Denpasar dari Bulan Januari sampai Bulan Juli 2014. Variabel terikat pada penelitian ini adalah karakteristik kecelakaan lalu lintas, sedangkan variabel bebasnya adalah jenis laka (kecelakaan lalu lintas), waktu, lokasi laka, jenis kelamin, dan umur. Dari 371 data kecelakaan lalu lintas di Kota Denpasar selama tahun 2014, diperoleh gambaran karakteristik kecelakaan lalu lintas yaitu kecelakaan ringan 24% atau 87 orang, kecelakaan sedang sebesar 53% atau 198 orang, dan kecelakaan berat sebesar 23% atau 86 orang.

#### 3.1 Pembentukan CART Awal

CART awal dibentuk dengan menggunakan metode pemilahan Indeks Gini. Metode ini memisahkan kelas yang anggota kelasnya terbesar lebih dahulu atau yang merupakan kelas terpenting dalam simpul tersebut. Berikut ini, diberikan kemungkinan pemilah pada masing-masing variabel bebas:

- 1) Jenis laka ( $x_1$ ) dengan 6 kategori nominal, kemungkinan pemilah  $2^{6-1} - 1 = 31$
- 2) Waktu ( $x_2$ ) dengan 2 kategori nominal, mempunyai kemungkinan pemilah  $2^{2-1} - 1 = 1$
- 3) Lokasi status jalan ( $x_3$ ) dengan 5 kategori nominal, mempunyai kemungkinan pemilah  $2^{5-1} - 1 = 15$
- 4) Jenis kelamin ( $x_4$ ) dengan 2 kategori nominal, mempunyai kemungkinan pemilah  $2^{2-1} - 1 = 1$

Umur ( $x_5$ ) merupakan data kontinu. Dari 371 orang, terdapat  $n$  nilai amatan berbeda, maka terdapat  $n-1$  kemungkinan pemilah.

Pemilah yang memberikan nilai indeks gini tertinggi adalah pemilah terbaik yang digunakan

sebagai pemilah pertama dalam pembentukan pohon. Dari lima variabel bebas, pemilahan pertama kali dilakukan terhadap simpul akar/utama yang didasarkan pada jenis laka dengan nilai indeks gini 0,03252. Nilai indeks gini masing-masing variabel ditampilkan dalam Tabel 2.

Tabel 2. Nilai Indeks Gini

	Variabel bebas	Pemisah	Indeks gini	N kiri	N kanan
1	Jenis Laka	depan-belakang,	0,03252	359	12
		depan-depan,			
		depan-samping,			
		pejalan kaki,			
		samping-samping			
2	Lokasi Laka	Jalan Kabupaten,	0,02304	277	94
		Jalan Kecamatan,			
		Jalan Kodya			
3	Jenis Kelamin	Perempuan	0,01267	123	248
4	Umur	39.50000	0,00566	270	101
5	Waktu	Padat	0,00554	97	274

Peubah jenis laka terpilih sebagai pemilah utama karena memberikan nilai indeks gini tertinggi dari variabel lain. Dengan kata lain bahwa, variabel bebas jenis laka merupakan peubah utama yang berperan penting dalam pembentukan CART.

Pemilahan ini memisahkan simpul utama menjadi kelompok jenis kecelakaan depan-belakang, depan-depan, depan-samping, pejalan-kaki, samping-samping, dan kelompok jenis kecelakaan tunggal. Hasil pemilahan pada simpul ini, jumlah objek sebelah kiri sebanyak 359 orang dan sebelah kanan sebanyak 12 orang. Dari 359 orang tersebut, terdapat karakteristik kecelakaan berat sebesar 74 orang, kecelakaan ringan 87 orang, dan kecelakaan sedang 198

orang. Sedangkan dari 12 orang pada simpul kanan tersebut hanya terdapat 12 orang dengan karakteristik kecelakaan berat. Selanjutnya simpul dua dipilah berdasarkan lokasi laka, dan seterusnya muncul variabel-variabel bebas jenis kelamin, umur, dan waktu laka. Secara umum semua variabel bebas muncul sebagai pemilah.

Penghentian pembentukan CART dilakukan ketika banyaknya amatan pada simpul kurang dari 5 atau dalam simpul terdapat amatan yang homogen. Maka didapatkan CART awal dengan 42 simpul terminal.

### 3.2 Pemangkasan Pohon

Hasil dari pembentukan CART dengan metode pemilahan Indeks Gini berukuran sangat besar yaitu 42 simpul. Ukuran pohon yang besar ini dapat menimbulkan *overfitting*. Untuk mengatasinya maka dicari pohon dugaan yang layak dengan memangkas pohon tersebut.

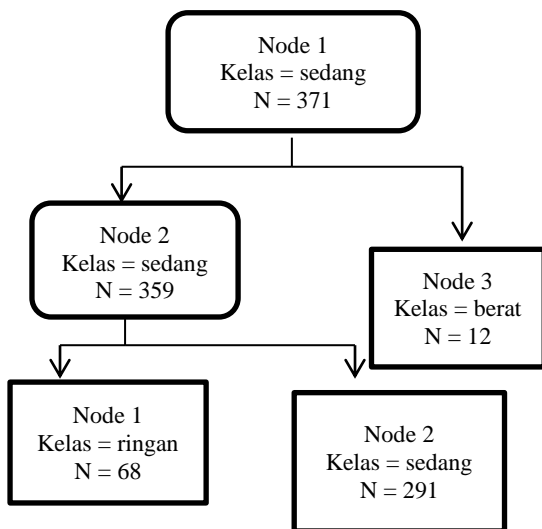
Pemangkasan CART merupakan suatu upaya untuk menentukan ukuran CART yang terbaik bagi CART yang terbentuk. Proses pemangkasan pohon dilakukan apabila memenuhi persamaan  $R(t) = R(t_{left}) + R(t_{right})$ , maka  $t_{right}$  dan  $t_{left}$  dipangkas. Berdasarkan perhitungan program yang digunakan diperoleh 18 sub pohon yang dihasilkan dari proses pemangkasan.

Setelah CART dipangkas, ditentukan CART optimum menggunakan penduga silang lipat  $V$  (*Cross Validation V-Fold Estimate*). CART optimum didapatkan berdasarkan nilai  $R^{cv}$  minimum yaitu  $0,83540 \pm 0,03440$  atau  $0,80100 \leq R^{cv} \leq 0,86980$  dengan parameter kompleksitas  $\alpha = 0,020682$ . Berdasarkan kriteria ini, didapatkan CART optimum dengan tiga simpul akhir. Letak pemangkasan CART tersebut dapat dilihat pada Tabel 3.

Tabel 3. Letak Pemangkasan CART

Pohon	Jumlah simpul terminal	Cross-Validated Relative Cost (biaya kesalahan)	Resubstitution Relative Cost (penduga pengganti)	kompleksitas parameter
1	42	0,91886 +/- 0,04079	0,48357	0,000000
11	15	0,87560 +/- 0,04126	0,59757	0,005648
12	14	0,89038 +/- 0,04101	0,60893	0,007584
13	13	0,89474 +/- 0,04107	0,62069	0,007851
14	9	0,85753 +/- 0,04151	0,67477	0,009023
15	8	0,85646 +/- 0,04143	0,68998	0,010150
16	7	0,85240 +/- 0,04070	0,71065	0,013791
17	6	0,85310 +/- 0,04058	0,73220	0,014379
18	5	0,84997 +/- 0,03953	0,75824	0,017367
19**	3	0,83540 +/- 0,03440	0,82025	0,020682
20	1	1,00000 +/- 0,00003	1,00000	0,059926

Berdasarkan  $R^{cv}$  minimum sebesar 0,83540+/-0,03440 diperoleh CART optimum dari proses pemangkasan dengan simpul-simpul terminal yang dihasilkan dan digambarkan pada Gambar 1.



Gambar 1. CART Optimum dengan Tiga Simpul Terminal

Dalam proses pemangkasan berdasarkan  $R^{cv}$  minimum, didapatkan CART dengan tiga simpul terminal. Dengan peubah yang masuk dalam CART tersebut adalah jenis laka dan lokasi laka. Jenis laka menjadi pemilah utama dalam pembentukan CART. Ketiga simpul terminal yang dihasilkan dapat diuraikan sebagai berikut:

1. Simpul terminal pertama, terdiri dari 12 orang yang dikelompokkan sebagai karakteristik kecelakaan lalu lintas berat dengan jenis laka tunggal.
2. Simpul terminal kedua, terdiri dari 68 orang yang dikelompokkan sebagai karakteristik kecelakaan lalu lintas ringan dengan jenis laka depan-belakang, depan-depan, depan-samping, pejalan kaki, samping-samping dan lokasi laka jalan Kabupaten, jalan Kecamatan.
3. Simpul terminal ketiga, terdiri dari 291 orang yang dikelompokkan sebagai karakteristik kecelakaan lalu lintas sedang dengan jenis laka depan-belakang, depan-depan, depan-samping, pejalan kaki, samping-samping dan lokasi laka jalan Kodya.

#### 4. KESIMPULAN

Berdasarkan hasil dan pembahasan dapat disimpulkan bahwa CART optimum yang terbentuk menghasilkan tiga simpul terminal yaitu simpul terminal pertama, terdiri dari 12 orang yang dikelompokkan sebagai karakteristik kecelakaan lalu lintas berat dengan jenis laka tunggal, simpul terminal kedua, terdiri dari 68 orang yang dikelompokkan sebagai karakteristik kecelakaan lalu lintas ringan dengan jenis laka depan-belakang, depan-depan, depan-samping, pejalan kaki, samping-samping dan lokasi laka jalan Kabupaten, jalan Kecamatan, serta Simpul

terminal ketiga, terdiri dari 291 orang yang dikelompokkan sebagai karakteristik kecelakaan lalu lintas sedang dengan jenis laka depan-belakang, depan-depan, depan-samping, pejalan kaki, samping-samping dan lokasi laka jalan kodya dan variabel bebas yang menjadi pemilah utama dalam pembentukan CART adalah jenis laka dengan nilai indeks gini 0,03252.

#### Daftar Pustaka

- [1] Lusyanti, Merlina. 2010. Perbandingan Metode Regresi Logistik Dengan Metode Pohon Klasifikasi Pada Data Polikotomus (Studi Kasus Pada Faktor Yang Berpengaruh Terhadap Keberhasilan Pengobatan Akupuntur Pada Penderita Obesitas Di LP3A Surabaya). *Skripsi*. Surabaya: Institut Teknologi Sepuluh Nopember (tidak dipublikasikan).
- [2] Pratiwi, F. E., dan Zain, I. 2014. Klasifikasi Pengangguran Terbuka Menggunakan CART (Classification and Regression Tree) di Provinsi Sulawesi Utara. *Sains dan Seni Pomit*, Jurusan Statistika, FMIPA, Institut Teknologi Sepuluh Nopember.
- [3] Afidah, L. N. 2011. Pola Tingkat Keparahan Korban Kecelakaan Lalu Lintas dengan Menggunakan Regresi Logistik Multinomial (Studi Kasus Kecelakaan Lalu Lintas di Surabaya). *Skripsi*. Surabaya: Institut Teknologi Sepuluh Nopember (tidak dipublikasikan).
- [4] Timofeev, R. 2004. *Classification and Regression Trees (CART) Theory and Applications*. Berlin: Center of Applied Statistics and Economics Humboldt University.
- [5] Breimen, L. F. 1984. *Classification and Regression Tree*. New York: Chapman And Hall.