

Klasifikasi Keparahan Kecelakaan Lalu Lintas Berbasis Decision Tree C4.5 dan Teknik SMOTE

Githa Alfiansyah, Hendra, Raisul Azhar

Universitas Bumigora, Mataram, Indonesia

Correspondence : raisulazhar@universitasbumigora.ac.id

Abstrak

Kecelakaan lalu lintas merupakan masalah serius yang berdampak besar terhadap keselamatan masyarakat, khususnya di Provinsi Nusa Tenggara Barat (NTB). Dalam beberapa tahun terakhir, angka kecelakaan di NTB menunjukkan tren fluktuatif dengan kecenderungan peningkatan pada kategori kecelakaan berat. Kondisi ini menjadi isu penting yang harus segera mendapat perhatian dari pihak berwenang maupun masyarakat. Tingginya angka kecelakaan, terutama yang berakibat fatal, menegaskan perlunya upaya penanggulangan berbasis data yang terarah dan berkelanjutan. Dalam upaya menurunkan tingkat kecelakaan dan meningkatkan keselamatan lalu lintas, penelitian ini bertujuan untuk mengklasifikasikan tingkat keparahan kecelakaan lalu lintas menggunakan algoritma Decision Tree (C4.5) yang dikombinasikan dengan teknik SMOTE untuk mengatasi ketidakseimbangan data antar kelas. Data yang digunakan merupakan data sekunder dari Badan Pusat Statistik NTB, mencakup periode 2018 hingga 2023. Model divalidasi menggunakan 15-Fold Cross Validation dan menghasilkan akurasi sebesar 98,33%. Hasil klasifikasi menunjukkan bahwa atribut "Jumlah" merupakan variabel paling berpengaruh dalam membentuk aturan klasifikasi, dengan pembagian kategori keparahan ke dalam tiga kelas: Rendah, Sedang, dan Tinggi. Temuan ini menunjukkan bahwa model memiliki performa tinggi dan dapat dimanfaatkan sebagai alat bantu dalam menyusun kebijakan pencegahan kecelakaan, mengidentifikasi wilayah rawan, serta mendukung pengambilan keputusan yang lebih tepat dalam upaya menurunkan risiko kecelakaan di wilayah NTB.

Kata kunci: klasifikasi, kecelakaan lalu lintas, Decision Tree, SMOTE, Cross Validation, NTB.

Abstract

Traffic accidents are a serious problem that has a major impact on public safety, especially in the Province of West Nusa Tenggara (NTB). In recent years, the number of accidents in NTB has shown a fluctuating trend with an increasing tendency in the category of serious accidents. This condition is an important issue that must immediately receive attention from the authorities and the public. The high number of accidents, especially those that result in fatalities, emphasizes the need for targeted and sustainable data-based mitigation efforts. In an effort to reduce the accident rate and improve traffic safety, this study aims to classify the severity of traffic accidents using the Decision Tree algorithm (C4.5) combined with the SMOTE technique to overcome data imbalances between classes. The data used is secondary data from the NTB Central Statistics Agency, covering the period 2018 to 2023. The model was validated using 15-Fold Cross Validation and produced an accuracy of 98.33%. The classification results show that the "Number" attribute is the most influential variable in forming classification rules, with the division of severity categories into three classes: Low, Medium, and High. These findings indicate that the model has high performance and can be used as a tool in formulating accident prevention policies, identifying vulnerable areas, and supporting more appropriate decision-making in an effort to reduce the risk of accidents in the NTB region.

Keywords: classification, traffic accidents, Decision Tree, SMOTE, Cross Validation, NTB.

1. Pendahuluan

Saat ini, kepemilikan kendaraan pribadi kini telah menjadi bagian dari kebutuhan utama masyarakat, baik roda dua maupun empat [1]. Hal ini berdampak pada meningkatnya jumlah kecelakaan lalu lintas yang kini menjadi persoalan global. Kecelakaan dapat terjadi kapan saja dan di mana saja, dengan dampak mulai dari luka ringan hingga kematian dan kerugian ekonomi [2]. Berdasarkan laporan Polda NTB, tercatat 1.664 kasus kecelakaan terjadi di Lombok dan Sumbawa sepanjang Januari–September 2024. Selain itu, korban meninggal akibat kecelakaan di NTB mencapai 304 orang, meningkat 15,36% dari tahun sebelumnya. Fakta ini menunjukkan bahwa kecelakaan terus meningkat. Diperlukan pendekatan sistematis berbasis data untuk mengantisipasi hal tersebut.

Di tengah meningkatnya angka kecelakaan, belum banyak kajian yang mengklasifikasikan tingkat keparahannya secara sistematis. Padahal, klasifikasi ini penting agar penanganan dapat disesuaikan dengan tingkat risiko kejadian. Beberapa variabel seperti usia korban, jenis kendaraan, dan waktu kejadian terbukti berpengaruh terhadap tingkat keparahan kecelakaan [3]. Informasi ini dapat membantu dalam pengelompokan korban ke dalam kategori ringan, sedang, dan berat. Dengan klasifikasi seperti ini, pola kecelakaan dapat dipahami secara lebih menyeluruh. Hal tersebut akan mendukung penyusunan kebijakan yang lebih tepat sasaran. Oleh karena itu, pendekatan klasifikasi berbasis data sangat diperlukan.

Analisis klasifikasi digunakan untuk mengolah data kecelakaan lalu lintas secara sistematis berdasarkan variabel seperti jenis kejadian, lokasi, hingga waktu kecelakaan [4]. Hasil klasifikasi mampu mengungkap pola tersembunyi dalam data yang berkaitan dengan tingkat keparahan korban [5]. Informasi ini dapat dimanfaatkan untuk menilai karakteristik kecelakaan dan mengidentifikasi faktor penyebab secara objektif. Dari kejadian tersebut maka perlu adanya analisa daerah rawan kecelakaan, Dari analisa ini diharapkan langkah penanggulangan kecelakaan lalu lintas akan lebih mudah sehingga risiko kecelakaan di masa mendatang bisa diminimalkan [6]. Penelitian ini menerapkan pendekatan berbasis klasifikasi yang merupakan bagian dari teknik data mining untuk mengolah dan menganalisis data secara sistematis. Model analisis dalam studi ini dibangun menggunakan algoritma *Decision Tree*, yang dimanfaatkan sebagai metode utama dalam proses klasifikasi data [7]. *Decision Tree* dikenal luas sebagai salah satu metode paling umum dalam dunia data mining, khususnya dalam mendukung proses pengambilan keputusan. Metode ini tidak memerlukan pengelolaan pengetahuan awal dan efektif dalam menangani data berdimensi besar secara efisien [8].

Algoritma *Decision Tree* dapat digunakan untuk mengelompokkan tingkat keparahan kecelakaan lalu lintas di wilayah NTB, dengan mempertimbangkan variabel-variabel penting seperti tipe jalan, kondisi cuaca, serta waktu terjadinya peristiwa [9]. Algoritma ini dapat digunakan untuk membentuk aturan klasifikasi yang kemudian dikembangkan menjadi aplikasi klasifikasi [10]. Hasil klasifikasi ini memberikan gambaran pola kecelakaan yang membantu pihak berwenang merumuskan kebijakan keselamatan lalu lintas dan memudahkan pengambilan keputusan untuk mengurangi tingkat kecelakaan [11]. Beberapa penelitian sebelumnya telah mengintegrasikan metode *Decision Tree* dan SMOTE untuk mengatasi ketidakseimbangan kelas pada data kecelakaan, namun hanya berfokus pada klasifikasi dua kelas tanpa melihat kompleksitas tingkat keparahan korban secara lebih detail [12]. Di sisi lain, pendekatan lain menggunakan Regresi Logistik Ordinal dan Fuzzy K-NN dalam klasifikasi tingkat keparahan, tetapi belum menerapkan teknik penyeimbangan data maupun perbandingan desain model klasifikasi untuk meningkatkan akurasi [3]. Penelitian ini mengisi celah tersebut dengan menggabungkan pendekatan klasifikasi berbasis *Decision Tree* dan teknik SMOTE secara komprehensif untuk memprediksi tingkat keparahan korban kecelakaan berdasarkan data yang tidak seimbang.[13]

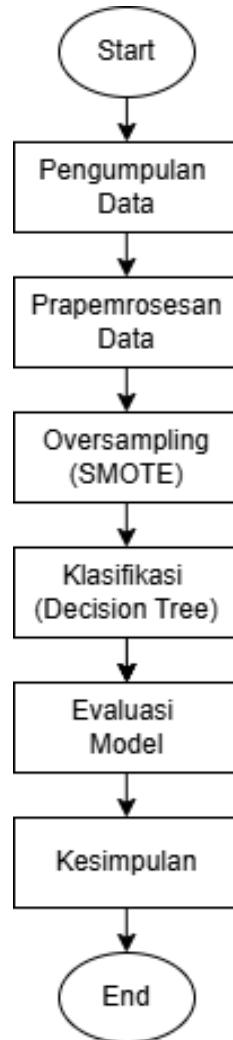
Penelitian ini membatasi ruang lingkupnya agar analisis lebih fokus, yaitu pada klasifikasi tingkat kecelakaan di NTB berdasarkan data 2018–2023. Kecelakaan akan dikategorikan menurut tingkat keparahan (ringan, sedang, berat) dengan mempertimbangkan variabel seperti waktu, lokasi, dan jumlah korban. Data berasal dari instansi resmi untuk menjamin validitas [14]. Tujuan dari penelitian ini adalah untuk mengklasifikasikan tingkat keparahan kecelakaan lalu lintas di Provinsi Nusa Tenggara Barat menggunakan algoritma *Decision Tree* (C4.5) dengan bantuan teknik SMOTE untuk mengatasi ketidakseimbangan data antar kelas. Dengan demikian, penelitian ini tidak hanya berkontribusi dalam aspek akademis, tetapi juga memberikan manfaat praktis bagi pihak kepolisian dan pemerintah daerah dalam meningkatkan keselamatan lalu lintas [15].

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan fokus pada klasifikasi tiga tingkat keparahan, yaitu kategori rendah, sedang, dan tinggi. Tujuan utama penelitian adalah mengidentifikasi pola tingkat keparahan kecelakaan lalu lintas di Provinsi Nusa Tenggara Barat (NTB) berdasarkan data dari Badan Pusat Statistik (BPS) dalam kurun waktu 5 tahun trakhir, dimulai dari tahun 2018 - 2023. Data yang digunakan merupakan data sekunder yang diperoleh dari publikasi resmi BPS dan instansi terkait. Data ini mencakup variabel-variabel utama seperti kode provinsi, satuan lalu lintas (Polres), jumlah laporan kecelakaan, dan tahun terjadinya kecelakaan. Keunggulan penggunaan data sekunder ini terletak pada kelengkapan, cakupan waktu yang panjang, serta reliabilitasnya karena bersumber dari instansi resmi pemerintah.

2.1. Deskripsi dan Langkah-Langkah Penelitian

Penelitian ini disusun secara tersusun menggunakan data kecelakaan 5 tahun terakhir untuk mengevaluasi dan mengakaji tingkat kecelakaan lalu lintas di wilayah NTB. Adapun tahapan adalah sebagai berikut :



Gambar 1. *Flowchart* Alur Penelitian.

1. Pengumpulan Data

Penelitian ini menggunakan data sekunder yang diperoleh dari publikasi resmi Badan Pusat Statistik (BPS) Provinsi Nusa Tenggara Barat [3]. Total data yang digunakan berjumlah 60 entri, yang masing-masing memuat tiga atribut utama, yaitu satuan wilayah kepolisian (Polres), tahun kejadian, serta jumlah kasus kecelakaan. Ketiga atribut tersebut dijadikan sebagai dasar dalam proses klasifikasi tingkat keparahan kecelakaan lalu lintas.

2. Prapemrosesan Data

Data diperiksa agar tidak ada duplikasi, formatnya seragam, dan informasinya konsisten. Langkah ini penting agar data bersih dan siap untuk dianalisis. Seperti dijelaskan, "data penelitian harus melewati tahap proses cleaning, transformasi dan feature selection" [2]. Tujuannya adalah agar data yang digunakan benar-benar bersih dan tidak mengandung kesalahan yang bisa memengaruhi hasil akhir.

Tabel 2. Data setelah melewati prapemrosesan

Satuan Polisi	Tahun	Jumlah
Polres Kabupaten Lombok Barat	2018	60
Polres Kota Bima	2018	115
Polres Kabupaten Bima	2018	113
Polres Kabupaten Dompu	2018	66
Polres Kabupaten Sumbawa Besar	2018	195

Tabel 1 menunjukkan sebagian data kecelakaan lalu lintas yang telah melewati proses prapemrosesan. Atribut Satuan Polisi, Tahun, dan Jumlah dipertahankan sebagai fitur utama. Data ini telah siap digunakan dalam tahap klasifikasi dengan metode *Decision Tree*.

3. Oversampling (SMOTE)

Karena distribusi kategori kecelakaan (rendah, sedang, tinggi) tidak seimbang, dilakukan proses oversampling menggunakan teknik SMOTE (*Synthetic Minority Oversampling Technique*). Metode ini bertujuan menyeimbangkan jumlah data antar kelas dengan mensintesis data minoritas sehingga model tidak bias terhadap kelas mayoritas.[12]

4. Klasifikasi (*Decision Tree*)

Proses klasifikasi dalam penelitian ini dilakukan dengan memanfaatkan algoritma Decision Tree jenis C4.5. Setelah seluruh data melalui tahapan prapemrosesan dan penyeimbangan menggunakan teknik SMOTE, data kemudian dimasukkan ke dalam proses pelatihan model. Atribut-atribut yang digunakan sebagai input terdiri atas "Jumlah", "Tahun", dan "Satuan Polisi", sedangkan variabel target yang akan diprediksi adalah "Kategori" yang dibagi ke dalam tiga kelas: Rendah, Sedang, dan Tinggi.

Algoritma C4.5 bekerja dengan cara mengevaluasi setiap atribut untuk menentukan atribut terbaik yang dapat memisahkan data secara optimal, menggunakan perhitungan gain ratio. Proses klasifikasi dilakukan secara bertahap, membentuk struktur pohon keputusan dari akar menuju simpul-simpul terminal, dengan cabang yang dibentuk berdasarkan nilai ambang atribut yang dipilih.

Tahapan pelatihan dan pengujian model dilakukan di lingkungan RapidMiner Studio dengan membandingkan antara *5-Fold*, *10-Fold*, *15-Fold Cross Validation*. Secara matematis, perhitungan gain ratio pada algoritma C4.5 menggunakan konsep entropy untuk mengukur ketidakpastian dalam data. Rumus-rumus dasar yang digunakan dalam pemilihan atribut dapat dijelaskan sebagai berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \cdot Entropy(S_i) \quad (1)$$

Rumus diatas digunakan untuk pemilihan atribut terbaik untuk pemisahan data dilakukan dengan menghitung nilai *Gain* atau *Information Gain*. Rumus pertama menggambarkan proses perhitungan Gain suatu atribut *A* terhadap himpunan kasus *S*. *Gain* dihitung dari selisih antara nilai *Entropy* awal dari himpunan *S*, yaitu tingkat ketidakpastian sebelum data dipisahkan, dengan jumlah tertimbang *entropy* dari setiap partisi *S_i* setelah pemisahan berdasarkan atribut *A*. Bobot masing-masing partisi dihitung berdasarkan proporsi jumlah kasus dalam partisi tersebut terhadap total jumlah kasus dalam *S*, yaitu $\frac{|S_i|}{|S|}$. Semakin tinggi nilai *Gain*, berarti atribut tersebut semakin baik dalam mengurangi ketidakpastian data, sehingga lebih layak dipilih sebagai simpul pemisah dalam pohon keputusan.

Sementara itu, perhitungan nilai entropi dapat dilihat pada persamaan berikut:

$$Entropy(S) = - \sum_{i=1}^n p_i \cdot \log_2 p_i \quad (2)$$

Sementara itu, rumus kedua digunakan untuk menghitung *Entropy* dari himpunan data *S*, yaitu ukuran ketidakteraturan atau ketidakpastian dari distribusi kelas dalam data tersebut. Entropy dihitung dengan menjumlahkan hasil kali antara proporsi masing-masing kelas *p_i* terhadap total data dengan logaritma basis dua dari proporsi tersebut, lalu dikalikan dengan -1. Entropy bernilai 0 jika seluruh data termasuk dalam satu kelas, dan bernilai maksimum jika distribusi kelas merata. Dalam metode C4.5, nilai *entropy* ini menjadi dasar dalam perhitungan *Gain* dan juga *Gain Ratio*, sehingga perannya sangat penting dalam membentuk struktur pohon keputusan yang optimal dan akurat.

5. Evaluasi Model

Proses evaluasi digunakan sebagai tahap akhir untuk menilai sejauh mana model yang dikembangkan mampu memberikan hasil yang akurat dan konsisten. yang telah dibentuk dengan algoritma *Decision Tree* dan teknik *oversampling* SMOTE. Proses ini menggunakan *RapidMiner Studio* sebagai alat bantu, yang memungkinkan evaluasi menyeluruh terhadap performa model. *RapidMiner* digunakan untuk menampilkan visualisasi hasil klasifikasi sekaligus menghitung metrik

evaluasi seperti akurasi, presisi, serta mengidentifikasi jumlah prediksi yang benar dan salah. Akurasi diartikan sebagai proporsi prediksi yang sesuai dari keseluruhan data yang diuji.[12]

6. Kesimpulan

Berdasarkan hasil evaluasi, ditarik kesimpulan mengenai pola distribusi kecelakaan di wilayah NTB. Penelitian ini juga menyusun rekomendasi strategis kepada pihak terkait untuk meningkatkan keselamatan lalu lintas dan menekan angka kecelakaan.

3. Hasil dan Pembahasan

3.1. Deskripsi Data

Dataset yang digunakan berisi informasi tentang kejadian kecelakaan lalu lintas yang terdiri dari beberapa atribut input (fitur) dan satu atribut target. Setiap atribut memiliki peran penting dalam membentuk model klasifikasi untuk menentukan tingkat keparahan kecelakaan. Berikut penjelasannya:

Tabel 2. Atribut

No	Atribut	Tipe Data	Deskripsi
1	Satuan Polisi	Nominal	Identitas Lokasi atau Wilayah
2	Tahun	Numerik	Waktu Kejadian
3	Jumlah	Numerik	Jumlah Korban
4	Kategori	Nominal	Label klasifikasi yang menunjukkan tingkat keparahan kecelakaan berdasarkan data korban, seperti "Rendah", "Sedang", atau "Tinggi".

Tabel 2 menunjukkan bahwa sebagian besar atribut merupakan data numerik yang merepresentasikan jumlah korban dan kondisi kecelakaan, sedangkan atribut Kategori digunakan sebagai label klasifikasi untuk menentukan tingkat keparahan kecelakaan. Atribut-atribut ini menjadi dasar dalam membangun model klasifikasi menggunakan metode *Decision Tree*.

3.2. Prapemrosesan Data

Data harus melalui proses prapemrosesan untuk memastikan kelengkapan, konsistensi, dan kualitas data. Proses ini meliputi pemilihan atribut, penentuan label, penanganan ketidakseimbangan data, dan penyesuaian format atribut agar sesuai dengan kebutuhan algoritma.

Tabel 3. Data bersih setelah prapemrosesan

Satuan Polisi	Tahun	Jumlah	Kategori
Polres Kabupaten Dompu	2021	54	Rendah
Polres Kabupaten Bima	2018	113	Sedang
Polres Kabupaten Lombok Timur	2018	388	Tinggi

Tabel 3 Kategori target ditambahkan setelah proses klasifikasi dengan algoritma C4.5, berdasarkan nilai ambang yang ditentukan oleh model.

3.3. Penerapan SMOTE

Distribusi awal data keparahan kecelakaan lalu lintas menunjukkan ketidakseimbangan antar kelas. Kategori "Sedang" memiliki proporsi yang dominan, sementara kategori "Rendah" dan "Tinggi" hanya mewakili sebagian kecil dari keseluruhan data. Kondisi ini berpotensi menyebabkan model klasifikasi bias terhadap kelas mayoritas. Untuk mengatasi hal tersebut, digunakan metode oversampling SMOTE (Synthetic Minority Oversampling Technique). SMOTE bekerja dengan membentuk sampel baru pada kelas minoritas berdasarkan tetangga terdekat, tanpa melakukan duplikasi langsung. Berikut adalah distribusi data sebelum dan sesudah dilakukan proses SMOTE:

Tabel 4. Distribusi kelas sebelum dan sesudah SMOTE

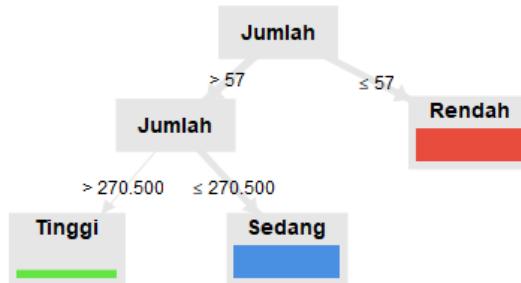
Kategori	Jumlah (Sebelum SMOTE)	Jumlah (Setelah SMOTE)
Rendah	3	47
Sedang	47	47
Tinggi	10	47

Dengan distribusi yang telah seimbang, model klasifikasi dapat dilatih secara optimal tanpa terpengaruh dominasi kelas tertentu.

3.4. Klasifikasi (*Decision Tree*)

Proses klasifikasi pada penelitian ini diterapkan melalui pendekatan algoritma *Decision Tree*, yang membentuk model pohon keputusan berdasarkan data atribut input untuk menentukan kelas target. Algoritma ini bekerja dengan membagi data berdasarkan nilai atribut tertentu yang paling berpengaruh terhadap klasifikasi target.

Model klasifikasi yang digunakan telah melalui proses prapemrosesan dan penyeimbangan data (SMOTE) untuk menghasilkan struktur pohon yang optimal. Atribut utama yang digunakan dalam pembentukan pohon adalah Jumlah, yang mewakili banyaknya kejadian atau korban pada tiap satuan wilayah.



Gambar 2. Visualisasi *Decision Tree*

Gambar 2 menunjukkan hasil struktur pohon keputusan yang dibentuk oleh model. Berdasarkan hasil tersebut, algoritma membagi data ke dalam tiga simpul keputusan berdasarkan nilai atribut Jumlah.

Tabel 5. Aturan Klasifikasi Berdasarkan Pohon Keputusan

No	Kondisi	Kategori Hasil
1	Jumlah ≤ 57	Rendah
2	Antara Jumlah $57 >$ dan ≤ 270.5	Sedang
3	Jumlah > 270.5	Tinggi

Berdasarkan pohon keputusan yang dihasilkan oleh model klasifikasi, atribut Jumlah menjadi faktor utama dalam menentukan tingkat keparahan kecelakaan. Jika nilai Jumlah kurang dari atau sama dengan 57, maka diklasifikasikan sebagai Rendah. Nilai antara 58 hingga 270.5 dikategorikan sebagai Sedang, dan selebihnya dikategorikan sebagai Tinggi.

3.5. Evaluasi Model

Evaluasi dilakukan untuk mengukur seberapa baik model klasifikasi membedakan tingkat keparahan kecelakaan lalu lintas. Model dibangun dengan algoritma *Decision Tree C4.5* menggunakan data yang telah diproses dan diseimbangkan dengan teknik SMOTE untuk mengatasi dominasi kelas mayoritas. Evaluasi ini penting sebagai dasar kelayakan model dalam mendukung keputusan terkait keselamatan lalu lintas.

1) Analisis Skema Validasi dan SMOTE

Perbandingan skema validasi dilakukan untuk mengidentifikasi konfigurasi *k-Fold Cross Validation* yang paling efektif dalam menghasilkan model klasifikasi dengan performa terbaik. Dalam penelitian ini, dilakukan pengujian dengan tiga nilai *k* yang berbeda, yaitu *k* = 5, *k* = 10, dan *k* = 15, untuk mengevaluasi pengaruh jumlah lipatan terhadap akurasi dan konsistensi hasil klasifikasi.

Tabel 6. Hasil Perbandingan *K-Fold*

K-Fold	Akurasi	Standar Deviasi (%)
K = 5	95.00%	$\pm 7.45\%$
K = 10	95.00%	$\pm 11.25\%$
K = 15	98.33%	$\pm 6.45\%$

Berdasarkan hasil evaluasi dengan variasi nilai k pada skema *k-Fold Cross Validation*, diketahui bahwa nilai $k = 15$ menghasilkan akurasi tertinggi sebesar 98.33% dengan deviasi standar $\pm 6.45\%$, yang berarti model memiliki performa yang lebih konsisten antar fold. Sementara itu, nilai $k = 10$ dan $k = 5$ sama-sama menghasilkan akurasi 95.00%, dengan deviasi standar yang lebih besar, yaitu $\pm 11.25\%$ dan $\pm 7.45\%$.

Oleh karena itu, meskipun nilai $k = 10$ sering direkomendasikan sebagai konfigurasi default dalam banyak literatur, dalam konteks dataset ini, penggunaan $k = 15$ memberikan hasil yang lebih optimal dan stabil. Maka, skema $k = 15$ dipilih sebagai pendekatan evaluasi utama dalam penelitian ini.

2) Confusion Matrix

Tabel 7. *Confusion Matrix*

	True Sedang	True Tinggi	True Rendah	Class Precision
Pred. Sedang	47	1	0	97.92%
Pred. Tinggi	0	9	0	100.00%
Pred. Rendah	0	0	3	100.00%
Class Recall	100.00%	90.00%	100.00%	

Hasil *confusion matrix* menunjukkan bahwa model mampu mengklasifikasikan data dengan akurasi tinggi pada semua kategori. Kelas Rendah dan Sedang memiliki tingkat ketepatan sempurna, ditunjukkan oleh nilai recall dan *precision* yang mencapai 100%. Sementara itu, pada kelas Tinggi, satu data teridentifikasi keliru sebagai Sedang, sehingga recall turun menjadi 90%. Meski demikian, semua prediksi pada kelas Tinggi tetap tepat sasaran, tercermin dari *precision* yang tetap 100%. Secara keseluruhan, model bekerja secara konsisten dan andal dalam mengenali pola tingkat keparahan kecelakaan.

Penelitian ini menghasilkan model klasifikasi tingkat keparahan kecelakaan lalu lintas di NTB menggunakan algoritma Decision Tree (C4.5), 15-Fold Cross Validation, dan SMOTE. Model ini mencapai akurasi 98,33% dengan nilai *precision*, *recall*, serta *F1-score* yang tinggi secara konsisten pada setiap kategori, menunjukkan kinerja yang andal meskipun data tidak seimbang. Dibandingkan dengan penelitian Franseda et al. yang menggunakan CHAID dan mencapai akurasi 88% serta *F1-score* 0,86, model ini terbukti lebih akurat dan seimbang. Dengan demikian, Informasi tersebut dapat digunakan sebagai dasar dalam menyusun kebijakan pencegahan, seperti peningkatan pengawasan lalu lintas, penempatan rambu di titik rawan, serta edukasi keselamatan berkendara yang ditargetkan sesuai tingkat risiko di tiap daerah.

4. Kesimpulan

Penelitian ini menghasilkan sebuah model klasifikasi untuk mengelompokkan tingkat keparahan kecelakaan lalu lintas di Provinsi Nusa Tenggara Barat dengan memanfaatkan algoritma *Decision Tree* (C4.5) dan teknik SMOTE. Model ini dibangun berdasarkan data sekunder dari BPS NTB tahun 2018–2023 yang telah melalui tahapan prapemrosesan dan penyeimbangan kelas. Hasil klasifikasi membagi data ke dalam tiga tingkat keparahan, yaitu Rendah, Sedang, dan Tinggi, dengan atribut "Jumlah" menjadi faktor utama dalam pengambilan keputusan klasifikasi. Pengujian model dilakukan menggunakan *k-Fold Cross Validation* dengan berbagai nilai k , dan menunjukkan bahwa $k=15$ memberikan performa terbaik, dengan akurasi mencapai 98,33% dan deviasi yang paling rendah dibandingkan skema lainnya.

Lebih dari sekadar akurasi model, hasil klasifikasi ini memberikan kontribusi nyata terhadap upaya penanggulangan kecelakaan. Dengan mengetahui wilayah atau waktu kejadian yang masuk dalam kategori risiko tinggi, instansi terkait seperti Dinas Perhubungan dan kepolisian dapat melakukan tindakan preventif yang lebih terfokus, seperti peningkatan pengawasan di titik rawan, pemasangan rambu tambahan, dan edukasi keselamatan berlalu lintas. Model ini juga dapat digunakan untuk menyusun prioritas intervensi berbasis data, sehingga kebijakan yang diterapkan lebih tepat sasaran. Selain sebagai alat bantu teknis, model ini berperan penting dalam perencanaan strategis untuk menekan angka kecelakaan di wilayah NTB.

Daftar Pustaka

- [1] A. K. Nalendra, M. Mujiono, R. Akhsani, and A. S. W. Utama, "Implementasi Algoritma K-Mean dalam Pengelompokan Data Kecelakaan di Kabupaten Kediri," *VOCATECH Vocat. Educ. Technol. J.*, vol. 1, no. 2, pp. 53–60, Apr. 2020, doi: 10.38038/vocatech.v1i2.28.

-
- [2] C. A. Salsabila, F. Yulianto, and T. A. Y. Siswa, "Implementasi Metode Naive Bayes Untuk Klasifikasi Kecelakaan Lalu Lintas di Kota Samarinda," *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5890.
 - [3] C. Silvia *et al.*, "Ketepatan Klasifikasi Tingkat Keparahan Korban Kecelakaan Lalu Lintas Menggunakan Metode Regresi Logistik Ordinal dan Fuzzy K-Nearest Neighbor In Every Class," vol. 4, pp. 441–451, 2015, [Online]. Available: <http://ejournal-s1.undip.ac.id/index.php/gaussian>
 - [4] N. Betris, D. Maretta, F. Yulianto, T. Azhima, and Y. Siswa, "Identifikasi Jenis Kecelakaan Lalu Lintas di Kota Samarinda Menggunakan Metode Random Forest," 2025.
 - [5] I. G. A. M. S. I Gede Agus Jiwadiana, I Komang Gde Sukarsa, "Klasifikasi Karakteristik Kecelakaan Lalu Lintas di Kota Denpasar dengan Pendekatan Classification And Regression Trees(CART)," 2015.
 - [6] D. Anitasari, F. Yulianto, T. Azhima, and Y. Siswa, "Klasifikasi Kecelakaan Lalu Lintas di Kota Samarinda Menggunakan Algoritma K-Nearest Neighbor," 2025.
 - [7] M. R. Qisthiano, P. A. Prayesy, and I. Ruswita, "Penerapan Algoritma Decision Tree dalam Klasifikasi Data Prediksi Kelulusan Mahasiswa," *G-Tech J. Teknol. Terap.*, vol. 7, no. 1, pp. 21–28, Jan. 2023, doi: 10.33379/gtech.v7i1.1850.
 - [8] R. Ade Putranto and T. Wuryandari, "Perbandingan Analisis Klasifikasi Antara Decision Tree dan Support Vector Machine Multiclass Untuk Penentuan Jurusan Pada Siswa Sma," vol. 4, pp. 1007–1016, 2015, [Online]. Available: <http://ejournal-s1.undip.ac.id/index.php/gaussian>
 - [9] D. H. M. Surya Prangga, Rito Goejantoro, Memi Nor Hayati, Siti Mahmuda, "Perbandingan Akurasi Klasifikasi Menggunakan Algoritma Quest Pada Skenario Data Kodifikasi dan Non-Kodifikasi," 2024.
 - [10] I. Sutoyo, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik," vol. 14, no. 2, 2018, [Online]. Available: www.bsi.ac.id
 - [11] J. T. Kumalasari, A. Merdekawati, and A. Hidayati, "Klasifikasi Multi Class Pada Metode Kerja Jarak Jauh Menggunakan Algoritma Decision Tree dan Imbalance Data," *J. Inf. Syst. Applied, Manag. Account. Res.*, vol. 8, no. 1, p. 109, Jan. 2024, doi: 10.52362/jisamar.v8i1.1350.
 - [12] A. Franseda, W. Kurniawan, S. Anggraeni, and W. Gata, "Integrasi Metode Decision Tree dan SMOTE untuk Klasifikasi Data Kecelakaan Lalu Lintas," *J. Sist. dan Teknol. Inf.*, vol. 8, no. 3, p. 282, Jul. 2020, doi: 10.26418/justin.v8i3.40982.
 - [13] F. S. Said Noor Abdullah, Siska Kurnia Gusti, Fitri Wulandari, "Klasifikasi Tingkat Keparahan Korban Kecelakaan Lalu Lintas Menggunakan Naive Bayes," 2024.
 - [14] Z. Vebryan Maharani, A. Lestari Arman Syah, and W. Ode Azzahra Astiani Hasiba, "Prediksi Jumlah Kecelakaan Lalu Lintas di Sulawesi Tenggara Menggunakan Metode Algoritma Regresi Linear," 2024.
 - [15] K. Octavia, F. Yulianto, T. Azhima, and Y. Siswa, "Penentuan Jenis Kecelakaan Lalu Lintas di Kota Samarinda Menggunakan Metode Decision Tree," 2025.