

PENENTUAN JENIS KECELAKAAN LALU LINTA DI KOTA SAMARINDA MENGUNAKAN METODE *DECISION TREE*

Khanisa Octavia¹, Fendy Yulianto², Taghfirul Azhima Yoga Siswa³

^{1, 2, 3} Teknik Informatika, Universitas Muhammadiyah Kalimantan Timur

Alamat e-mail : 2111102441172@umkt.ac.id¹, fy415@umkt.ac.id²
tay758@umkt.ac.id³

ABSTRACT

Traffic accidents are one of the main causes of death in Indonesia, so comprehensive research is needed to determine the causal factors. The aim of this research is to classify the level of traffic accidents in Samarinda City using the Decision Tree method. The data used is the result of data collection from the Samarinda City Sector Police taken from 2021 to 2024. Data collection, preprocessing, method training, and evaluation are the steps carried out in this research process. The Decision Tree method was chosen because it has the advantage of a simple and easy to understand structure. The evaluation was carried out using K-Fold Cross Validation to ensure the method's performance remains stable and avoids overfitting. The classification results in the research show that the Decision Tree method produces an accuracy of 93% with evaluation metrics such as precision, Recall, F1-Score which also show good performance. It is hoped that the results of this research will provide a useful contribution in creating data-based safety regulations and can help reduce accident rates in Samarinda City.

Keywords: Classification, Accidents, Decision Tree, K-Fold Cross Validation, Samarinda City

ABSTRAK

Kecelakaan lalu lintas merupakan salah satu penyebab utama kematian di Indonesia, sehingga diperlukan penelitian yang menyeluruh untuk menentukan faktor-faktor penyebabnya. Tujuan dari penelitian ini yaitu mengklasifikasi tingkat kecelakaan lalu lintas di Kota Samarinda dengan menggunakan metode *Decision Tree*. Data yang digunakan adalah hasil pengumpulan data dari Kepolisian Sektor Kota Samarinda yang diambil dari tahun 2021 hingga tahun 2024. Pengumpulan data, *Preprocessing*, pelatihan metode, dan evaluasi merupakan langkah-langkah yang dilakukan dalam proses penelitian ini. Metode *Decision Tree* dipilih karena memiliki keunggulan berupa struktur yang sederhana dan mudah dipahami, evaluasi dilakukan menggunakan *K-Fold Cross Validation* untuk memastikan performa metode tetap stabil dan menghindari *Overfitting*. Hasil klasifikasi pada penelitian menunjukkan bahwa metode *Decision Tree* menghasilkan akurasi sebesar 93% dengan metrik evaluasi seperti presisi, *Recall*, *F1-Score* yang juga

menunjukkan performa yang baik. Diharapkan bahwa hasil penelitian ini akan memberikan kontribusi yang bermanfaat dalam menciptakan peraturan keselamatan yang berbasis data dan dapat membantu menurunkan tingkat kecelakaan di Kota Samarinda.

Kata Kunci: Klasifikasi, Kecelakaan, *Decision Tree*, *K-Fold Cross Validation*, Kota Samarinda.

A. Pendahuluan

Kecelakaan lalu lintas adalah kejadian yang terjadi secara tidak terduga dengan melibatkan berbagai jenis kendaraan maupun pengguna jalan lainnya. Kondisi jalan buruk serta kecepatan kendaraan tinggi sehingga kerap menjadi faktor utama dapat memicu terjadinya kecelakaan (Siregar et al., 2023). Kecelakaan lalu lintas menjadi penyebab kematian tertinggi ketiga di Indonesia, dengan jumlah korban yang terus meningkat setiap tahunnya. Sehingga masih diperlukan peningkatan kesadaran hukum masyarakat Indonesia terhadap peraturan lalu lintas jalan agar mengurangi risiko kecelakaan lalu lintas (Gordon, 2019).

Kecelakaan lalu lintas tidak dapat dibiarkan begitu saja karena akan berpotensi menimbulkan kerugian dan korban jiwa yang tidak diinginkan, sehingga penting untuk memahami dan mengembangkan kesadaran individu agar dapat mencegah terjadinya kecelakaan lalu lintas (Agung et al., 2022). Kecelakaan lalu lintas sering kali terjadi dikarenakan adanya pengendara yang melanggar peraturan, sehingga untuk mengurangi jumlah kecelakaan, pihak berwenang perlu memberikan sanksi pidana kepada pengendara

yang terlibat dalam kecelakaan lalu lintas (Sari & Saleh, 2022)

Kecelakaan lalu lintas sering kali terjadi di kota – kota yang ada di Indonesia, salah satunya seperti yang terjadi di Kota Samarinda dengan 1004 kasus kecelakaan pada tahun 2021 hingga 2024, angka ini terbilang besar sehingga diperlukan analisa lebih lanjut untuk mengetahui penyebab kecelakaan lalu lintas (Ryamizard, 2021). Kecelakaan lalu lintas dapat diidentifikasi dengan berbagai macam cara seperti melakukan inspeksi langsung ke lapangan atau dengan menggunakan bantuan sistem dalam mengelompokkan jenis kecelakaannya, dalam melakukan pengelompokan dengan bantuan sistem dapat menggunakan konsep data *Mining* yang terdiri dari berbagai macam metode seperti *Random Forest*, *Naïve Bayes* dan *Decision Tree* (Lesmana, 2019).

Penelitian yang dilakukan oleh (Chen & Chen, 2020) dengan judul *Modeling Road Accident Severity with Comparisons Of Logistic Regression, Decision Tree and Random Forest* menggunakan metode *Decision Tree* CART, *Logistic Regression*, *Random Forest* didapatkan hasil akurasi pada metode *Decision Tree* CART

sebesar 72.65%, *Logistic Regression* sebesar 73.07%, *Random Forest* 73.38%. Selain itu penelitian selanjutnya yang dilakukan oleh Cicek et al., (2023) dengan data *Comparison Of Traffic Accident Injury Severity Prediction Models With Explainable Machine Learning* melakukan perbandingan metode terdiri dari metode *Decision Tree*, *K-Nearest Neighbor* (KNN), *Naïve Bayes* dan didapatkan hasil pada metode KNN sebesar 95,58%, *Naïve Bayes* sebesar 89,14%, serta *Decision Tree* sebesar 96,13%.

Metode *Decision Tree* memiliki keunggulan berupa metode yang sederhana dan mudah dipahami, dengan metode pengelompokan yang serupa dengan struktur pohon keputusan, sehingga memudahkan untuk menganalisis hasil penelitian (Tangkelayuk, 2022). Menurut penelitian sebelumnya yang dilakukan oleh (Puspita et al., 2021) tentang Analisis Sentimen Pengguna Layanan BPJS menunjukkan bahwa metode *Decision Tree* lebih unggul dengan akurasi 96.13% di bandingkan dengan metode *Naïve Bayes* sebesar 89.14%.

Pada penelitian-penelitian sebelumnya, *Decision Tree* digunakan untuk mengklasifikasi kasus kecelakaan lalu lintas, perbedaan dari penelitian ini dengan penelitian lainnya adalah parameter yang digunakan untuk menentukan kinerja performa model dalam klasifikasi kecelakaan lalu lintas berdasarkan tingkat kecelakaan. Pada penelitian ini mengambil kasus kecelakaan lalu lintas di Kota

Samarinda, data yang diperoleh berasal dari Kepala Satuan Lalu Lintas Kepolisian Sektor Kota (Polsek) Samarinda, bertujuan agar memahami struktur kecelakaan untuk membantu kebijakan keselamatan berbasis data (Cicek et al., 2023).

Confusion Matrix digunakan untuk menilai kinerja model klasifikasi dengan mengidentifikasi jumlah prediksi yang benar dan salah untuk setiap kategori, dapat menghitung berbagai metrik evaluasi yang memberikan wawasan mendalam mengenai metrik utama akurasi, presisi, *Recall* dan *F1-Score* (Taghfirul Yoga Azhima et al., 2023). *Confusion Matrix* digunakan karena kemampuannya dalam mengelola permasalahan ketidakseimbangan data, memberikan nilai evaluasi secara keseluruhan, membantu dalam menentukan ketidakakuratan prediksi, dan berfungsi sebagai standar umum dalam penelitian (Sari & Haranto, 2019).

Pada penelitian ini metode *Decision Tree* akan digunakan untuk melakukan klasifikasi kecelakaan lalu lintas di kota Samarinda, sehingga metode *Decision Tree* diharapkan dapat mengklasifikasi jenis kecelakaan secara tepat dan memiliki akurasi yang baik. Penulis berharap hasil penelitian ini dapat bermanfaat bagi pihak berwenang dan memberikan himbauan kepada seluruh masyarakat Kota Samarinda untuk mengurangi angka kecelakaan di kota Samarinda.

B. Metode Penelitian

Data yang digunakan adalah hasil pengumpulan data dari Kepolisian Sektor Kota Samarinda yang diambil dari tahun 2021 hingga tahun 2024. Pengumpulan data, *Preprocessing*, pelatihan metode, dan evaluasi merupakan langkah-langkah yang dilakukan dalam proses penelitian ini. Metode *Decision Tree* dipilih karena memiliki keunggulan berupa struktur yang sederhana dan mudah dipahami. *Decision tree* memiliki beberapa seperti ID3, C4.5, CART, metode tersebut sangat populer didalam Klasifikasi ini termasuk dari data mining, data *Desicion Tree* dinyatakan dalam bentuk tabel dengan atribut dan *record*. Model yang telah dibuat akan dihitung tingkat Akurasi prediksinya menggunakan rumus (Franseda et al., 2020). Evaluasi dilakukan menggunakan *K-Fold Cross Validation* untuk memastikan performa metode tetap stabil dan menghindari *Overfitting*. *K-Fold Cross Validation* adalah Metode yang umum digunakan untuk mengevaluasi kinerja model dalam pembelajaran *Machine Learning*, dimana bertujuan mengukur seberapa baik model dapat melakukan normalisasi terhadap data penelitian (Lestari et al., 2020).

C. Hasil Penelitian dan Pembahasan Hasil

Berikut adalah tampilan hasil dari program yang telah dirancang untuk mengklasifikasikan Tingkat Kecelakaan di Kota Samarinda, hasil implementasi dianalisis berupa

penerapan metode *Decision Tree*, rasio, *K-Fold Cross Validation*, dan evaluasi penjelasan tentang langkah-langkah yang dilakukan disajikan berikut ini.

1. Transformasi Data

Label encoding dipakai buat ngubah data jadi vektor pas tahap Transformasi data di pemodelan data pakai metode *Decision Tree*. Data yang udah diubah bisa dilihat di Tabel 1.

Tabel 1. Biner

Atribut	Spesifikasi	Label
Cuaca	Hujan disertai Angin Kencang	1
	Angin Kencang	1
Kondisi Permukaan Jalan	Berawan	0
	Berombak	1
	Baik	0
	Licin	1
	Belubang	1
Kelas Jalan	Basah	1
	I (Jalan Besar utk beban s/d 10 ton & max 18 m Panjang Ran)	1
	II (Jalan Sedang utk beban s/d 10 ton & 12 m Panjang Ran)	1
	III (Jalan Kecil utk max beban 8 ton & 9 m Panjang Ran)	0
Kemiringan Jalan	Datar	0
	Menanjak	1
Tipe Jalan	1/1 (1 Lajur 1 Arah)	0
	2/2 TB (2 Lajur 2 Arah Tanpa Batas Median)	0
	4/2 B (4 Lajur 2 Arah dengan Batas Median)	1
Batas Kecepatan di Lokasi	20	0
	30	0
	40	0
	50	1
	60	1
	80	1
Status Jalan	Jalan Provinsi	1
	Jalan Kota	0
Status Jalan	Jalan Nasional	1
	Jalan Desa	0

Tabel 1 melihatkan perubahan data menjadi label biner yang melihatkan pengaruhnya terhadap Tingkat Kecelakaan, (0) memiliki nilai tidak berpengaruh, dan (1) memiliki nilai berpengaruh. Tingkat Kecelakaan di Klasifikasikan sebagai (0) ringan, (1) sedang, dan (2) berat.

2. Pembagian Data

Data yang digunakan dalam pemodelan ini dibagi menjadi 70% untuk data latih dan 30% untuk data uji, seperti yang terlihat pada Gambar 1.

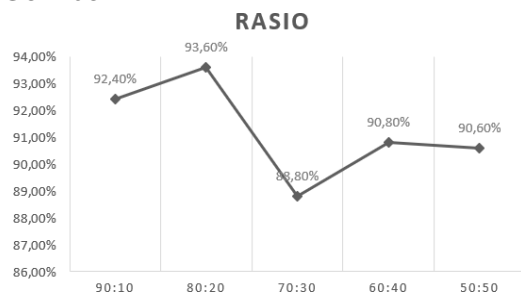
Jumlah data latih : 702
 Jumlah data uji : 302

Gambar 1. Jumlah Data

Pada Gambar 1 data yang ditampilkan adalah 702 untuk data *Training* dan 302 untuk data *Testing*. Pembagian data dengan 70% untuk data latih dan 30% untuk data uji ini menghasilkan tingkat Akurasi yang tinggi.

3. Pengujian Rasio Data

Pengujian Rasio dilakukan pembagian data yang berbeda yaitu 90:10, 80:20, 70:30, 60:40, 50:50. Proses ini nilai rasio yang digunakan pada pengujian ini dapat dilihat pada Gambar 2.



Gambar 2. Pengujian Rasio

Pada Gambar 2 akurasi rasio 80:20 dengan nilai 93,60% adalah Rasio pembagian tertinggi dari pengujian Rasio dengan 20% data *Testing* dan 80% data *Training*, maka pada pengujian selanjutnya Rasio yang digunakan adalah 80:20.

4. Nilai Awal K-Fold Cross Validation

Dalam *K-Fold Cross Validation*, setiap iterasi menggunakan satu bagian sebagai data uji, sementara bagian lainnya sebagai data latih. Dalam penelitian ini, K-Fold dengan nilai 5 digunakan saat menginisialisasi *K-Fold Cross Validation* dapat dilihat pada Gambar 3.

	Fold	Index	Actual (Biner)	Predicted (Biner)
0	1	0	1	1
1	1	1	1	1
2	1	2	0	0
3	1	3	0	0
4	1	4	1	1
...	...	...	...	...
5015	5	999	0	0
5016	5	1000	0	1
5017	5	1001	1	1
5018	5	1002	1	1
5019	5	1003	1	1

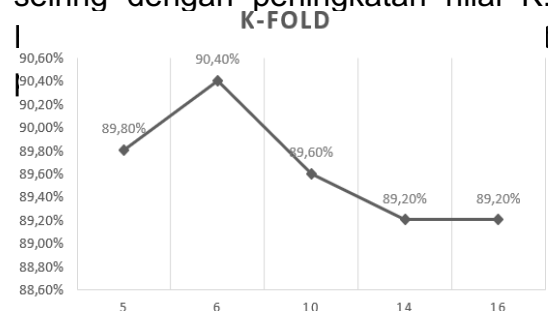
[5020 rows x 4 columns]

Gambar 3. Hasil Pengujian K-Fold 5

Pada Gambar 3 menunjukkan hasil evaluasi model menggunakan *K-Fold Cross-Validation* dengan tabel berisi 5020 baris dan 4 kolom *Fold*, *Index*, *Actual*, dan *Predicted*. Tabel ini digunakan untuk mengevaluasi performa model pada setiap lipatan, termasuk menghitung metrik evaluasi seperti akurasi, presisi, *Recall*, dan *F1-score*.

5. Pengujian K-Fold Cross Validation

K-Fold adalah Metode dalam *Cross Validation* yang dapat digunakan untuk mengevaluasi kinerja model *Machine Learning*. Pengujian ini dilakukan dengan menggunakan nilai berbeda-beda, yaitu 5, 6, 10, 14 dan 16. Setiap nilai K dilakukan Proses pengujian diulang sebanyak 5 kali dan menunjukan bahwa Akurasi model dapat berubah seiring dengan peningkatan nilai K.



Gambar 4. Pengujian K-Fold Cross Validation

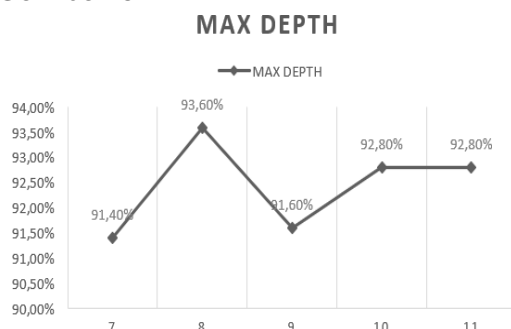
Pada Gambar 4 Akurasi nilai K 6 dengan nilai 90.40% adalah nilai K tertinggi dari pengujian *K-Fold*, dan nilai K yang digunakan untuk pengujian selanjutnya adalah 6.

6. Nilai Awal Parameter Decision Tree

Setiap parameter memiliki peran dalam menyusun, membentuk, dan meningkatkan kinerja Metode *Decision Tree* maka dari itu menetapkan parameter awal sangatlah penting dalam menggunakan Metode *Decision Tree*. Berdasarkan penelitian (Rian Oktafiani et al., 2024) nilai parameter *Max Depth* tertinggi terdapat pada angka 7, nilai ini digunakan sebagai Nilai Awal parameter.

7. Max_Depth

Pada pengujian parameter *Max_Depth* digunakan untuk membatasi kedalaman maksimal pohon keputusan. Pengujian parameter *Max_Depth* ini menggunakan nilai 7, 8, 9, 10, 11, hasil dari pengujian dapat dilihat pada Gambar 5.



Gambar 5. Pengujian Parameter

Pada Gambar 5 dapat dilihat bahwa nilai parameter *Max_Depth* yang memiliki Akurasi tertinggi adalah 8 dengan Akurasi sebesar 93,60%, sehingga nilai parameter *Max_Depth* yang digunakan pada hasil akhir adalah 8.

8. Perhitungan Entropy Total

Entropy adalah konsep dari teori informasi yang mengukur tingkat ketidakpastian atau ketidakteraturan dalam sebuah dataset. Dalam pembelajaran mesin, terutama pada metode pohon keputusan (*Decision Tree*), *Entropy* digunakan untuk mengevaluasi seberapa baik sebuah atribut dapat memisahkan data ke dalam berbagai kategori. Menghitung *Entropy* menggunakan Persamaan 2.1.

$$Entropy = \left(\left(-\frac{5}{10} \right) \times \log_2 \left(\frac{5}{10} \right) + \left(-\frac{3}{10} \right) \times \log_2 \left(\frac{3}{10} \right) + \left(-\frac{2}{10} \right) \times \log_2 \left(\frac{2}{10} \right) \right) = 1,485475297$$

Variabel yang digunakan analisis ini, termasuk atribut penting yang memiliki pengaruh signifikan pada hasil perhitungan *Entropy*, serangkaian peran diperlukan untuk menentukan kualitas model Klasifikasi data. Variabel yang dimaksud mungkin melibatkan berbagai faktor, termasuk variabel Kondisi Cahaya hingga variabel Status Jalan, yang semuanya berkontribusi terhadap kompleksitas perhitungan. Data kecelakaan lalu lintas dapat dilihat pada Tabel 2.

Tabel 2. Dataset

Kondisi Cahaya	Cuaca	Kelas Jalan	Tipe Jalan	Kondisi Permukaan Jalan	Batas Kecepatan	Kemiringan Jalan	Status Jalan	Tingkat Kecelakaan
----------------	-------	-------------	------------	-------------------------	-----------------	------------------	--------------	--------------------

Terang	Cerah	I (Jalan Besar utk beban 10 ton & max 18m Panjang Ran)	2/2 TB(2 Lajur/2 Arah Tanpa Batas Median)	Licin	20	Datar	Jalan Provinsi	Ringan
Gelap	Hujan	I (Jalan Besar utk beban 10 ton & max 18m Panjang Ran)	4/2 B (4 Lajur/2 Arah dengan Batas Median)	Licin	50	Menanjak	Jalan Provinsi	Berat
Gelap	Cerah	III (Jalan Kecil utk max beban 8 ton & 9m Panjang Ran)	2/2 TB(2 Lajur/2 Arah Tanpa Batas Median)	Berombak	30	Datar	Jalan Desa	Ringan
Gelap	Cerah	II (Jalan Sedang utk beban s/d 10 ton & 12m Panjang Ran)	4/2 B (4 Lajur/2 Arah dengan Batas Media)	Baik	60	Menanjak	Jalan Kota	Sedang
Kondisi Cahaya	Cuaca	Kelas Jalan	Tipe Jalan	Kondisi Permukaan Jalan	Batas Kecepatan	Kemiringan Jalan⁶	Status Jalan	Tingkat Kecelakaan
Terang	Cerah	III (Jalan Kecil utk max beban 8 ton & 9m Panjang Ran)	2/2 TB(2 Lajur/2 Arah Tanpa Batas Median)	Baik	30	Menanjak	Jalan Kota	Ringan
Gelap	Hujan	I (Jalan Besar utk beban 10 ton & max 18m Panjang Ran)	2/2 TB(2 Lajur/2 Arah Tanpa Batas Median)	Licin	50	Datar	Jalan Nasional	Sedang

Pada Tabel 2 menampilkan data yang digunakan untuk menghitung *Entropy*, data memiliki 10 data yang berisi berbagai atribut seperti Kondisi Cahaya, Cuaca, Kelas Jalan, Tipe

Jalan, Kondisi Permukaan Jalan, Batas Kecepatan, Kemiringan Jalan, Status Jalan, Tingkat Kecelakaan. Hasil perhitungan *Entropy* terdapat pada Tabel 3.

Tabel 3. Perhitungan *Entropy*

Atribut	Jumlah	Ringan	Sedang	Berat	<i>Entropy</i>
Total	10	5	3	2	1,485475297

Berdasarkan Tabel 3 menunjukkan hasil perhitungan dari *Entropy* Keseluruhan untuk semua dataset yang ditampilkan pada Tabel 2. Pada perhitungan tersebut jumlah datanya adalah 10 dengan 5 data diklasifikasikan sebagai Ringan, 3

data diklasifikasikan sebagai Sedang, dan 2 data diklasifikasikan sebagai Berat. Nilai *Entropy* dari perhitungan tersebut adalah 1,485475297, nilai *Entropy* ini digunakan sebagai acuan untuk menghitung *Gain* dari setiap atribut.

9. Perhitungan *Entropy* dan *Gain* Tiap Atribut

Data dalam dataset dibagi berdasarkan setiap atribut, kemudian dilakukan perhitungan *Entropy* dan *Gain* untuk masing-masing atribut. *Gain* menggambarkan sejauh mana pembagian data berdasarkan atribut tersebut dapat mengurangi ketidakpastian. Perhitungan *Entropy* untuk setiap atribut dilakukan dengan menggunakan Persamaan 2.1.

$$Entropy = \left(\left(-\frac{1}{3} \right) \times \log_2 \left(\frac{1}{3} \right) + \left(-\frac{1}{3} \right) \times \log_2 \left(\frac{1}{3} \right) + \left(-\frac{1}{3} \right) \times \log_2 \left(\frac{1}{3} \right) \right) = 1,584962501$$

Perhitungan *Gain* pada setiap atribut dengan menggunakan Persamaan 2.2, menjadi penentu jumlah informasi yang didapat dari setiap atribut, sehingga bisa membantu keputusan yang lebih akurat dan efektif dalam analisis data. Dapat dilihat pada Tabel 4.

$$Gain = (0) - \left(\left(\frac{4}{10} \right) \times 0 \right) - \left(\left(\frac{6}{10} \right) \times 1,459147917 \right) = 0,512710019$$

Tabel 4. Perhitungan *Entropy* dan *Gain*

	Keterangan	Jumlah Kasus	Ringan	Sedang	Berat	<i>Entropy</i>	<i>Gain</i>
Total		10	5	3	2	1,4854753	
							0,512710019
Kondisi Cahaya	Terang	4	4	0	0	0	
	Gelap	6	1	3	2	1,45914792	
							1,485475297
Cuaca	Cerah	6	5	1	0	0	
	Hujan	3	0	2	1	0	
	Hujan disertai angin kencang	1	0	0	1	0	
	Angin Kencang	0	0	0	0	0	
	Berawan	0	0	0	0	0	
							0,885475297
Kelas Jalan	I (Jalan Besar utk beban 10 ton & max 18 m Panjang Ran)	4	1	1	2	1,5	
	II (Jalan Sedang utk beban s/d 10 ton & 12 m Panjang Ran)	2	0	2	0	0	
	III (Jalan Kecil utk max beban 8 ton & 9 m Panjang Ran)	4	4	0	0	0	
							0,609986547
Tipe Jalan	1/1 (1 Lajur/1 Arah)	2	2	0	0	0	
	2/2 TB (2 Lajur/2 Arah Tanpa Batas Median)	6	3	2	1	1,45914792	
	4/2 B (4 Lajur/2 Arah dengan Batas Median)	2	0	1	1	0	
							0,885475297
Kondisi Permukaan Jalan	Licin	4	1	2	1	1,5	
	Berombak	3	3	0	0	0	
	Berlubang	0	0	0	0	0	
	Basah	1	0	0	1	0	
	Baik	2	1	1	0	0	
							1,485475297
Batas Kecepatan	20	2	2	0	0	0	
	30	2	2	0	0	0	
	40	1	1	0	0	0	
	50	4	0	2	2	0	

	60	1	0	1	0	0	
	80	0	0	0	0	0	
							0,72451125
Kemiringan Jalan	Datar	5	3	2	0	0	
	Menanjak	5	2	1	2	1,52192809	
							0,72451125
Status Jalan	Jalan Provinsi	5	2	1	2	1,52192809	
	Jalan Desa	2	2	0	0	0	
	Jalan Kota	2	1	1	0	0	
	Jalan Nasional	1	0	1	0	0	

Pada Tabel 4 menunjukkan bahwa *Entropy* dan *Gain* digunakan untuk menentukan atribut terbaik untuk membuat pohon keputusan. *Entropy* digunakan untuk mengukur tingkat ketidakpastian, sementara *Gain* untuk mengukur sejauh mana suatu atribut dapat mengurangi ketidakpastian, dengan atribut yang memiliki *Gain* tertinggi dianggap paling efektif dalam membagi data dan membentuk pohon keputusan.

10. Nilai Awal Gain dan Entropy

Dengan Metode *Decision Tree* telah dilakukan perhitungan nilai *Gain* pada setiap atribut dan *Entropy* Keseluruhan pada dataset Kecelakaan Kota Samarinda. Hasil perhitungan *Gain* dan *Entropy* yang memberikan data sangat penting terkait tingkat keefektifan atribut dalam mengurangi ketidakpastian dan meningkatkan akurasi klasifikasi dapat dilihat pada Gambar 6.

Entropy Total: 1.109

Gain Seluruh Atribut Kondisi Cahaya: 0.12

Gain Seluruh Atribut Cuaca: 0.084

Gain Seluruh Atribut Kelas Jalan: 0.098

Gain Seluruh Atribut Tipe Jalan: 0.1
Gain Seluruh Atribut Kondisi Permukaan Jalan: 0.098

Gain Seluruh Atribut Kondisi Permukaan Jalan: 0.098

Gain Seluruh Atribut Batas Kecepatan Dilokasi: 0.206
Gain Seluruh Atribut Kemiringan Jalan: 0.04

Gain Seluruh Atribut Kemiringan Jalan: 0.04

Gain Seluruh Atribut Status Jalan: 0.097

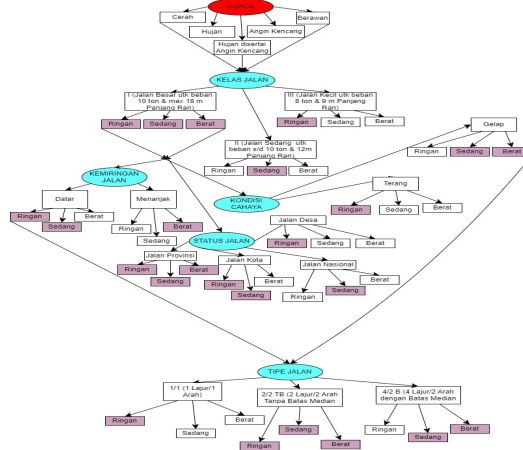
Gambar 6. *Entropy* Keseluruhan dan *Gain* Dataset

Pada Gambar 6 nilai *Entropy* Keseluruhannya adalah 1.109,

dengan nilai *Gain* Kondisi Cahaya sebesar 0.12, *Gain* Cuaca sebesar 0.084, *Gain* Kelas Jalan sebesar 0.098, *Gain* Tipe Jalan sebesar 0.1, *Gain* Kondisi Permukaan Jalan sebesar 0.098, *Gain* Batas Kecepatan Dilokasi sebesar 0.206, *Gain* Kemiringan Jalan sebesar 0.04, dan *Gain* Status Jalan sebesar 0.097.

11. Pohon Keputusan

Atribut dengan nilai *Gain* tertinggi dipilih untuk menjadi node dalam pohon keputusan, langkah ini dilakukan berulang kali untuk tiap node, dengan tiap node memiliki atribut yang paling sesuai, ketika terbentuk pohon keputusan digunakan untuk menghasilkan prediksi. Langkah-langkah ini diulangi untuk setiap cabang baru hingga pohon dapat dilihat pada gambar 7.



Gambar 7. Pohon Keputusan

Pada Gambar 7 merupakan pohon keputusan yang digunakan untuk mengklasifikasikan tingkat kecelakaan lalu lintas ke dalam kelas ringan, sedang atau berat, dengan menentukan faktor seperti cuaca, kelas jalan, kemiringan jalan, status jalan, kondisi cahaya dan tipe jalan, dalam menentukan keputusan akhir.

12. Data Label, Aktual dan Klasifikasi

Terdapat Data Label, Aktual dan Klasifikasi yang telah dihasilkan oleh pemodelan yang dapat dilihat pada Gambar 8.

147	Ringan	Ringan
148	Ringan	Sedang
149	Ringan	Ringan
150	Sedang	Sedang

Gambar 8. Hasil Data Aktual dan Klasifikasi

Pada Gambar 8 perbandingan nilai aktual dan prediksi untuk setiap data uji, contoh data uji 147 dengan label aktual 'Ringan' berhasil diprediksi dengan benar, dan pada data uji nomor 148 yang seharusnya memiliki label 'Ringan' tetapi justru diprediksi 'Sedang', hal ini menunjukkan seberapa baik metode *Decision Tree* dalam melakukan klasifikasi.

13. Evaluasi

Model pohon keputusan yang telah dibuat dievaluasi untuk menilai kinerjanya. Penilaian dapat dilakukan menggunakan metrik seperti Akurasi, Presisi, *Recall*, dan *F1-Score*. Berikut ini adalah *Confusion Matrix* untuk menentukan Tingkat Kecelakaan Lalu Lintas dapat dilihat pada Tabel 5.

Tabel 5. Confusion Matrix

Kelas	TP	TN	FP	FN
Ringan (0)	6	0	0	7
Sedang (1)	4	0	0	9
Berat (2)	3	0	0	10

Pada Tabel 5 melihat hasil evaluasi Metode *Decision Tree* menggunakan *Confusion Matrix* agar dapat mengklasifikasikan Kecelakaan Lalu Lintas ke dalam tiga kelas Ringan (0), Sedang (1), Berat (2). Pada kelas Ringan (0) terdapat *True Positive* (TP) 6, *True Negative* (TN) 0, *False Positive* (FP) 0, *False Negative* (FN) 7. Kelas Sedang (1) terdapat *True Positive* (TP) 4, *True Negative* (TN) 0 *False Positive* (FP) 0, *False Negative* (FN) 9. Kelas Berat (2)

terdapat *True Positive* (TP) 3, *True Negative* (TN) 0 *False Positive* (FP) 0, *False Negative* (FN) 10.

Setelah mendapatkan nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), *False Negative* (FN) , selanjutnya model dievaluasi dengan menghitung Akurasi, Presisi, *Recall* dan *F1-Score*. Berikut adalah perhitungan Akurasi, Presisi, *Recall*, *F1-Score*.

Hasil Kelas Negatif (0)

Perhitungan Presisi tiap atribut menggunakan Persamaan 2.4, yang berfungsi sebagai acuan utama untuk memastikan bahwa hasil analisis yang dihasilkan memenuhi standar ketepatan yang diharapkan.

$$\begin{aligned}\text{Presisi} &= \frac{TP}{TP + FP} \times 100\% \\ &= \frac{6}{6 + 0} \times 100\% \\ &= \frac{6}{6} \times 100\% = 1.00 \\ &= 100\%\end{aligned}$$

Perhitungan *Recall* tiap atribut menggunakan Persamaan 2.5, yang berfungsi sebagai acuan utama dalam menilai performa model, terutama dalam hal mengidentifikasi atribut yang signifikan.

$$\begin{aligned}\text{Recall} &= \frac{TP}{TP + FN} \times 100\% \\ &= \frac{6}{6 + 0} \times 100\% \\ &= \frac{6}{6} \times 100\% = 1.00 \\ &= 100\%\end{aligned}$$

Perhitungan *F1-Score* tiap atribut menggunakan Persamaan 2.6, proses ini melibatkan penggabungan nilai Presisi dan *Recall* menjadi satu metrik untuk mengevaluasi keseimbangan antara keduanya.

$$\begin{aligned}F1 - \text{Score} &= 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \\ &= 2 \times \frac{1.00 \times 1.00}{1.00 + 1.00} \\ &= 2 \times \frac{1.00}{2.00} = 1.00 \\ &= 100\%\end{aligned}$$

Hasil Kelas Positif (1)

Perhitungan Presisi tiap atribut menggunakan Persamaan 2.4, yang berfungsi sebagai acuan utama untuk

memastikan bahwa hasil analisis yang dihasilkan memenuhi standar ketepatan yang diharapkan.

$$\begin{aligned}\text{Presisi} &= \frac{TP}{TP + FP} \times 100\% \\ &= \frac{4}{4 + 0} \times 100\% \\ &= \frac{4}{4} \times 100\% = 1.00 \\ &= 100\%\end{aligned}$$

Perhitungan *Recall* tiap atribut menggunakan Persamaan 2.5, yang berfungsi sebagai acuan utama dalam menilai performa model, terutama dalam hal mengidentifikasi atribut yang signifikan.

$$\begin{aligned}\text{Recall} &= \frac{TP}{TP + FN} \times 100\% \\ &= \frac{4}{4 + 0} \times 100\% \\ &= \frac{4}{4} \times 100\% = 1.00 \\ &= 100\%\end{aligned}$$

Perhitungan *F1-Score* tiap atribut menggunakan Persamaan 2.6, proses ini melibatkan penggabungan nilai Presisi dan *Recall* menjadi satu metrik untuk mengevaluasi keseimbangan antara keduanya.

$$\begin{aligned}F1 - \text{Score} &= 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \\ &= 2 \times \frac{1.00 \times 1.00}{1.00 + 1.00} \\ &= 2 \times \frac{1.00}{2.00} = 1.00 \\ &= 100\%\end{aligned}$$

Hasil Positif Kelas (2)

Perhitungan Presisi tiap atribut menggunakan Persamaan 2.4, yang berfungsi sebagai acuan utama untuk memastikan bahwa hasil analisis

yang dihasilkan memenuhi standar ketepatan yang diharapkan.

$$\begin{aligned}\text{Presisi} &= \frac{TP}{TP + FP} \times 100\% \\ &= \frac{3}{3 + 0} \times 100\% \\ &= \frac{3}{3} \times 100\% = 1.00 \\ &= 100\%\end{aligned}$$

Perhitungan *Recall* tiap atribut menggunakan Persamaan 2.5, yang berfungsi sebagai acuan utama dalam menilai performa model, terutama dalam hal mengidentifikasi atribut yang signifikan.

$$\begin{aligned}\text{Recall} &= \frac{TP}{TP + FN} \times 100\% \\ &= \frac{3}{3 + 0} \times 100\% \\ &= \frac{3}{3} \times 100\% = 1.00 \\ &= 100\%\end{aligned}$$

Perhitungan *F1-Score* tiap atribut menggunakan Peersamaan 2.6, proses ini melibatkan penggabungan nilai Presisi dan *Recall* menjadi satu metrik untuk mengevaluasi keseimbangan antara keduanya.

$$\begin{aligned}F1 - Score &= 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \\ &= 2 \times \frac{1.00 \times 1.00}{1.00 + 1.00} \\ &= 2 \times \frac{1.00}{2.00} = 1.00 \\ &= 100\%\end{aligned}$$

14. Pengujian Awal dan Akhir

Hasil pengujian awal dan akhir menunjukkan perubahan pada rasio, *K-Fold Cross Validation*, parameter *Max Depth*. Terdapat peningkatan hasil setelah melakukan pengujian dapat dilihat pada Tabel 6.

Tabel 6. Hasil Pengujian Awal dan Akhir

Aspek	Pengujian Awal	Pengujian Akhir	Perbandingan
Rasio	70% data latih dan 30% data uji (Widiastiwi & Ernawati, 2021)	Rasio terbaik dari pengujian yaitu 80:20	Pada pengujian awal menggunakan pembagian data yang tertinggi dari jurnal, sementara pada pengujian akhir menggunakan pembagian data yang telah diuji pada model.
<i>K-Fold Cross Validation</i>	<i>K-Fold</i> dengan K = 5 Hamdani et al., (2022)	<i>K-Fold</i> terbaik dari pengujian yaitu K = 6	Pada pengujian awal menggunakan nilai K yang tertinggi dari jurnal, sementara pada pengujian akhir menggunakan nilai K yang telah diuji pada model.
Parameter <i>Max_Depth</i>	<i>Max_Depth</i> = 7	<i>Max_Depth</i> terbaik dari pengujian yaitu 8	Pada pengujian awal menggunakan <i>Max_Depth</i> yang tertinggi dari jurnal, sementara pada pengujian akhir menggunakan <i>Max_Depth</i> yang telah diuji pada model.
Hasil Akurasi	Mendapatkan hasil akurasi sebesar 92%	Mendapatkan hasil akurasi sebesar 93%	Pada pengujian akhir menunjukkan peningkatan performa sebesar 1% yang memberikan hasil terbaik
Hasil <i>Precesion</i>	Kelas Ringan 96% Kelas Sedang 88% Kelas Berat 72%	Kelas Ringan 99% Kelas Sedang 86% Kelas Berat 100%	Pada pengujian awal terdapat hasil Kelas Ringan 96%, Sedang 88%, Berat 72%, sedangkan untuk pengujian akhir terdapat peningkatan 3% pada Kelas Ringan, pada Kelas Sedang terdapat penurunan 2% dan pada Kelas Berat terdapat peningkatan sebesar 28%.
Hasil <i>Recall</i>	Kelas Ringan 93%	Kelas Ringan 87%	Pada pengujian awal terdapat hasil Kelas Ringan 93%, Sedang 89%, Berat 100%,

	Kelas 89%	Sedang	Kelas 99%	Sedang	sedangkan untuk pengujian akhir terdapat penurunan 6% pada Kelas Ringan, pada Kelas Sedang terdapat peningkatan 10% dan pada Kelas Berat tidak terdapat peningkatan ataupun penurunan.
	Kelas 100%	Berat	Kelas 100%	Berat	
Hasil <i>F1-Score</i>	Kelas 94%	Ringan	Kelas 93%	Ringan	Pada pengujian awal terdapat hasil Kelas Ringan 94%, Sedang 88%, Berat 84%, sedangkan untuk pengujian akhir terdapat penurunan 1% pada Kelas Ringan, pada Kelas Sedang terdapat peningkatan 4% dan pada Kelas Berat terdapat peningkatan sebesar 16%.
	Kelas 88%	Sedang	Kelas 92%	Sedang	
	Kelas 84%	Berat	Kelas 100%	Berat	

Pada Tabel 6 menunjukkan perbandingan hasil pengujian awal dan akhir seperti rasio, *K-Fold Cross Validation*, parameter *Decision Tree* dan evaluasi model, dengan tabel ini dapat disimpulkan bahwa penyempurnaan model mendapatkan peningkatan.

Pembahasan

Hasil penelitian menunjukkan bahwa pembagian data dan *K-Fold Cross Validation* memiliki pengaruh signifikan terhadap performa model. Pengujian awal menggunakan pembagian data 70:30 dengan nilai $K = 5$ menghasilkan akurasi sebesar 92%, sedangkan pada pengujian akhir pembagian data 80:20 dengan nilai $K = 6$ mendapatkan hasil 93%. Hal ini menunjukkan bahwa pemilihan rasio data dan jumlah *Fold* lebih optimal dapat meningkatkan performa model.

Penyesuaian parameter *Decision Tree* juga memberikan hasil yang baik terhadap performa model. Pada pengujian awal parameter *Max_Depth* 7 mendapatkan hasil 92% dan pada pengujian akhir menggunakan *Max_Depth* 8 mendapatkan hasil 93%, evaluasi menunjukkan peningkatan pada *Precision* dimana kelas ringan

meningkat dari 86% menjadi 90% kelas sedang dari 85% menjadi 88% dan kelas berat dari 72% menjadi 74%. Pada *Recall* kelas ringan tetap 93%, kelas sedang meningkat dari 93% menjadi 97%, dan kelas berat tetap 100% , *F1-Score* juga menunjukkan peningkatan signifikan dengan kelas ringan naik dari 94% menjadi 95%, kelas sedang 92% menjadi 98% dan kelas berat 84% menjadi 100%. Hal ini menunjukkan bahwa parameter yang disesuaikan dengan baik dapat meningkatkan kemampuan model secara keseluruhan.

E. Kesimpulan

Penelitian Klasifikasi Kecelakaan Lalu Lintas di Kota Samarinda menggunakan metode *Decision Tree* menghasilkan beberapa kesimpulan sebagai berikut : (i) Penelitian ini menunjukkan bahwa metode *Decision Tree* dapat digunakan secara efektif untuk klasifikasi tingkat kecelakaan lalu lintas di Kota Samarinda, dengan hasil 93% dan evaluasi yang konsisten seperti presisi, *Recall*, dan *F1-Score*. (ii) Parameter terbaik untuk model klasifikasi ditemukan pada *Max_Depth* 8, *K-Fold Cross Validation* 6 dan rasio 80:20. Hasil ini

menunjukkan bahwa metode *Decision Tree* mampu memberikan hasil yang stabil dan akurat.

DAFTAR PUSTAKA

- Agung, J., Madiung, B., & Makkawaru, Z. (2022). Analisis Pelaksanaan UU Nomor 22 Tahun 2009 Terhadap Resiko Kecelakaan Lalu Lintas Pada Jalan Nasional Provinsi Sulawesi Barat. *Indonesian Journal of Legality of Law*, 4(2), 117–123.
<https://doi.org/10.35965/ijlf.v4i2.1467>
- Chen, M. M., & Chen, M. C. (2020). Modeling road accident severity with comparisons of logistic regression, decision tree and random forest. *Information (Switzerland)*, 11(5).
<https://doi.org/10.3390/INFO11050270>
- Cicek, E., Akin, M., Uysal, F., & Topcu Aytas, R. M. (2023). Comparison of traffic accident injury severity prediction models with explainable machine learning. *Transportation Letters*, 15(9), 1043–1054.
<https://doi.org/10.1080/19427867.2023.2214758>
- Franseda, A., Kurniawan, W., Anggraeni, S., & Gata, W. (2020). Integrasi Metode Decision Tree dan SMOTE untuk Klasifikasi Data Kecelakaan Lalu Lintas. *Jurnal Sistem Dan Teknologi Informasi (Justin)*, 8(3), 282.
<https://doi.org/10.26418/justin.v8i3.40982>
- Gordon, G. (2019). BAB 2 Tinjauan Pustaka. *Pontificia Universidad Catolica Del Peru*, 8(33), 44.
- Hamdani, H., Hatta, H. R., Puspitasari, N., Septiarini, A., & Henderi. (2022). Dengue classification method using support vector machines and cross-validation techniques. *IAES International Journal of Artificial Intelligence*, 11(3), 1119–1129.
<https://doi.org/10.11591/ijai.v11.i3.pp1119-1129>
- Lesmana, A. (2019). [3]View of Penggunaan Metode Klasifikasi Data Mining Untuk Prediksi Kecelakaan Lalu Lintas Berbasis IOT _ Tugas Mahasiswa Program Studi Informatika.pdf.
- Lestari, S., Akmaludin, A., & Badrul, M. (2020). Implementasi Klasifikasi Naive Bayes Untuk Prediksi Kelayakan Pemberian Pinjaman Pada Koperasi Anugerah Bintang Cemerlang. *PROSISKO: Jurnal Pengembangan Riset Dan Observasi Sistem Komputer*, 7(1), 8–16.
<https://doi.org/10.30656/prosisko.v7i1.2129>
- Puspita, R., & Widodo, A. (2021). Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS. *Jurnal Informatika Universitas Pamulang*, 5(4), 646.
<https://doi.org/10.32493/informatika.v5i4.7622>

- Ryamizard, M. I. (2021). Analisa Kecelakaan Lalu Lintas Studi Kasus Kota Malang. Institut Teknologi Sepuluh November. <https://doi.org/10.35957/jatisi.v9i2.2048>.
- Sari, B. W., & Haranto, F. F. (2019). Implementasi Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Telkom Dan Biznet. *Jurnal Pilar Nusa Mandiri*, 15(2), 171–176. <https://doi.org/10.33480/pilar.v15i2.699>
- Sari, N., & Saleh, K. (2022). Tinjauan yuridis penerapan sanksi pidana pada kecelakaan lalu lintas yang menyebabkan korban jiwa menurut pasal 310 undang-undang nomor 22 tahun 2009 tentang lalu lintas dan angkutan jalan. *Jurnal Politik Dan Pemerintahan Daerah*, 4(2), 282-292.
- Siregar, A. Z., Awaluddin, M., & Wahyuddin, Y. (2023). Identifikasi Tingkat Rawan Kecelakaan Lalu Lintas Menggunakan Metode Kernel Density dan K-Medoids (Studi Kasus: Kecamatan Depok dan Kalasan, Kabupaten Sleman). *Jurnal Ilmiah Geomatika*, 03(1), 23–35. <http://jurnal.upnyk.ac.id/index.php/imagi>
- Tangkelayuk, A. (2022). The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes, dan Decision Tree. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(2), 1109–1119.
- Widiastiwi, Y., & Ernawati, I. (2021). Klasifikasi Penyakit Batu Ginjal Menggunakan Algoritma Decision Tree C4 . 5 Dengan Membandingkan Hasil Uji Akurasi. *Jurnal IKRA-ITH INFORMATIKA*, 5(2), 128.