

**PEMODELAN KLASIFIKASI CHAID DENGAN PENDEKATAN
SMOTE UNTUK MENANGANI DATA TIDAK SEIMBANG
PADA KASUS KECELAKAAN LALU LINTAS KOTA
PARIAMAN**

SKRIPSI

Diajukan Sebagai Salah Satu Persyaratan Guna Memperoleh Gelar Sarjana

Statistika



Oleh
MANJA DANOVA PUTRI
NIM. 19337055

**PROGRAM STUDI SARJANA STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI PADANG
2024**

PERSETUJUAN SKRIPSI

PEMODELAN KLASIFIKASI CHAID DENGAN PENDEKATAN SMOTE UNTUK MENANGANI DATA TIDAK SEIMBANG PADA KASUS KECELAKAAN LALU LINTAS KOTA PARIAMAN

Nama : Manja Danova Putri
NIM : 19337055
Program Studi : S1 Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Padang, 31 Januari 2024

Mengetahui:
Ketua Departemen Statistika



Dr. Yenni Kurniawati, S.Si., M.Si
NIP. 198402232010122005

Disetujui Oleh:
Pembimbing



Dina Fitria, S.Pd., M.Si
NIP. 172019

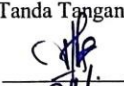


PENGESAHAN LULUS UJIAN SKRIPSI

Nama : Manja Danova Putri
NIM : 19337055
Program Studi : S1 Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

PEMODELAN KLASIFIKASI CHAID DENGAN PENDEKATAN SMOTE UNTUK MENANGANI DATA TIDAK SEIMBANG PADA KASUS KECELAKAAN LALU LINTAS KOTA PARIAMAN

Dinyatakan lulus setelah dipertahankan di depan Tim Penguji Skripsi
Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Padang

Padang, 31 Januari 2024

Tim Penguji	Nama	Tanda Tangan
Ketua	: Dina Fitria, S.Pd., M.Si	
Anggota	: Dr. Yenni Kurniawati, S.Si., M.Si	
Anggota	: Zilrahmi, S.Pd., M.Si	

SURAT PERNYATAAN TIDAK PLAGIAT

Saya yang bertandatangan di bawah ini:

Nama : Manja Danova Putri
NIM : 19337055
Program Studi : S1 Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan bahwa, skripsi saya dengan judul **“Pemodelan Klasifikasi CHAID dengan Pendekatan SMOTE untuk Menangani Data Tidak Seimbang Pada Kasus Kecelakaan Lalu Lintas Kota Pariaman”** adalah benar merupakan hasil karya saya dan bukan merupakan plagiat dari karya orang lain atau pengutipan dengan cara-cara yang tidak sesuai dengan etika yang berlaku dalam tradisi keilmuan.

Apabila suatu saat terbukti saya melakukan plagiat maka saya bersedia diproses dan menerima sanksi akademis maupun hukum sesuai dengan hukum dan ketentuan yang berlaku, baik di institusi UNP maupun di masyarakat dan negara.

Demikian pernyataan ini saya buat dengan penuh kesadaran rasa tanggung jawab sebagai anggota masyarakat ilmiah.

Diketahui oleh,
Ketua Departemen Statistika,



Dr. Yenni Kurniawati, S.Si., M.Si
NIP. 198402232010122005

Saya yang menyatakan,



Manja Danova Putri
NIM. 19337055

Pemodelan Klasifikasi CHAID Dengan Pendekatan SMOTE untuk Menangani Data Tidak Seimbang Pada Kasus Kecelakaan Lalu Lintas Kota Pariaman

Manja Danova Putri

ABSTRAK

Kecelakaan lalu lintas merupakan salah satu permasalahan serius secara global yang terus meningkat. Cedera dan kerusakan fisik akibat kecelakaan lalu lintas dapat mempengaruhi kehidupan seseorang secara drastis, seperti, kemampuan untuk bergerak, bekerja, dan menjalankan aktivitas sehari-hari, sehingga memerlukan bantuan perawatan jangka panjang. Selain itu, tingginya jumlah korban akibat kecelakaan lalu lintas akan memberikan kerugian cukup besar terhadap perekonomian keluarga maupun perekonomian nasional. Pada penelitian ini dilakukan perbandingan klasifikasi terhadap karakteristik yang mencirikan kondisi korban kecelakaan lalu lintas Kota Pariaman menggunakan metode CHAID sebelum dan sesudah diterapkan SMOTE.

CHAID merupakan salah satu metode dalam teknik klasifikasi untuk membangun model klasifikasi berbasis pohon keputusan. CHAID membagi rangkaian data menjadi subgrup-subgrup berdasarkan hubungan antara variabel dependen (Y) dengan variabel independen (X). Sedangkan, SMOTE merupakan salah satu teknik yang diterapkan dalam rangka menangani ketidakseimbangan distribusi data dalam dataset.

Hasil klasifikasi CHAID memperoleh dua variabel yang signifikan terhadap kondisi korban yaitu usia korban dan jenis kecelakaan. Sedangkan setelah penerapan SMOTE diperoleh empat variabel yang signifikan yaitu jenis kecelakaan, usia korban, peran korban dan jenis pekerjaan. Hasil evaluasi menunjukkan bahwa pemodelan klasifikasi CHAID dengan penerapan SMOTE lebih baik dibandingkan dengan pemodelan klasifikasi CHAID tanpa SMOTE. Setelah penerapan SMOTE, terjadi peningkatan terhadap nilai *spesifisitas* yaitu 0,67 menjadi 0,85. Sedangkan nilai akurasi dan *sensitifitas* mengalami penurunan. Meskipun nilai *sensitifitas* mengalami penurunan, namun peningkatan terhadap nilai *spesifisitas* membuktikan bahwa penerapan SMOTE berhasil meningkatkan kemampuan model untuk mengklasifikasikan kelas minoritas dengan baik. Selain itu dengan tingkat akurasi sebesar 85% model tersebut secara keseluruhan menunjukkan hasil klasifikasi yang baik.

Kata Kunci : CHAID, Kecelakaan Lalu Lintas, SMOTE

Classification Using CHAID Algorithm with SMOTE Approach to Handle Imbalance Data at the Traffic Accident Cases in Pariaman

Manja Danova Putri

ABSTRACT

Traffic accidents are a serious global issue that continues to increase. Injuries and physical damages resulting from traffic accidents can have a drastic impact on a person's life, such as their ability to move, work, and carry out daily activities, requiring long-term care assistance. Additionally, the high number of accident victims can lead to significant economic losses for both families and the national economy. This research compares the classification of characteristics that describe the condition of traffic accident victims in Pariaman City using the CHAID method before and after the application of SMOTE.

CHAID is one of the methods in classification techniques used to build a decision tree-based classification model. CHAID divides a dataset into subgroups based on the relationship between the dependent variable (Y) and the independent variable (X). On the other hand, SMOTE is one of the techniques applied to address the imbalance distribution of data in a dataset.

The CHAID classification results obtained two significant variables related to the victim's condition, namely the age of the victim and the type of accident. Meanwhile, after implementing SMOTE, four significant variables were obtained, namely the type of accident, age of the victim, victim's role, and occupation. The evaluation results showed that the CHAID classification modeling with SMOTE outperformed the CHAID classification modeling without SMOTE. After implementing SMOTE, there was an improvement in the specificity value, from 0.67 to 0.85. However, the accuracy and sensitivity values decreased. Although the sensitivity value decreased, the increase in the specificity value proves that the application of SMOTE successfully improves the model's ability to classify the minority class accurately. Furthermore, with an overall accuracy rate of 85%, the model demonstrates good classification results.

Keywords : CHAID, Traffic Accidents, SMOTE

KATA PENGANTAR

Puji syukur peneliti ucapkan atas kehadiran Allah SWT karena berkat limpahan rahmat dan karuniannya kepada peneliti, sehingga peneliti dapat menyelesaikan Skripsi ini yang berjudul: **“Pemodelan Klasifikasi CHAID dengan Pendekatan SMOTE untuk Menangani Data Tidak Seimbang Pada Kasus Kecelakaan Lalu Lintas Kota Pariaman”**. Shalawat beserta salam peneliti hadiahkan untuk nabi Muhammad SAW, sebagai suri tauladan bagi seluruh umat.

Dalam proses penulisan skripsi ini, peneliti banyak memperoleh bantuan dari berbagai pihak, baik berupa kritik, saran, bimbingan ataupun bantuan lainnya. Oleh karena itu pada kesempatan ini peneliti ingin menyampaikan ucapan terimakasih kepada:

1. Ibu Dr. Yenni Kurniawati, M.Si, selaku ketua Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang sekaligus dosen pembahas.
2. Ibu Dina Fitria, M.Si selaku dosen pembimbing yang telah meluangkan waktunya untuk membimbing dan memberikan masukan-masukan yang bermanfaat bagi peneliti.
3. Ibu Zilrahmi, S.Pd, M.Si selaku dosen pembahas.
4. Bapak/Ibu dosen Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang.
5. Bapak/Ibu POLRES Kota Pariaman yang telah membantu dan mengizinkan peneliti untuk melakukan penelitian dan pengambilan data.

6. Teristimewa untuk orang tua yang telah mencurahkan segala kasih sayang, do'a, waktu serta dukungan penuh kepada peneliti.
7. Semua sahabat, teman dan rekan-rekan Statistika 2019 yang telah banyak membantu dan memberi dukungan kepada peneliti.

Semoga segala bentuk bimbingan, bantuan, dukungan, serta motivasi yang telah diberikan segala pihak mendapat balasan dari Allah SWT.

Dalam penyusunan Skripsi ini peneliti telah berusaha semaksimal mungkin untuk memberikan yang terbaik. Namun peneliti menyadari bahwa Skripsi ini masih memiliki banyak kekurangan karena keterbatasan pengetahuan dan kemampuan yang peneliti miliki. Untuk itu diharapkan kritik dan saran yang bersifat membangun. Semoga Skripsi ini dapat bermanfaat bagi peneliti dan semua pihak.

Padang, Januari 2024

Manja Danova Putri

DAFTAR ISI

ABSTRAK.....	i
ABSTRACT	ii
KATA PENGANTAR.....	iii
DAFTAR ISI	v
DAFTAR TABEL	vii
DAFTAR GAMBAR.....	ix
DAFTAR LAMPIRAN	x
BAB I PENDAHULUAN	1
A. Latar Belakang Masalah	1
B. Batasan Masalah	8
C. Rumusan Masalah	8
D. Tujuan Penelitian	9
E. Manfaat Penelitian	9
BAB II KAJIAN TEORITIS	10
A. Kecelakaan Lalu Lintas	10
B. Data Tidak Seimbang	21
C. <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	23
D. Klasifikasi	24
E. <i>Chi-Square Automatic Interation Detection (CHAID)</i>	26
BAB III METODOLOGI PENELITIAN	37
A. Jenis Penelitian.....	37

B. Jenis dan Sumber Data	37
C. Variabel Penelitian	37
D. Struktur Data.....	40
E. Tahapan Analisis Data.....	40
F. Diagram Alir	43
BAB IV HASIL DAN PEMBAHASAN.....	46
A. Deskripsi Data.....	46
B. Analisis Data.....	54
C. Pembahasan	74
BAB V PENUTUP	79
A. Kesimpulan	79
B. Saran.....	79
DAFTAR PUSTAKA.....	81
LAMPIRAN	84

DAFTAR TABEL

Tabel	Halaman
1. <i>Confusion Matrix</i>	25
2. Tabel Kontingensi Uji <i>Chi-Square</i>	29
3. Variabel Penelitian	37
4. Struktur Data Penelitian.....	40
5. Perbandingan Jumlah dan Proporsi Variabel Kondisi Korban	54
6. Tabulasi Silang Waktu Kejadian Terhadap Kondisi Korban.....	55
7. Frekuensi Harapan Waktu Kejadian Terhadap Kondisi Korban	56
8. Nilai <i>Chi-Square</i> Waktu Kejadian Terhadap Kondisi Korban	57
9. Uji <i>Chi-Square</i> Pertama.....	57
10. Uji <i>Chi-Square</i> Kedua untuk <i>Node 1</i>	59
11. Uji <i>Chi-Square</i> Kedua untuk <i>Node 2</i>	59
12. Segmentasi Karakteristik Korban Kecelakaan Lalu Lintas Sebelum SMOTE ...	62
13. Persentase Setiap Segmen Kondisi Korban Sebelum SMOTE	62
14. <i>Confusion Matrix</i> Klasifikasi CHAID	63
15. Jumlah dan Proporsi Kondisi Korban Sebelum dan Sesudah SMOTE	64
16. Tabulasi Silang Waktu Kejadian Terhadap Kondisi Korban Pada Data Setelah SMOTE	65
17. Frekuensi Harapan Waktu Kejadian Terhadap Kondisi Korban Pada Data Setelah SMOTE	66

18. Nilai <i>Chi-Square</i> Waktu Kejadian Terhadap Kondisi Korban Pada Data Setelah SMOTE	66
19. Uji <i>Chi-Square</i> Pertama Pada Data Setelah SMOTE	67
20. Segmentasi Karakteristik Korban Kecelakaan Lalu Lintas Setelah SMOTE	70
21. Persentase Setiap Segmen Kondisi Korban Setelah SMOTE	71
22. <i>Confusion Matrix</i> Klasifikasi CHAID Setelah SMOTE	74
23. Perbandingan Evaluasi Klasifikasi CHAID Sebelum dan Sesudah SMOTE.....	75

DAFTAR GAMBAR

Gambar	Halaman
1. Rasio Jumlah Kendaraan dan Jumlah Kecelakaan Lalu Lintas di Provinsi Sumatera Barat Berdasarkan Kabupaten/Kota Tahun 2022.....	3
2. Persentase Kondisi Korban Kecelakaan Lalu Lintas Kota Pariaman Tahun 2022 ..5	
3. Struktur Diagram Pohon CHAID.....	35
4. Diagram Alir Penelitian.....	43
5. Diagram Alir CHAID.....	44
6. Diagram Alir SMOTE	45
7. Karakteristik Kondisi Korban Kecelakaan Lalu Lintas	46
8. Karakteristik Waktu Kejadian Berdasarkan Kondisi Korban.....	47
9. Karakteristik Kendaraan Korban Berdasarkan Kondisi Korban.....	48
10. Karakteristik Jenis Kecelakaan Berdasarkan Kondisi Korban.....	49
11. Karakteristik Usia Korban Berdasarkan Kondisi Korban	50
12. Karakteristik Jenis Kelamin Berdasarkan Kondisi Korban.....	51
13. Karakteristik Jenis Pekerjaan Berdasarkan Kondisi Korban.....	51
14. Karakteristik Kepemilikan SIM Berdasarkan Kondisi Korban.....	52
15. Karakteristik Peran Korban Berdasarkan Kondisi Korban	53
16. Diagram Pohon CHAID Awal.....	58
17. Diagram Pohon CHAID	60
18. Diagram Pohon CHAID Awal Setelah SMOTE	68
19. Diagram Pohon CHAID Setelah SMOTE.....	68

DAFTAR LAMPIRAN

Lampiran	Halaman
1. Data Korban Kecelakaan Lalu Lintas Kota Pariaman Tahun 2022	85
2. <i>Syntax</i> dan Output <i>R</i> Pembagian Data <i>Training</i> dan Data <i>Testing</i>	86
3. <i>Syntax</i> dan Output <i>Python</i> Pemodelan Klasifikasi CHAID.....	87
4. <i>Syntax</i> dan Output <i>Python</i> Penerapan SMOTE Pada Data <i>Training</i>	89
5. <i>Syntax</i> dan Output <i>Python</i> Penerapan SMOTE Pada Data <i>Testing</i>	91
6. <i>Syntax</i> dan Output <i>Python</i> Pemodelan Klasifikasi SMOTE CHAID.....	93
7. Output SPSS, Tabulasi Silang Masing-Masing Variabel Independen dengan Variabel Dependen (Data Sebelum SMOTE).....	95
8. Perhitungan Nilai Frekuensi Harapan Variabel Waktu Kejadian Terhadap Kondisi Korban (Data Sebelum SMOTE)	97
9. Perhitungan Nilai <i>Chi-Square</i> Variabel Waktu Kejadian Terhadap Kondisi Korban (Data Sebelum SMOTE)	99
Output SPSS, Nilai <i>Chi-Square</i> Masing-Masing Variabel Dependen dengan Variabel Independen (Data Sebelum SMOTE)	101
10. Output SPSS, Tabulasi Silang Masing-Masing Variabel Dependen dengan Variabel Independen (Data Setelah SMOTE)	103
11. Perhitungan Nilai Frekuensi Harapan Variabel Waktu Kejadian Terhadap Kondisi Korban (Data Setelah SMOTE).....	105
12. Perhitungan Nilai <i>Chi-Square</i> Variabel Waktu Kejadian Terhadap Kondisi Korban (Data Setelah SMOTE)	107

13. Output SPSS, Nilai *Chi-Square* Masing-Masing Variabel Dependen dengan Variabel Independen (Data Setelah SMOTE) 109

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

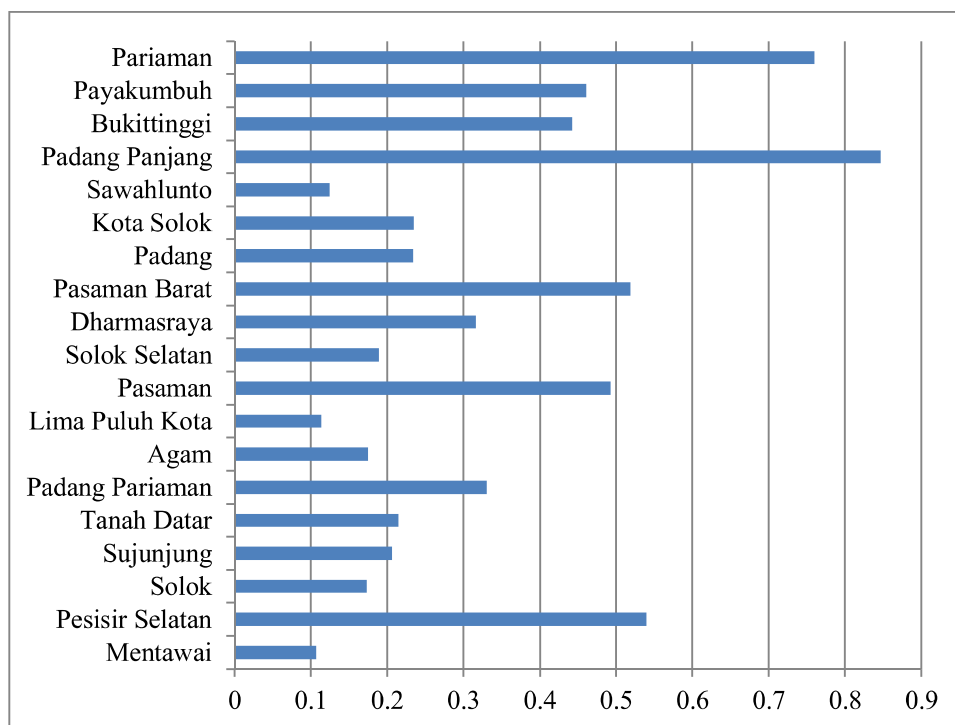
Perkembangan sarana transportasi merupakan aspek penting dalam menunjang keberhasilan pembangunan terutama dalam mendukung kegiatan perekonomian masyarakat dan meningkatkan pengembangan wilayah baik daerah pedesaan maupun perkotaan. Di Indonesia kendaraan bermotor menjadi sarana transportasi paling populer yang digunakan masyarakat karena berperan langsung dalam memudahkan aktivitas manusia untuk berpindah dari suatu tempat ke tempat lain. Keberagaman aktivitas dan kebutuhan manusia yang didukung oleh kemajuan ilmu pengetahuan dan teknologi menyebabkan kendaraan bermotor mengalami perkembangan yang semakin pesat baik dari segi kualitas maupun kuantitasnya. Berdasarkan data Badan Pusat Statistik (BPS) Indonesia, pertumbuhan jumlah kendaraan bermotor di Indonesia periode tahun 2017-2021 mengalami peningkatan sebesar 4,53 persen pertahun (Badan Pusat Statistik, 2021). Peningkatan tersebut akan berpengaruh terhadap volume dan kepadatan jalan raya sehingga mengakibatkan rentan terjadinya kecelakaan lalu lintas.

Kecelakaan lalu lintas didefinisikan sebagai suatu peristiwa yang tidak diduga dan tidak disengaja yang melibatkan kendaraan dengan atau tanpa pengguna jalan lain, sehingga mengakibatkan korban manusia atau kerugian harta benda (Undang-Undang No.22 Tentang Lalu Lintas Dan Angkutan Jalan, 2009). Kecelakaan lalu lintas dapat mengakibatkan terjadinya hal yang sangat fatal terhadap korban, mulai dari luka ringan, luka berat, bahkan mengakibatkan kematian. 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 terus meningkat dan mencapai angka tertinggi sebesar 1,35 juta jiwa pada tahun 2016. Angka ini 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 (Lampiran Peraturan Presiden Republik Indonesia Nomor 1 Tahun 2022). Pada periode tahun 2010-2020 jumlah korban kecelakaan lalu lintas di Indonesia masih tergolong tinggi. Hal ini disebabkan oleh tingginya paparan risiko, laju pertumbuhan jumlah penduduk, jumlah kendaraan, dan jumlah perjalanan yang menggunakan kendaraan bermotor. Berdasarkan data Kepolisian Negara Republik Indonesia, jumlah korban kecelakaan lalu lintas periode tahun 2010-2020 berkisar antara 147.798-197.560 jiwa, dan korban meninggal dunia berkisar antara 23.529-32.657 jiwa. Korban meninggal dunia tersebut didominasi oleh kalangan usia 15-34 tahun.

Di Indonesia, Sumatera Barat merupakan salah satu provinsi yang memiliki angka kecelakaan lalu lintas cukup tinggi. Pada tahun 2022 tercatat sebanyak 2.956 kejadian kecelakaan lalu lintas yang terjadi di Sumatera Barat. Kejadian tersebut mengakibatkan 6.000 orang menjadi korban dengan korban luka ringan sebanyak 5.156 orang, luka berat sebanyak 323 orang, dan 521 orang lainnya meninggal dunia. Jika dibandingkan dengan jumlah kendaraan, maka rasio antara jumlah kendaraan dan jumlah kecelakaan lalu lintas pada masing-masing Kabupaten/Kota di Provinsi Sumatera Barat dapat dilihat pada Gambar 1.



Gambar 1. Rasio Jumlah Kendaraan dan Jumlah Kecelakaan Lalu Lintas berdasarkan Kabupaten/Kota di Provinsi Sumatera Barat Tahun 2022.

Berdasarkan Gambar 1 dapat dilihat bahwa Kota Padang Panjang memiliki rasio tertinggi antara jumlah kendaraan dan jumlah kecelakaan lalu lintas di Provinsi Sumatera Barat pada tahun 2022. Sementara itu diurutkan kedua adalah Kota Pariaman. Akan tetapi jika dilihat dari jumlah korban kecelakaan lalu lintasnya, Kota Pariaman memiliki jumlah korban kecelakaan lalu lintas yang lebih tinggi dibandingkan Kota Padang Panjang. Dimana pada tahun tersebut terdapat sebanyak 338 orang yang menjadi korban pada kasus kecelakaan lalu lintas di Kota Pariaman, sedangkan di Kota Padang Panjang jumlah Korban kecelakaan lalu lintas hanya sebanyak 272 orang.

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 et al., 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.

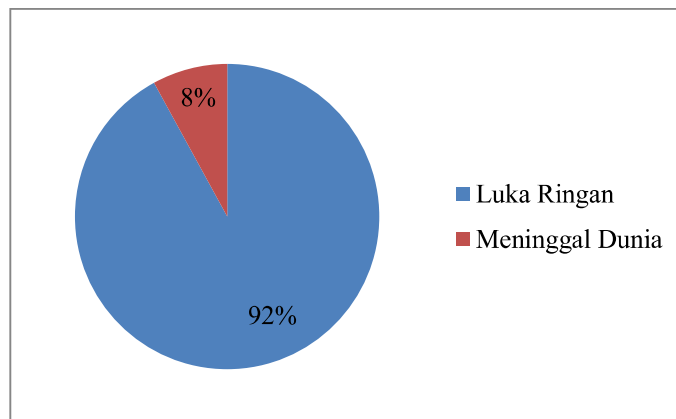
Jika dilihat dari segi ekonomi tingginya jumlah korban akibat kecelakaan lalu lintas akan memberikan kerugian yang cukup besar baik terhadap perekonomian keluarga maupun terhadap sistem perekonomian nasional (Suseno, 2022). Banyaknya jumlah anggota keluarga yang meninggal dunia maupun mengalami cacat permanen pada usia produktif, akan berdampak terhadap 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.

Dari data kecelakaan lalu lintas yang tercatat dalam Laporan Polisi di POLRES Kota Pariaman, karakteristik kecelakaan lalu lintas dapat dilihat berdasarkan waktu kejadian, kendaraan korban, jenis kecelakaan, usia korban, jenis kelamin, jenis pekerjaan, kepemilikan Surat Izin Mengemudi (SIM), dan peran korban saat terjadinya kecelakaan. Untuk mengetahui karakteristik yang menggambarkan permasalahan ini dapat digunakan pendekatan dengan teknik klasifikasi.

Chi-Square Automatic Interaction Detection (CHAID) merupakan salah satu metode dalam teknik klasifikasi. CHAID merupakan metode untuk membentuk segmentasi yang membagi sebuah sampel menjadi dua atau lebih kelompok berdasarkan kriteria tertentu, kemudian diteruskan dengan membagi kelompok tersebut menjadi kelompok yang lebih kecil berdasarkan variabel independen yang lain (Hasna

& Kunto, 2006). Metode ini akan membagi data menjadi kelompok-kelompok yang lebih kecil berdasarkan keterkaitan antara variabel dependen dengan variabel independen.

CHAID memiliki keunggulan dalam mengeksplorasi data berjumlah besar, yang seluruh peubahnya bertipe kategori baik dengan skala nominal maupun ordinal. Analisis ini tidak membutuhkan asumsi apapun karena menggunakan statistik *chi-square* dalam uji hipotesisnya (Miftahuddin, 2012). Output yang dihasilkan berupa pohon keputusan, sehingga dapat diketahui variabel independen yang paling signifikan diantara variabel independen lainnya. Persentase korban kecelakaan lalu lintas di Kota Pariaman tahun 2022 berdasarkan kondisi korban dapat dilihat pada Gambar 2.



Gambar 2. Persentase Kondisi Korban Kecelakaan Lalu Lintas Kota Pariaman Tahun 2022.

Dari Gambar 2 dapat dilihat bahwa terdapat ketidakseimbangan antara jumlah korban kecelakaan lalu lintas yang mengalami luka ringan dengan korban yang meninggal dunia. Ketidakseimbangan ini akan memberikan dampak yang tidak baik pada hasil klasifikasi (Siringoringo, 2019). Model akan cenderung memprediksi sampel sebagai kelas mayoritas dan mengabaikan kelas minoritas, sehingga

mengakibatkan tingginya nilai akurasi dari hasil klasifikasi (Fitriani dkk, 2021). Akurasi yang tinggi menunjukkan bahwa model berhasil memprediksi dengan benar sebagian sampel dalam dataset. Namun tidak dapat memberikan gambaran terhadap kemampuan model dalam mengklasifikasikan kelas minoritas. Sementara itu, *sensitivitas* akan bernilai rendah karena model memiliki kecenderungan untuk mengabaikan kelas minoritas, sedangkan *spesifisitas* akan bernilai tinggi karena model cenderung memprediksi sebagian besar sampel sebagai kelas mayoritas.

Penanganan pada kelas data yang tidak seimbang dapat diatasi dengan cara melakukan *oversampling* pada kelas minoritas atau *undersampling* pada kelas mayoritas. Namun *oversampling* pada kelas minoritas akan menyebabkan model menjadi *overfitting* karena dilakukan dengan cara menduplikasi data yang telah ada sebelumnya sehingga pengklasifikasian akan terkena informasi yang sama. Sedangkan metode *undersampling* pada kelas mayoritas dapat mengakibatkan hilangnya informasi penting pada data karena bekerja dengan cara mengurangi data pada kelas mayoritas (Ni Putu Yulika Trisna dkk, 2021). Oleh karena itu dibutuhkan suatu metode yang dapat digunakan untuk mengatasi permasalahan data yang tidak seimbang ini.

Synthetic Minority Oversampling Technique (SMOTE) merupakan salah satu algoritma turunan teknik *oversampling* untuk mengatasi ketidakseimbangan data dalam pembelajaran mesin (*Machine Learning*). SMOTE bekerja dengan cara membuat replikasi dari data minoritas sehingga jumlah data antar kelas menjadi seimbang. Replikasi ini disebut sebagai data sintesis (Sitohang dkk., 2022). Data sintesis dibuat dengan melakukan interpolasi antara sampel-sampel minoritas yang ada. Data sintesis tersebut kemudian ditambahkan pada data asli untuk membentuk dataset

yang seimbang secara keseluruhan. Dataset yang telah seimbang dapat digunakan untuk melatih model agar dapat mempelajari dan memahami karakteristik kelas minoritas dengan lebih baik. Dengan demikian model cenderung menjadi lebih baik dalam mengklasifikasikan dan memprediksi sampel-sampel dari kelas minoritas.

Penelitian sebelumnya mengenai pemodelan klasifikasi CHAID dengan pendekatan SMOTE pernah dilakukan oleh Santi dkk, (2022) dan memperoleh hasil bahwa, penerapan SMOTE mampu meningkatkan performa pohon klasifikasi CHAID dalam mengklasifikasikan pasien berdasarkan kategorinya, hasil tersebut ditunjukkan dengan peningkatan nilai *sensitivitas* menjadi 26,667, sehingga dapat disimpulkan bahwa model klasifikasi SMOTE CHAID memiliki performa lebih baik dibandingkan dengan pemodelan CHAID tanpa SMOTE. Penelitian yang dilakukan oleh Hidayati dan Arcana, (2020) menunjukkan hasil bahwa pendekatan SMOTE lebih baik digunakan dengan nilai *sensitivitas* sebesar 48,3% dan nilai presisi sebesar 75,9%. Selain itu penelitian yang dilakukan oleh Syamsi dan Nugraha, (2019) memiliki tingkat akurasi sebesar 86,3% untuk penerapan metode CHAID dengan pendekatan SMOTE.

Dengan menggunakan model klasifikasi *Chi-Square Automatic Interaction Detection* (CHAID) dan pendekatan *Synthetic Minority Oversampling Technique* (SMOTE) sebagai penyeimbang data pada kelas minoritas maka penelitian ini diberi judul **“Pemodelan Klasifikasi CHAID dengan Pendekatan SMOTE Untuk Menangani Data Tidak Seimbang Pada Kasus Kecelakaan Lalu Lintas Kota Pariaman”**.

B. Batasan Masalah

Agar tidak terjadi penyimpangan dalam permasalahan yang diteliti, maka dibutuhkan batasan-batasan dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan adalah data korban kecelakaan lalu lintas pada kendaraan bermotor yang diperoleh dari POLRES Kota Pariaman tahun 2022.
2. Pada penelitian ini akan diteliti karakteristik kecelakaan lalu lintas yang mencirikan kondisi korban. Karakteristik tersebut meliputi waktu kejadian, kendaraan korban, jenis kecelakaan, usia korban, jenis kelamin, jenis pekerjaan, kepemilikan SIM, dan peran korban.

C. Rumusan Masalah

Berdasarkan latar belakang permasalahan diatas, maka dapat dirumuskan masalah yang akan diteliti adalah sebagai berikut:

1. Bagaimana performansi ketepatan kinerja dari hasil klasifikasi kondisi korban kecelakaan lalu lintas di Kota Pariaman tahun 2022 menggunakan metode CHAID sebelum dan sesudah diterapkan SMOTE?
2. Apa saja karakteristik yang memiliki hubungan yang signifikan dengan kondisi korban kecelakaan lalu lintas di Kota Pariaman tahun 2022?

D. Tujuan Penelitian

Berdasarkan rumusan masalah diatas, maka diperoleh tujuan penelitian ini adalah sebagai berikut:

1. Untuk mengetahui performansi ketepatan kinerja dari hasil klasifikasi kondisi korban kecelakaan lalu lintas di Kota Pariaman tahun 2022 menggunakan metode CHAID sebelum dan sesudah diterapkan SMOTE.

2. Untuk mengetahui apa saja karakteristik yang memiliki hubungan yang signifikan dengan kondisi korban kecelakaan lalu lintas di Kota Pariaman tahun 2022.

E. Manfaat Penelitian

Adapun manfaat yang dapat diambil dari penelitian ini adalah sebagai berikut:

1. Dapat menambah wawasan dan ilmu pengetahuan bagi penulis tentang pemahaman terhadap analisis CHAID dan SMOTE.
2. Hasil penelitian diharapkan dapat menambah informasi serta dapat dijadikan sebagai bahan evaluasi oleh POLRES Kota Pariaman dalam pengendalian peraturan lalu lintas dan kebijakan lainnya yang berkaitan dengan kecelakaan lalu lintas.
3. Dapat dijadikan sebagai referensi untuk penelitian selanjutnya yang menggunakan metode CHAID, SMOTE dan permasalahan terkait kasus kecelakaan lalu lintas.