

Classification the Characteristics of Traffic Accident Victims in Pariaman Using the Chi-square Automatic Interaction Detection Algorithm

Manja Danova Putri, Dina Fitria*, Yenni Kurniawati, Zilrahmi

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

*Corresponding author: dinafitria@fmipa.unp.ac.id

Submitted : 24 Oktober 2023

Revised : 17 Januari 2024

Accepted : 25 Januari 2024

ABSTRACT

Traffic accidents are incidents that occur when motor vehicles collide on the road, resulting in damage to vehicles and road infrastructure, as well as the potential for material loss, injuries, physical damage, and even death for those involved. From the recorded data of traffic accidents in the Police Report of Resort Police Pariaman, the characteristics of traffic accidents can be observed based on the time of occurrence, the vehicles involved, the types of accidents, the age of the victims, gender, occupation, possession of a Driver's License (SIM), and the role of the victims at the time of the accident. To determine the characteristics that describe this issue, an approach with classification techniques can be used. Chi-Square Automatic Interaction Detection (CHAID) is one of the methods in classification techniques. In this study, CHAID was conducted to classify the characteristics of traffic accidents in Kota Pariaman. The results of the study showed that the age of the victims and the types of accidents are the most significant variables that influence the condition of traffic accident victims. The evaluation results using the confusion matrix showed an accuracy rate of 91%. This indicates that the model can classify the data well.

Keywords: CHAID, Classification, Decision Tree, Traffic Accident.



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Kecelakaan lalu lintas merupakan peristiwa yang terjadi ketika kendaraan bermotor mengalami tabrakan atau insiden lainnya di jalan raya yang mengakibatkan rusaknya kendaraan dan infrastruktur jalan serta dapat menyebabkan cedera, kerusakan fisik, bahkan kematian bagi pengemudi, penumpang, pejalan kaki, atau orang yang terlibat dalam kecelakaan tersebut. Menurut World Health Organization (2018) kecelakaan lalu lintas termasuk salah satu permasalahan serius secara global, hal ini diakibatkan oleh tingginya angka kematian akibat kecelakaan lalu lintas yang mencapai angka tertinggi sebesar 1,35 juta jiwa pada tahun 2016. Angka tersebut menempati urutan kedelapan penyebab kematian di dunia untuk semua kelompok umur, dan menjadi penyebab utama kematian pada kalangan usia 5-29 tahun.

Indonesia menempati posisi keempat sebagai negara dengan jumlah korban kecelakaan lalu lintas tertinggi di kawasan Asia Tenggara. Pada tahun tahun 2010-2020 jumlah korban kecelakaan lalu lintas berkisar antara 147.798-197.560 jiwa, dengan korban meninggal dunia didominasi oleh kalangan usia 15-34 tahun (Hidayati dkk., 2022). Tingginya jumlah korban akibat kecelakaan lalu lintas akan berdampak terhadap berbagai aspek kehidupan. Cedera dan kerusakan fisik akibat kecelakaan lalu lintas dapat mengubah kehidupan seseorang secara drastis (Sabet dkk., 2016). Hal ini akan mempengaruhi kemampuan seseorang untuk bergerak, bekerja, dan menjalankan aktivitas sehari-hari sehingga memerlukan bantuan perawatan jangka panjang. Selain itu trauma, gangguan kecemasan, atau depresi yang mungkin terjadi dapat mempengaruhi kesehatan mental, kualitas hidup dan interaksi sosial seseorang. Jika dilihat dari segi ekonomi tingginya jumlah korban akibat kecelakaan lalu lintas akan memberikan kerugian yang cukup besar terhadap perekonomian keluarga maupun terhadap sistem perekonomian nasional (Suseno, 2022). Banyaknya jumlah anggota keluarga yang meninggal dunia maupun mengalami cacat permanen pada usia produktif, akan memberikan dampak pada tingginya potensi suatu keluarga kehilangan tulang punggung dalam mencari nafkah sehingga rentan mengalami kemiskinan. Dengan demikian sangat penting melakukan evaluasi untuk mengenali

karakteristik yang mempengaruhi kondisi korban kecelakaan lalu lintas. Salah satu metode yang dapat digunakan adalah metode dengan teknik klasifikasi.

Klasifikasi merupakan teknik pengelompokan objek/data ke dalam kelas atau kategori yang telah ditentukan sebelumnya berdasarkan atribut atau fitur-fitur yang ada (Yuliasari dkk., 2021). Tujuan dari teknik ini adalah untuk mengenali pola atau karakteristik tertentu dalam data yang dapat digunakan untuk memprediksi atau mengidentifikasi kelas atau kategori yang sesuai. *Chi-Square Automatic Interaction Detection* (CHAID) merupakan salah satu metode dalam teknik klasifikasi yang digunakan untuk membangun model klasifikasi berbasis pohon keputusan (*decision tree*). CHAID membagi rangkaian data menjadi subgrup-subgrup berdasarkan hubungan antara variabel dependen dengan variabel independen (Hasna dan Kunto, 2006). Metode ini menggunakan uji *chi-square* untuk mengukur hubungannya atau signifikansi antara kedua variabel tersebut. Jika uji *chi-square* menunjukkan hasil yang signifikan, maka pemisahan populasi dilakukan dengan membuat cabang-cabang pada pohon keputusan berdasarkan variabel independen yang paling signifikan. Proses ini dilakukan secara berulang sampai tidak ada lagi variabel independen yang signifikan atau sampel dalam setiap kelompok sudah homogen terhadap variabel dependen.

Penerapan klasifikasi CHAID telah dilakukan oleh para peneliti sebelumnya dengan menggunakan objek penelitian yang berbeda-beda. CHAID pernah digunakan untuk mengklasifikasikan karakteristik yang mempengaruhi pengangguran di Provinsi Banten tahun 2022 (Fikham dan Achmad, 2023). Hasil penelitian tersebut menunjukkan bahwa status dalam rumah tangga merupakan variabel yang paling signifikan dalam mempengaruhi pengangguran di Provinsi tersebut. Selain itu CHAID juga diaplikasikan untuk mengklasifikasikan penyakit *Tuberculosis Relapse* dan ditemukan bahwa variabel yang paling signifikan dalam mempengaruhi penyakit tersebut adalah tipe diagnosis dan umur pasien (Santi dkk., 2022). Dalam artikel ini CHAID digunakan untuk mengklasifikasikan karakteristik yang mempengaruhi kondisi korban pasca peristiwa kecelakaan lalu lintas. Dari hasil klasifikasi akan diperoleh gambaran mengenai karakteristik-karakteristik apa saja yang berperan dalam mempengaruhi kondisi korban setelah mengalami kecelakaan lalu lintas di kota Pariaman pada tahun 2022.

II. METODE PENELITIAN

A. Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang bersumber dari Laporan Polisi kasus kecelakaan lalu lintas di tingkat Kepolisian Resor (POLRES) Kota Pariaman tahun 2022. Adapun variabel yang digunakan dalam penelitian ini diambil berdasarkan data korban kecelakaan lalu lintas yang terdapat dalam laporan tersebut yang meliputi, kondisi korban (Y), waktu kejadian (X_1), kendaraan korban (X_2), jenis kecelakaan (X_3), usia korban (X_4), jenis kelamin (X_5), jenis pekerjaan (X_6), kepemilikan SIM (X_7), dan peran korban (X_8).

B. Teknik Analisis Data

Dalam artikel ini analisis data dilakukan menggunakan metode CHAID dengan bantuan *SPSS* dan *R Studio*. Adapun tahapan analisis dalam penelitian ini adalah sebagai berikut:

1. Analisis Statistika Deskriptif.

Analisis statistika deskriptif adalah metode yang digunakan untuk menggambarkan dan meringkas data secara numerik. Tujuan utamanya adalah untuk memberikan pemahaman yang jelas mengenai pola, karakteristik, dan sifat data yang diamati. Analisis ini melibatkan penggunaan berbagai teknik statistik untuk menyajikan, mengorganisir, serta menggambarkan data sehingga dapat diinterpretasikan dengan lebih baik.

2. Data *Training* dan Data *Testing*.

Data *training* adalah himpunan data yang digunakan untuk melatih atau membangun model. Data ini digunakan untuk menemukan pola data sehingga nantinya dapat digunakan untuk memprediksi data baru. Sedangkan data *testing* merupakan himpunan data yang digunakan untuk menguji model setelah proses pelatihan menggunakan data *training*. Data *training* dan *testing* dibandingkan untuk memeriksa apakah model akhir yang digunakan bekerja dengan benar.

3. Uji *Chi-Square*

Uji *chi-square* merupakan teknik dalam analisis statistik yang digunakan untuk menguji asosiasi atau hubungan antara dua variabel.

Hipotesis dalam pengujian *chi-square*:

H_0 : $n_{ij} = n_i n_j$ (tidak terdapat hubungan antara kategori i pada variabel dependen dengan kategori j pada variabel independen).

H_1 : $n_{ij} \neq n_i n_j$ (terdapat hubungan antara kategori i pada variabel dependen dengan kategori j pada variabel independen).

Statistik uji dalam pengujian chi-square:

$$x^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \text{ dimana } E_{ij} = \frac{n_i \cdot n_j}{n} \quad (1)$$

Keterangan :

- n_{ij} : banyaknya pengamatan pada baris ke-i dan kolom ke-j
- E_{ij} : nilai harapan pengamatan pada baris ke-i dan kolom ke-j
- n_i : total banyaknya pengamatan pada baris ke-i
- n_j : total banyaknya pengamatan pada kolom ke-j
- n : total banyaknya pengamatan

Dasar pengambilan keputusan dalam uji *chi-square* adalah tolak H_0 jika nilai $X^2_{hitung} > X^2_{tabel}$ atau $p - value < \alpha$.

4. Koreksi Bonferroni

Koreksi *bonferroni* merupakan proses koreksi yang digunakan ketika melakukan serangkaian pengujian hipotesis secara bersamaan (simultan). Tujuan dari koreksi ini adalah untuk mengontrol tingkat kesalahan tipe I yaitu, menolak hipotesis nol meskipun sebenarnya hipotesis tersebut benar. Dengan menerapkan koreksi *bonferroni* dapat mengurangi risiko terjadinya kesalahan tersebut dengan menyesuaikan tingkat signifikansi yang digunakan dalam setiap pengujian hipotesis. Koreksi *bonferroni* dilakukan setelah penggabungan dan diperolehnya variabel independen yang signifikan. Nilai *bonferroni* dapat dihitung menggunakan persamaan 2.

$$\alpha' = \frac{\alpha}{M} \quad (2)$$

Keterangan:

- α : Tingkat signifikansi awal
- α' : Tingkat signifikansi yang disesuaikan setelah koreksi *bonferroni*
- M : Jumlah total uji hipotesis yang dilakukan

5. Algoritma CHAID

Algoritma CHAID bertujuan untuk membentuk diagram pohon CHAID yang dapat digunakan untuk memprediksi nilai dari variabel dependen. Secara garis besar algoritma CHAID dibagi dalam tiga tahap, yaitu tahap penggabungan (*merging*) tahan pemisahan (*splitting*), dan tahap penghentian (*stopping*) (Baron and Phillips, 1994).

a. Tahap Penggabungan

Pada tahap ini akan diperiksa signifikansi dari masing-masing kategori variabel independen terhadap variabel dependen menggunakan uji *chi-square*. Penggabungan ini dilakukan terhadap kategori-kategori variabel independen yang tidak signifikan untuk digabungkan menjadi satu kategori.

b. Tahap Pemisahan

Tahap ini dilakukan untuk memilih variabel independen mana yang akan digunakan sebagai pemisah *node* atau *split node* terbaik. Pemisahan dilakukan dengan membandingkan nilai *chi-square* atau *p-value* pada masing-masing variabel independen. Variabel independen yang digunakan adalah variabel independen terbaik (yang paling signifikan), yaitu variabel independen dengan nilai *chi-square* terbesar atau *p-value* terkecil.

c. Tahap Penghentian

Tahap ini dilakukan untuk menghentikan proses pembentukan pohon keputusan. Pertumbuhan pohon harus terhenti jika tidak ada lagi variabel independen yang signifikan secara statistik.

6. Evaluasi Model

Untuk mengukur performansi ketepatan kinerja dari hasil klasifikasi umumnya dilakukan dengan menggunakan matriks konfusi (*Confusion Matrix*). Matriks ini menggambarkan seberapa baik *classifier* dalam mengklasifikasikan data ke dalam kelas yang benar serta seberapa sering model membuat kesalahan dalam proses klasifikasinya. Tabel *Confusion Matrix* dapat dilihat pada Tabel 1.

Tabel 1. Confusion Matrix

		Actual Value	
		Positif	Negatif
Prediction Value	Positif	TP (True Positif)	FP (False Positif)
	Negatif	FN (False Negatif)	TN (True Negatif)

Keterangan

TP : merupakan banyaknya kasus yang diprediksi benar sebagai positif oleh model dan memang benar positif dalam kenyataan.

TN : merupakan banyaknya kasus yang diprediksi benar sebagai negatif oleh model dan memang benar negatif dalam kenyataan.
 FP : merupakan banyaknya kasus yang diprediksi positif oleh model tetapi sebenarnya negatif dalam kenyataan.
 FN : merupakan banyaknya kasus yang diprediksi negatif oleh model tetapi sebenarnya positif dalam kenyataan.
 Berdasarkan tabel *Confusion Matrix* diatas maka dapat dihitung nilai akurasi.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

III. HASIL DAN PEMBAHASAN

A. Statistika Deskriptif

Analisis statistika deskriptif dilakukan untuk melihat frekuensi pada masing-masing kategori dalam setiap variabel pada data korban kecelakaan lalu lintas Kota Pariaman tahun 2022. Hasil tersebut dapat dilihat pada Tabel 2.

Tabel 2. Statistika Deskriptif

Variabel	Kategori	Frekuensi	Variabel	Kategori	Frekuensi
Kondisi Korban (Y)	Luka ringan	311	Usia Korban (X4)	Balita (0-5 tahun)	22
	Meninggal dunia	27		Kanak-kanak (6-11 tahun)	18
Waktu Kejadian (X1)	Dini hari (00.00-05.00)	6		Remaja (12-25 tahun)	128
	Pagi hari (05.01-11.00)	76		Dewasa (26-45 tahun)	80
	Siang hari (11.01-15.00)	72	Lansia (>45 tahun)	90	
	Sore hari (15.01-19.00)	86	Jenis Kelamin (X5)	Laki-laki	223
Malam hari (19.01-23.59)	98	Perempuan		115	
Kendaraan Korban (X2)	Mobil penumpang	9	Jenis Pekerjaan (X6)	PNS/TNI/POLRI	14
	Mobil barang	7		Wiraswasta	85
	Mobil bus	1	Petani/Nelayan	24	
	Sepeda motor	272	Ibu rumah tangga	55	
Jenis Kecelakaan (X3)	Pejalan kaki	49	Kepemilikan SIM (X7)	Pelajar/Mahasiswa	108
	Tabrak depan-depan	127		Lainnya	52
	Tabrak depan-belakang	29	Ada	70	
	Tabrak depan-samping	114	Tidak ada	268	
Peran Korban (X8)	Tabrak samping-samping	19	Peran Korban (X8)	Pengemudi	183
	Menabrak pejalan kaki	49		Penumpang	106
			Pejalan kaki	49	

Berdasarkan Tabel 2 dapat dilihat bahwa korban yang mengalami kecelakaan lalu lintas di kota pariaman pada tahun 2022 sebagian besar mengalami luka ringan. Dari jumlah keseluruhan korban 92% diantaranya adalah korban luka ringan, sedangkan 8% lainnya meninggal dunia. Peristiwa ini cenderung terjadi dimalam hari yaitu pada pukul 19.01-23.59 WIB, dimana dalam peristiwa tersebut korban sebagian besar mengendarai sepeda motor. Jenis kecelakaan yang paling banyak terjadi adalah tabrak depan-depan, dengan posisi kendaraan bergerak dari arah yang berlawanan dan saling bertabrakan pada bagian depan. Korban kecelakaan lalu lintas didominasi oleh kalangan usia remaja dengan rentang usia 12-25 tahun. Hal ini juga ditunjukkan dengan sebagian besar dari korban adalah seorang pelajar/mahasiswa yang usianya berkisar diantara rentang 12-25 tahun tersebut. Secara keseluruhan korban kecelakaan lalu lintas didominasi oleh jenis kelamin laki-laki. Selain itu mayoritas dari korban tidak memiliki Surat Izin Mengemudi, meskipun kebanyakan diantara mereka berperan sebagai pengemudi kendaraan bermotor.

B. Data Training dan Data Testing

Dataset dibagi menjadi data *training* dan data *testing* menggunakan *R Studio* dengan persentase 80:20. Untuk memperoleh hasil yang konsisten pembagian ini dilakukan menggunakan fungsi *set.seed* dan fungsi *sample* dari *library caret*. Dengan demikian, dari total 338 amatan diperoleh data *training* sebanyak 271 amatan dan data *testing* sebanyak 67 amatan.

C. Diagram Pohon CHAID

Tahap pertama untuk membentuk diagram pohon CHAID adalah melakukan uji *chi-square* untuk mengidentifikasi nilai signifikansi dari masing-masing variabel independen terhadap variabel dependen. uji ini akan dilakukan secara berulang sampai tidak ditemukan lagi variabel independen yang signifikan. Nilai *chi-square* untuk masing-masing variabel independen dapat dilihat pada Tabel 3.

Tabel 3. Uji *Chi-Square* Pertama

Variabel	X^2_{hitung}	X^2_{tabel}	Keterangan
Kondisi Korban * Waktu Kejadian	2,904	9,488	Tidak signifikan
Kondisi Korban * Kendaraan Korban	1,318	9,488	Tidak signifikan
Kondisi Korban * Jenis Kecelakaan	12,094	9,488	Signifikan
Kondisi Korban * Usia Korban	14,027	9,488	Signifikan
Kondisi Korban * Jenis Kelamin	0,004	3,841	Tidak signifikan
Kondisi Korban * Jenis Pekerjaan	7,484	12,592	Tidak Signifikan
Kondisi Korban * Kepemilikan SIM	5,374	3,841	Signifikan
Kondisi Korban * Peran Korban	5,711	5,991	Tidak signifikan

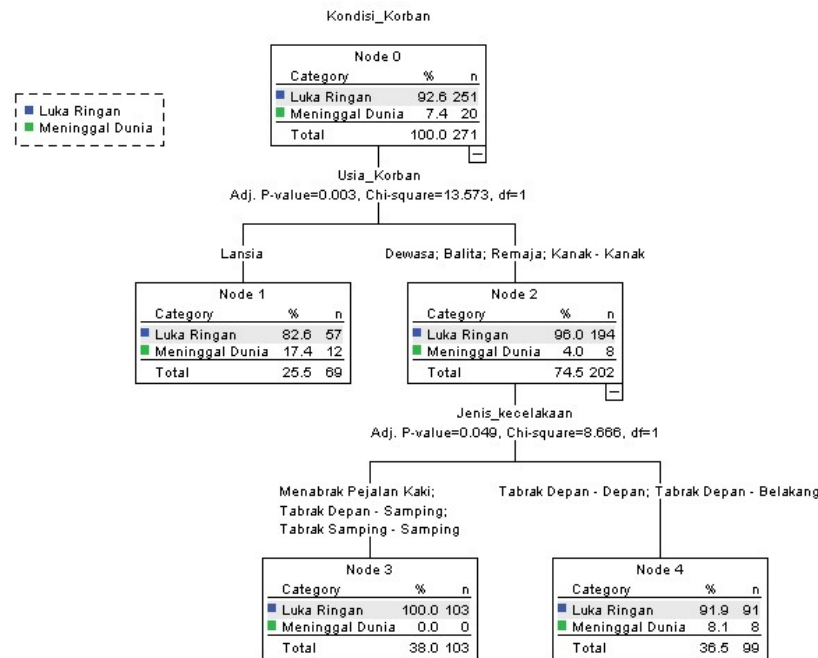
Berdasarkan hasil uji *chi-square* pada Tabel 3 diperoleh bahwa variabel usia korban merupakan variabel yang paling signifikan terhadap kondisi korban. Hal ini dapat dilihat dari hasil uji *chi-square* yang menunjukkan bahwa variabel usia korban memiliki nilai *chi-square* tertinggi diantara variabel independen lainnya, yakni sebesar 12,094. Oleh karena itu variabel usia korban dijadikan sebagai variabel splitting atau variabel pemisah pada simpul akar.

Tabel 4. Uji *Chi-Square* Kedua

Variabel	X^2_{hitung}	X^2_{tabel}	Keterangan
Kondisi Korban * Waktu Kejadian	5,521	9,488	Tidak signifikan
Kondisi Korban * Kendaraan Korban	3,907	9,488	Tidak signifikan
Kondisi Korban * Jenis Kecelakaan	9,438	9,488	Tidak signifikan
Kondisi Korban * Jenis Kelamin	0,251	3,841	Tidak signifikan
Kondisi Korban * Jenis Pekerjaan	5,298	12,592	Tidak signifikan
Kondisi Korban * Kepemilikan SIM	2,388	3,841	Tidak signifikan
Kondisi Korban * Peran Korban	4,543	5,991	Tidak signifikan

Berdasarkan hasil uji *chi-square* kedua pada tabel 4 tidak ditemukan adanya variabel independen yang memiliki nilai *chi-square* yang signifikan terhadap kondisi korban. Oleh karena itu dilakukan penggabungan kategori terhadap variabel jenis kecelakaan karena memiliki nilai *chi-square* tertinggi diantara variabel lainnya. Setelah dilakukan penggabungan diperoleh nilai X^2_{hitung} sebesar 8,666 dan nilai X^2_{tabel} menjadi 3,841. Karena $X^2_{hitung} > X^2_{tabel}$, sehingga variabel jenis kecelakaan dijadikan sebagai pemisah untuk variabel usia korban. Dengan demikian karena tidak terdapat lagi variabel independen yang memiliki nilai *chi-square* yang signifikan terhadap variabel dependen maka diagram pohon CHAID dapat dilihat pada Gambar 1.

Berdasarkan Gambar 1 terlihat bahwa pada variabel usia korban terjadi penggabungan terhadap kategori balita, remaja, kanak-kanak dan dewasa yang digabungkan menjadi satu kategori. Sehingga kategori dalam variabel tersebut diringkas menjadi 2 kelompok yaitu, kategori baita, remaja, kanak-kanak, dewasa dengan rentang usia 0-45 tahun dan kategori lansia dengan rentang usia >45 tahun. Begitupun pada variabel jenis kecelakaan, terjadi penggabungan terhadap kategori tabrak depan-samping, tabrak samping-samping, dan menabrak pejalan kaki. Sementara itu kategori tabrak depan-depan dan tabrak depan belakang juga digabungkan menjadi satu kategori. Dengan demikian variabel jenis kecelakaan juga diringkas menjadi 2 kelompok kategori. kelompok pertama terdiri dari tabrak depan-samping, tabrak samping-samoing dan menabrak pejalan kaki, sedangkan kelompok kedua terdiri dari tabrak depan-depan dan tabrak depan-belakang.



Gambar 1. Diagram Pohon CHAID

Diagram pohon tersebut terdiri dari 5 node dan memiliki 2 kedalaman (*depth*). Terdapat node 0 yang berfungsi sebagai simpul induk. Pada kedalaman pertama terdapat node 1 dan node 2. Pada kedalaman ini, node 1 menjadi terminal node karena tidak ada lagi percabangan dari node tersebut. Sementara itu node 2 memiliki cabang yang terdiri dari node 3 dan node 4 yang juga merupakan terminal node. Dengan jumlah terminal node sebanyak 3 maka diagram pohon tersebut dapat menghasilkan 3 segmentasi karakteristik dari korban kecelakaan lalu lintas. Segmentasi tersebut ditampilkan pada Tabel 5 dan Tabel 6.

Tabel 5. Segmentasi Karakteristik Korban Kecelakaan Lalu Lintas

Segmen	Karakteristik
1	Korban adalah lansia dengan usia >45 tahun.
2	Korban adalah balita, kanak-kanak, remaja, dan dewasa (usia 0-45 tahun) dengan jenis kecelakaan tabrak depan-samping, tabrak samping-samping, dan menabrak pejalan kaki.
3	Korban adalah Korban adalah balita, remaja, kanak-kanak, dan dewasa (usia 0-45 tahun) dengan jenis kecelakaan tabrak depan-depan dan tabrak depan belakang.

Tabel 6. Presentase Setiap Segmen Kondisi Korban kecelakaan Lalu Lintas

Segmen	Luka Ringan		Meninggal Dunia	
	Jumlah	Persentase	Jumlah	Persentase
1	57	82,6%	12	17,4%
2	103	100%	0	0
3	91	91,9%	8	8,1%

Berdasarkan tabel 5 dan tabel 6 dapat dilihat bahwa persentase terbesar dari korban kecelakaan lalu lintas yang mengalami luka ringan berada pada segmen kedua. Pada segmen tersebut korban adalah balita, kanak-kanak, remaja dan dewasa yang usianya 0-45 tahun dengan jenis kecelakaan yang terjadi adalah tabrak depan-samping, tabrak samping-samping, dan menabrak pejalan kaki. Artinya ketika korban kecelakaan lalu lintas di kota pariaman pada tahun 2022 adalah orang yang berusia 0-45 tahun dan jenis kecelakaan yang dialami adalah tabrak depan-samping, tabrak samping-samping, dan menabrak pejalan kaki, maka 100% korban tersebut mengalami luka ringan.

Sedangkan persentase terbesar dari korban kecelakaan lalu lintas yang meninggal dunia berada pada segmen pertama, yaitu lansia yang berusia diatas 45 tahun. Hal ini berarti bahwa sebagian besar korban kecelakaan lalu lintas yang meninggal dunia adalah lansia yang usianya diatas 45 tahun. Dalam kasus 20 orang korban meninggal dunia

akibat kecelakaan 12 diantaranya berusia di atas 45 tahun, sementara 8 orang lainnya berusia antara 0-45 tahun. Selain dari 12 orang lansia yang meninggal dunia tersebut juga terdapat 57 orang lainnya yang mengalami luka ringan.

Pada segmen 3 korban kecelakaan lalu lintas adalah balita, kanak-kanak, remaja, dan dewasa yang berusia diantara 0-45 tahun dengan jenis kecelakaan yang terjadi adalah tabrak depan-depan dan tabrak depan-belakang. Jumlah korban dalam segmen ini mencapai 99 orang. Dari jumlah tersebut 91 orang mengalami luka ringan, sementara 8 orang lainnya meninggal dunia.

D. Evaluasi Model

Evaluasi hasil klasifikasi dilakukan menggunakan bantuan *Python*. Berikut adalah nilai dari confusion matrix dapat dilihat pada tabel 5.

Tabel 5. *Confusion Matrix*

Prediction	Reference	
	Luka Ringan	Meninggal Dunia
Luka Ringan	59	1
Meninggal Dunia	5	2

Berdasarkan nilai dari *confusion matrix* pada tabel 5 dapat disimpulkan bahwa jumlah korban yang benar diprediksi luka ringan oleh model adalah sebanyak 59 orang, sedangkan korban yang diprediksi benar meninggal dunia oleh model adalah sebanyak 2 orang. Namun terdapat sebanyak 5 orang yang diprediksi meninggal dunia oleh model namun luka ringan pada kenyataannya, dan sebanyak 1 orang yang diprediksi luka ringan oleh model namun meninggal dunia pada kenyataannya. Hasil ini dapat dikatakan bahwa model berhasil mengklasifikasikan korban dengan baik. Hal ini juga dapat dilihat berdasarkan nilai akurasi dari hasil klasifikasi. Berdasarkan nilai dari *confusion matrix* tersebut diperoleh nilai akurasi sebesar 0,91, sehingga dapat dikatakan bahwa model dapat mengklasifikasikan data dengan baik, dengan ketepatan kinerja dari hasil klasifikasi mencapai 91%.

IV. KESIMPULAN

Berdasarkan hasil analisis yang dilakukan menggunakan metode *Chi-Square Automatic Interaction Detection* (CHAID) diperoleh hasil bahwa karakteristik utama yang paling signifikan dalam mempengaruhi kondisi korban pasca peristiwa kecelakaan lalu lintas adalah usia korban dan jenis kecelakaan. Dengan menggunakan *confusion matrix*, diperoleh nilai akurasi sebesar 91%. Hal ini menunjukkan bahwa model mampu melakukan klasifikasi data secara keseluruhan dengan baik.

DAFTAR PUSTAKA

- Baron, S., & Phillips, D. (1994). Attitude Survey Data Reduction Using CHAID: An Example in Shopping Centre Market Research. *Journal of Marketing Management*, 10(1-3), 75-88.
- Fikham, F. A. H., & Achmad, A. I. (2023). Penerapan Metode Chi-Squared Automatic Interaction Detection pada Segmentasi Karakteristik yang Mempengaruhi Pengangguran di Provinsi Banten Tahun 2022. *Bandung Conference Series: Statistics*, 3(2), 585-593.
- Hasna, S. N., & Kunto, Y. S. (2006). Analisis CHAID Sebagai Alat Bantu Statistika Untuk Segmentasi Pasar (Studi Kasus Pada Koperasi Syariah Al-Hidayah). *Jurnal Manajemen Pemasaran*, 1(2), 88-97.
- Hidayati, T. S., Siswanto, J., Hadi, S., & Ayu, B. P. S. B. R. (2022). Pendidikan Karakter SALUD (Sadar Lalu Lintas Dini).
- Sabet, F. P., Tabrizi, K. N., Khankeh, H. R., Saadat, S., Abedi, H. A., & Bastami, A. (2016). Road Traffic Accident Victims, Experiences of Return to Normal Life: A Qualitative Study. *Iran Red Crescent Medical Journal*, 18(4).
- Santi, V. M., Nafisah, L., & Meidianingsih, Q. (2022). Penerapan Metode SMOTE CHAID Dalam Klasifikasi Tuberkulosis Relapse. *Jurnal Statistika Dan Aplikasinya*, 6(1), 26-36.
- Suseno, H. (2022). Kecelakaan Menghambat Tumbuh dan Berkembang Suatu Bangsa. *Jurnal Indonesia Sosial Sains*, 3(9), 1345-1352.

World Health Organization. (2018). *Global Status Report On Road Safety*.

Yuliasari, P. D., Goejantoro, R., & Amijaya, F. D. T. (2021). Klasifikasi Rumah Tanga Miskin di Kecamatan Kaubun Tahun 2020 dengan Menggunakan Metode Improved Chi-Square Automatic Interaction. *Ekspensial*, 12(1), 83–92.