

Klasifikasi Parameter Penyebab Kecelakaan Lalu Lintas Dengan Algoritma Decission Tree Di Wilayah Polresta Kota Bengkulu

Ronna Mutiara Maheswari Wahyudi¹⁾, Herlina Latipa Sari²⁾, Desi Mahdalena³⁾

¹ Mahasiswa, Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dehasen Bengkulu
Alamat : Desa Saka Jaya, Blok B, Kecamatan Muara Enim, Kabupaten Muara Enim, Sumatera Selatan
email : ronnamutiara@gmail.com

^{2,3} Dosen Tetap, Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dehasen Bengkulu
Jl. Meranti Raya No. 32 Kota Bengkulu 38228 Telp. (0736) 22027, 26957 Fax. (0736) 341139;
e-mail: herlinalatipasari@unived.ac.id desimahdalena08@unived.ac.id

(Received: Juni 2026, Revised: Agustus 2026, Accepied: Oktober 2026)

Abstract-The Traffic accidents are a complex issue influenced by multiple factors, including human behavior, vehicle conditions, road characteristics, and environmental conditions. Historically, traffic accident data maintained by the Police Department of Bengkulu City have been utilized primarily for descriptive reporting, limiting their potential to reveal comprehensive patterns of accident causation. This study aims to identify the dominant factors contributing to traffic accidents, implement the Decision Tree C4.5 algorithm for accident classification, and evaluate the performance of the resulting classification model. This study used the Knowledge Discovery in Databases (KDD) methodology. The dataset consisted of historical traffic accident records obtained from the Jurisdiction of Bengkulu City Police. Data processing was conducted using Python in the Google Collaboratory environment through several stages, including exploratory data analysis (EDA), data preprocessing, model development, and performance evaluation. The results demonstrate that the proposed model successfully identified relationships among accident causative factors through the construction of a decision tree, generating 20 classification rules. Road Surface Condition was identified as the most influential predictor, with a feature importance value of 0.30, followed by Special Information (0.15), Speed Limit at the Accident Location (0.13), and Lighting Condition (0.10). Model evaluation showed an overall accuracy of 65.2%, with a weighted average F1-score of 75.9%, indicating satisfactory performance in classifying the majority classes. However, the macro average F1-score of 28.9% suggests that the model's performance on minority classes remains limited due to class imbalance in the dataset. This study is expected to provide a data-driven foundation for developing more effective traffic accident prevention and mitigation strategies in the Jurisdiction of Bengkulu City Police.

Keywords: Traffic Accidents, Data Mining, Decision Tree C4.5, Classification, the Jurisdiction of Bengkulu City Police.

Intisari-Kecelakaan lalu lintas merupakan permasalahan yang kompleks karena dipengaruhi oleh berbagai faktor, seperti faktor manusia, kendaraan, kondisi jalan, dan lingkungan. Data kecelakaan yang tersimpan di Polresta Kota Bengkulu selama ini masih dimanfaatkan secara deskriptif sehingga belum mampu mengungkap pola penyebab kecelakaan secara komprehensif. Penelitian ini bertujuan untuk mengidentifikasi parameter dominan penyebab kecelakaan lalu lintas, mengimplementasikan algoritma Decision Tree C4.5 dalam proses klasifikasi, serta mengevaluasi performa model yang dihasilkan. Metode penelitian yang digunakan adalah Knowledge Discovery in Database (KDD) yang. Dataset penelitian berupa data historis kecelakaan lalu lintas yang diperoleh dari Polresta Kota Bengkulu. Proses pengolahan data dilakukan menggunakan

Python pada lingkungan Google Colaboratory melalui tahapan exploratory data analysis,

Hasil penelitian menunjukkan bahwa model mampu mengidentifikasi pola hubungan antarparameter penyebab kecelakaan melalui pembentukan pohon keputusan dan menghasilkan 20 aturan. Variabel Kondisi Permukaan Jalan menjadi faktor yang paling dominan dengan nilai feature importance sebesar 0,30, kemudian Informasi Khusus sebesar 0,15, Batas Kecepatan di Lokasi sebesar 0,13, dan Kondisi Cahaya sebesar 0,10. Berdasarkan hasil evaluasi, model memperoleh tingkat akurasi keseluruhan sebesar 65,2%, dengan nilai weighted average F1-score sebesar 75,9% yang menunjukkan kemampuan model cukup baik dalam mengklasifikasikan kelas mayoritas. Sementara itu, nilai macro average F1-score sebesar 28,9% mengindikasikan bahwa performa model pada kelas minoritas masih belum optimal akibat ketidakseimbangan distribusi data. Penelitian ini diharapkan dapat menjadi dasar dalam penyusunan strategi pencegahan dan mitigasi kecelakaan lalu lintas yang lebih tepat sasaran dan berbasis data.

Kata kunci : kecelakaan lalu lintas, data mining, Decision Tree C4.5, klasifikasi, Polresta Kota Bengkulu

I. PENDAHULUAN

Kemajuan teknologi dan informasi yang mengalami perubahan begitu pesat, tidak hanya pada sektor industri dan komersial akan tetapi juga pada sektor pemerintahan, layanan publik dan sistem transportasi. Perkembangan teknologi dan informasi sangat membantu dalam pengolahan data berskala besar, otomatisasi keputusan, serta peningkatan efisiensi dengan memanfaatkan metode analitik nda Artificial Intelligence (AI). Penggunaan AI seperti data mining dan Machine Learning (ML) telah digunakan untuk menggali informasi, menemukan pola, memprediksi kejadian hingga sebagai bahas pertimbangan pengambilan berbasis data. Polresta Kota Bengkulu memiliki peran vital dalam mengumpulkan dan mengelola data kecelakaan lalu lintas. Ketersediaan data ini tidak hanya mencatat kuantitas kejadian, tetapi juga menyimpan variabel kompleks yang menjadi indikator utama risiko di jalan raya. Parameter-parameter tersebut mencakup faktor manusia (human error) seperti tingkat kelalaian, faktor kendaraan terkait kelaikan teknis, faktor infrastruktur yang melibatkan kondisi jalan (rata, berlubang, atau tikungan tajam), hingga faktor eksternal seperti kondisi cuaca dan intensitas cahaya (siang/malam). Dalam upaya menurunkan tingkat kecelakaan dan fatalitas korban, Polresta Kota Bengkulu membutuhkan pendekatan analitik

yang mampu mengklasifikasikan parameter tersebut secara akurat guna menghasilkan informasi strategis bagi pengambilan kebijakan. Menurut laporan statistik dari Satuan Lalu Lintas Polres Kota Bengkulu terdapat total 808 kejadian kecelakaan lalu lintas yang tercatat di wilayah Kota Bengkulu dalam rentang waktu 2020-2024 (Rozi et al., 2025). Selain itu data terkini dari media lokal menunjukkan bahwa periode Januari-September 2025, Satlantas Polresta Bengkulu mencatat 289 kasus kecelakaan lalu lintas (Tourindo & Lubis, 2025). Dengan rekam jejak kecelakaan lalu lintas yang tersedia, Polresta Kota Bengkulu menghadapi kendala dalam mengidentifikasi pola penyebab kecelakaan yang kompleks dan multi-faktor. Secara aktual terdapat celah antara meningkatnya jumlah kejadian dengan belum optimalnya mengidentifikasi serta mengklasifikasi parameter penyebab kecelakaan lalu lintas. Hal ini karena data yang tersedia selama ini hanya dianalisis secara deskriptif konvensional tanpa menggunakan teknik klasifikasi.

II. LANDASAN TEORI

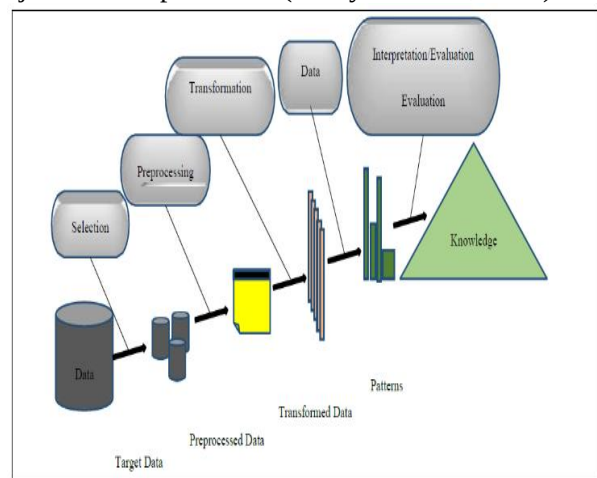
elitian terdahulu merupakan indikator penting dalam penyusunan tugas akhir karena akan memperkuar arah dan kontribusi penelitian yang akan dilakukan. Melalui kajian terhadap penelitian terdahulu, penulis dapat melakukan pendekatan terhadap tema dan tujuan penelitian.

Tabel 1. Penelitian Terdahulu

No	Author & Tahun	Metode yang digunakan	Objek Penelitian	Hasil Penelitian	Kelebihan & Kekurangan
1	(Saragih et al., 2024)	Decision Tree	Data kecelakaan lalu lintas di Indonesia	Algoritma Decision Tree mampu mengidentifikasi pola kecelakaan dan faktor dominan penyebab kecelakaan lalu lintas secara jelas dan terstruktur	Kelebihan: Interpretasi hasil mudah dipahami. Kekurangan: Tidak membahas perbandingan dengan algoritma lain dan belum fokus pada wilayah spesifik
2	(Alfiansyah et al., 2025)	Decision Tree C4.5 dan teknik SMOTE	Data kecelakaan lalu lintas di Indonesia	Implementasi SMOTE meningkatkan akurasi klasifikasi dan mengatasi ketidakseimbangan data kecelakaan	Kelebihan: Akurasi model meningkat signifikan. Kekurangan: Kompleksitas preprocessing tinggi dan hasil kurang mudah diinterpretasikan bagi non-teknis
3	(Atsir et al., 2025)	Random Forest	Data kecelakaan lalu lintas	Random Forest menghasilkan akurasi lebih tinggi dibandingkan Decision Tree tunggal	Kelebihan: Akurasi tinggi dan stabil. Kekurangan: Interpretabilitas rendah sehingga kurang cocok untuk pengambilan keputusan praktis
4	(Abdullah & Sipos, 2022)	Decision Tree	Analisis perilaku pengemudi	Faktor perilaku pengemudi menjadi penyebab utama kecelakaan lalu lintas berdasarkan hasil klasifikasi	Kelebihan: Fokus pada faktor manusia secara mendalam. Kekurangan: Tidak mempertimbangkan faktor jalan dan lingkungan secara komprehensif

Berdasarkan studi penelitian terdahulu, dapat disimpulkan bahwasanya algoritma *decision tree* telah banyak dieksplorasi dalam analisis kecelakaan lalu lintas dan terbukti mampu mengklasifikasikan faktor penyebab kecelakaan secara efektif. Namun sebagian penelitian masih bersifat global, tidak berfokus pada wilayah spesifik, serta belum mengoptimalkan interpretasi hasil untuk mendukung keputusan

kebijakan di tingkat kota. Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis (Sartika & Jumadi, 2023). Selanjutnya menurut Rahmat et al., (2021) mengungkapkan bahwa data mining merupakan pendekatan sistematis yang memanfaatkan algoritma komputasi untuk mengolah data berskala besar sehingga menghasilkan informasi strategis yang mendukung perencanaan dan evaluasi kebijakan. Data mining merupakan salah satu dari rangkaian *Knowledge Discovery in Database* (KDD). KDD berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah kumpulan data (Naldy & Andri, 2021)



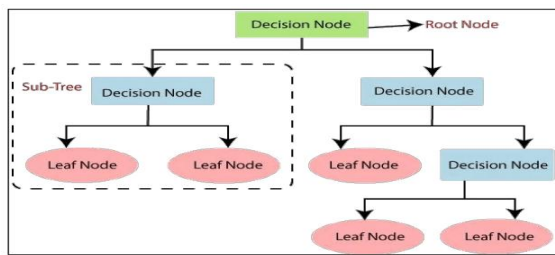
Gambar 2. Tahapan Data Mining

Klasifikasi

Klasifikasi adalah pengelompokan dari suatu rekaman (set pelatihan), di mana setiap rekaman berisi kumpulan atribut. Tujuan klasifikasi adalah untuk menemukan model dari atribut kelas, sehingga kemudian dapat memberikan nilai terhadap atribut lainnya (Ramadhon et al., 2024)

Decision Tree

Decision Tree adalah sebuah struktur pohon, dimana setiap node pohon merepresentasikan atribut yang telah diuji, setiap cabang merupakan suatu pembagian hasil uji, dan node daun (*leaf*) merepresentasikan kelompok kelas tertentu. Level node teratas dari sebuah *Decision Tree* adalah node akar (*root*) yang biasanya berupa atribut yang paling memiliki pengaruh terbesar pada suatu kelas tertentu. Pada umumnya *Decision Tree* melakukan strategi pencarian secara *top-down* untuk solusinya. Pada proses mengklasifikasi data yang tidak diketahui, nilai atribut akan diuji dengan cara melacak jalur dari node akar (*root*) sampai node akhir (daun) dan kemudian akan diprediksi kelas yang dimiliki oleh suatu data baru tertentu (Ananda et al., 2022).



Gambar 3. Konsep Decision Tree

Tinjauan Bahasa Pemrograman

python merupakan bahasa pemrograman tingkat tinggi yang yang dibuat oleh Guido Van Rossum dan dirilis pada tahun 1991. Python juga merupakan bahasa yang sangat populer belakangan ini. Selain itu python juga merupakan bahasa pemrograman yang multi fungsi contohnya python dapat digunakan untuk *Machine Learning* dan *Deep Learning*. Python dipilih sebagai penelitian karena python memiliki penulisan sintaksis yang mudah selain itu python juga memiliki library yang lengkap dan memiliki dukungan komunitas yang kuat karena python bersifat *open source*. Untuk menuliskan *source code python* anda dapat menggunakan IDE seperti *Visual Studio Code*, *Sublime Text*, *PyCharm* atau anda juga dapat menggunakan *IDE online* seperti *Jupyter notebook* dan *google colab* (Alfarizi et al., 2023).

Google Colab adalah layanan yang disediakan oleh Google dan bersifat gratis. Untuk membantu memanfaatkan *Python* dan *Pandas* dengan mudah, diperlukan juga *integrated development environment* yang paling mudah diperoleh. Keunggulan yang dimiliki *Google Colab* adalah pengguna tidak perlu menginstall *Python* dan *Pandas* kedalam komputer, serta tidak membutuhkan editor yang terinstall. Karena yang dibutuhkan hanyalah akun Google aktif (Hutagalung et al., 2023).

Flowchart yang sering disebut dengan diagram alir adalah suatu jenis diagram yang merepresentasikan algoritma atau langkah-langkah instruksi yang berurutan dalam sistem. flowchart dibuat dengan tujuan agar melihatkan setiap proses yang harus dilalui dalam suatu sistem (Febryana et al., 2024).

III. METODOLOGI PENELITIAN

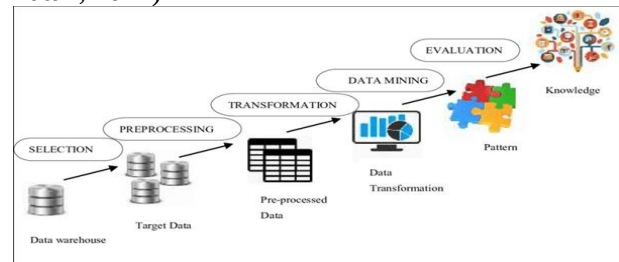
Subjek penelitian ini bukan data individu melainkan data histori kecelakaan lalu lintas yang tersedia di Polres Kota Bengkulu dalam beberapa tahun terakhir (2-3 tahun). Subjek penelitian bukan data individu melainkan data histori Data yang dianalisis meliputi:

1. Faktor manusia
2. Faktor kendaraan
3. Faktor jalan
4. Faktor lingkungan

Waktu penelitian dimulai dari bulan November 2025 sampai dengan Mei 2026. Penelitian dilakukan di Polres Kota Bengkulu yang beralamat di Jl. A. Yani, Malabero, Bengkulu, Kota Bengkulu, Bengkulu 38116.

Pada penelitian ini penulis, jenis penelitian yang digunakan dengan melakukan pendekatan eksperimen berbasis data mining. Penelitian bersifat *applied research*, karena hasilnya ditujukan untuk mendukung pengambilan keputusan praktis dalam bidang lalu lintas

Metode penelitian yang akan digunakan adalah *Knowledge Discovery In Database (KDD)*. KDD adalah metode teknis yang berguna untuk mencari dan mengidentifikasi pola (pattern) dalam data, pola yang sudah ditemukan bersifat sah dan baru sehingga dapat bermanfaat dan dapat dimengerti (Alfandi & Fatah, 2024)



Gambar 4. Tahapan KDD.

Adapun metode pengumpulan data yang penulis gunakan pada penelitian ini adalah sebagai berikut:

1. Observasi
Dalam pengumpulan data melalui observasi, penulis memperoleh gambaran terkait karakteristik data yang digunakan untuk proses analisis klasifikasi kecelakaan lalu lintas
2. Studi Dokumentasi
Studi dokumentasi ini dilakukan untuk memperoleh data histori kecelakaan lalu lintas di Polres Kota Bengkulu
3. Studi Pustaka
Metode dimana penulis mempelajari dan mencari data yang berasal dari buku dan referensi yang berhubungan dengan masalah yang ditulis.

Analisis Sistem Aktual

Berdasarkan dari observasi yang dilakukan dapat disimpulkan bahwasanya pengolahan data kecelakaan lalu lintas yang ada masih digunakan secara deskriptif atau tidak mencari pola sebab akibat serta belum dimanfaatkan untuk analisis untuk menghasilkan pengetahuan baru yang mudah dipahami

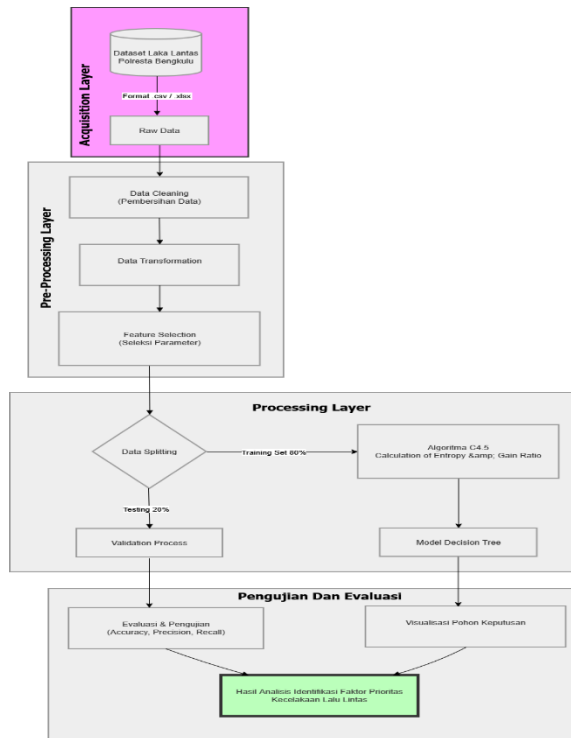
Analisis Kebutuhan Sistem

Sistem pada penelitian ini dirancang sebagai sistem klasifikasi berbasis data mining yang diimplementasikan menggunakan RapidMiner sebagai perangkat lunak pendukung model Decision Tree C4.5.

Perancangan Arsitektur Sistem Analisis Parameter Kecelakaan lalu Lintas

Arsitektur sistem merupakan kerangka kerja logis yang menggambarkan aliran data dan tahapan pemrosesan informasi dalam penelitian ini, mulai dari

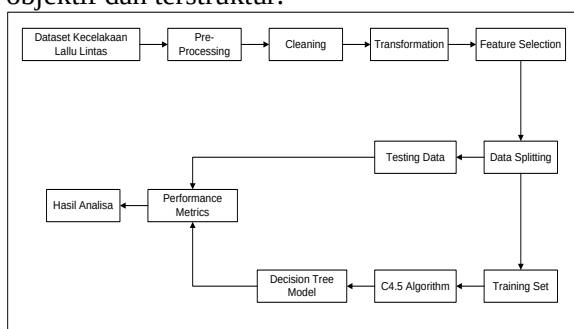
tahap akuisisi data mentah hingga dihasilkan sebuah pengetahuan (*knowledge*) yang dapat digunakan untuk pengambilan keputusan. Sistem klasifikasi parameter kecelakaan lalu lintas ini dirancang dengan mengadopsi standar industri proses data mining, yang mengintegrasikan pengolahan data lapangan dari Polresta Kota Bengkulu dengan kecerdasan algoritma Decision Tree C4.5.



Gambar 5. Rancangan Arsitektur Analisis Kecelakaan Lalu Lintas

Perancangan Analisis Data (Pipeline Data)

Perancangan analisis data (*Pipeline data*) ini akan menjelaskan langkah-langkah pengolahan data kecelakaan lalu lintas. Proses ini disusun secara sistematis agar analisis yang akan dilakukan bersifat objektif dan terstruktur.



Gambar 6. Rancangan Pipeline Data Analisis Kecelakaan Lalu Lintas

Pada tahap evaluasi, dilakukan penilaian terhadap kinerja model yang telah dibangun untuk mengukur seberapa baik model tersebut dalam klasifikasi parameter kecelakaan lalu lintas. Langkah-langkah evaluasi mencakup pengukuran kinerja menggunakan metrik seperti akurasi, presisi, recall,

F1-score, serta analisis confusion matrix untuk memahami jenis kesalahan yang dilakukan oleh model (Falah et al., 2025).

IV. HASIL DAN PEMBAHASAN

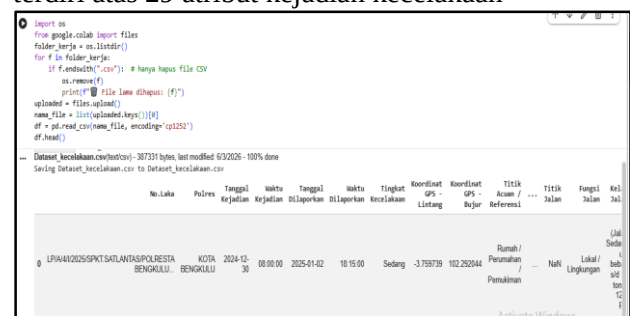
Implementasi sistem dalam penelitian ini difokuskan pada pemrosesan dan ekstraksi pengetahuan (*knowledge discovery*) dari data historis kecelakaan lalu lintas di wilayah hukum Kepolisian Resor Kota (Polresta) Bengkulu. Konseptualisasi sistem ini dirancang untuk menjembatani tumpukan data administratif mentah yang bersifat statis menjadi sebuah model prediktif dinamis berbasis algoritma Decision Tree (C4.5). Melalui integrasi teknologi data science, data laporan kecelakaan tidak lagi sekadar menjadi arsip pelaporan, melainkan ditransformasikan menjadi instrumen evaluasi strategis untuk mengidentifikasi pola kausalitas kecelakaan secara komprehensif. Pendekatan analitis ini menjadi krusial mengingat kompleksitas variabel lapangan—seperti faktor temporal, kondisi lingkungan, dan detail spesifik insiden—memerlukan penanganan komputasional yang terstruktur guna menghasilkan rekomendasi kebijakan keselamatan jalan (road safety) yang objektif dan berbasis data (*data-driven*).

Implementasi keseluruhan rancangan komputasi ini dieksekusi dalam lingkungan *Google Colab* menggunakan ekosistem bahasa pemrograman *Python* beserta pustaka spesifik seperti *Pandas*, *Scikit-Learn*, dan *Holidays Indonesia*. Pengembangan sistem dilakukan melalui pendekatan *pipeline end-to-end*, di mana setiap tahapan saling terikat mulai dari pengondisian awal data (*data preprocessing*), rekayasa fitur temporal (*feature engineering*), penyandian variabel kategorikal (*label encoding*), hingga pemodelan dan pengujian klasifikasi.

Tahapan ini bertujuan untuk memproses data mentah laporan kecelakaan lalu lintas menjadi model prediktif yang valid. Berikut adalah penjabaran langkah-langkah implementasi sistem yang dilakukan:

Import Dataset

Sistem diimplementasikan menggunakan *library Pandas* untuk memuat *Dataset_Kecelakaan.csv*. Data mentah yang bersumber dari laporan kepolisian ini terdiri atas 29 atribut kejadian kecelakaan



Gambar 7. Import Dataset Kecelakaan

Exploratory Data Analysis

Exploratory Data Analysis (EDA) adalah pendekatan investigasi awal pada data untuk memahami karakteristik data seperti missing value, duplikasi data, kesalahan format data serta menemukan korelasi antar variabel. Pada studi kasus penelitian ini terdapat missing value pada variabel titik acuan, jarak ke lokasi kejadian, informasi khusus, nomor jalan, dan titik jalan. Berikut hasil analisis menggunakan google colab

```
print("=== TAHAP 2: ANALISIS KARAKTERISTIK FITUR ===")
print("\n[A] Ringkasan Tipe Data & Missing Values:")
print(df.info())

print("\n[B] Kolom yang Memiliki Data Kosong (Missing Values):")
missing_info = df.isnull().sum()
print(missing_info[missing_info > 0])

print("\n[C] Distribusi Kelas pada Variabel Target (Penyebab):")
print(df['Penyebab'].value_counts())
print("\n[D] Persentase Distribusi Kelas Target (%):")
print(df['Penyebab'].value_counts(normalize=True) * 100)
print("\n")
```

Gambar 8. Source Code Proses EDA

```
[B] Kolom yang Memiliki Data Kosong (Missing Values):
Titik Acuan / Referensi      13
Jarak Ke Lokasi Kecelakaan  571
Informasi Khusus            153
Nomor Jalan                 669
Titik Jalan                 654
dtype: int64

[C] Distribusi Kelas pada Variabel Target (Penyebab):
Penyebab
Manusia      660
Jalan        22
Kendaraan     6
Alam (kabut,banjir,dll)  2
Name: count, dtype: int64

[D] Persentase Distribusi Kelas Target (%):
Penyebab
Manusia      95.652174
Jalan        3.188406
Kendaraan    0.869565
Alam (kabut,banjir,dll)  0.289855
Name: proportion, dtype: float64
```

Gambar 9. Hasil Analisis EDA

Rekayasa Fitur (*Feature Engineering*)

Pada tahap ini, data mentah waktu kejadian tidak dimasukkan secara langsung ke dalam model untuk menghindari high cardinality yang dapat menyebabkan overfitting pada algoritma C4.5. Sebagai gantinya, dilakukan rekayasa fitur (*Feature Engineering*) dengan mentransformasikan data jam mentah menjadi variabel kategori baru (Kategori_Waktu), yang membagi waktu kejadian menjadi tiga fase: Pagi (06:00–11:59), Siang (12:00–17:59), dan Malam (18:00–05:59).

```
df['Waktu Kejadian'] = df['Waktu Kejadian'].astype(str)
df['Jam'] = df['Waktu Kejadian'].str.split(':').str[0].astype(int)

def kelompokkan_waktu(jam):
    if 6 <= jam < 12:
        return 'Pagi'
    elif 12 <= jam < 18:
        return 'Siang'
    else:
        return 'Malam'

df['Kategori_Waktu'] = df['Jam'].apply(kelompokkan_waktu)

print("=== DISTRIBUSI KASUS BERDASARKAN KATEGORI WAKTU ===")
print(df['Kategori_Waktu'].value_counts())

print("\n=== PERSENTASE DISTRIBUSI KATEGORI WAKTU (%) ===")
print(df['Kategori_Waktu'].value_counts(normalize=True) * 100)
```

Gambar 10. Source Code Feature Engineering Waktu Kejadian

```
=== DISTRIBUSI KASUS BERDASARKAN KATEGORI WAKTU ===
Kategori_Waktu
Malam      269
Siang      221
Pagi       200
Name: count, dtype: int64

=== PERSENTASE DISTRIBUSI KATEGORI WAKTU (%) ===
Kategori_Waktu
Malam      38.985507
Siang      32.028986
Pagi       28.985507
```

Gambar 11. Hasil Analisa Waktu Kejadian

Data Pre Processing

Pada tahap ini melakukan analisa *missing values* pada atribut Informasi Khusus (seperti rekam jejak tabrak lari) dengan metode imputasi konstanta berupa label 'Tidak Ada Catatan Khusus'. Hal ini dilakukan untuk mempertahankan keutuhan sampel data (menghindari pembuangan baris data). Selanjutnya, kolom-kolom administratif dan fitur mentah seperti No.Laka, Nama Jalan, dan koordinat GPS dieliminasi karena tidak memiliki relevansi kausalitas dalam pembentukan pohon keputusan.

```
print("=== TAHAP 2: ANALISIS KARAKTERISTIK FITUR ===")
print("\n[A] Ringkasan Tipe Data & Missing Values:")
print(df.info())

print("\n[B] Kolom yang Memiliki Data Kosong (Missing Values):")
missing_info = df.isnull().sum()
print(missing_info[missing_info > 0])
```

Gambar 12. Source Code Cek Missing

```
[B] Kolom yang Memiliki Data Kosong (Missing Values):
Titik Acuan / Referensi      13
Jarak Ke Lokasi Kecelakaan  571
Informasi Khusus            153
Nomor Jalan                 669
Titik Jalan                 654
dtype: int64
```

Gambar 13. Variabel Data Kosong

Dari gambar 13. terlihat variabel atau kriteria yang terdapat data kosong (missing value) ada pada variabel titik acuan, jarak lokasi kecelakaan, informasi khusus, nomor jalan, dan titik jalan.

Transformasi Fitur dan Encoding

Algoritma *decision tree* C4.5 beroperasi menggunakan representasi matematis. Oleh karena itu, seluruh atribut bertipe teks, termasuk variabel target Penyebab (seperti faktor Manusia, Alam, Kendaraan), ditransformasikan menjadi representasi numerik menggunakan fungsi LabelEncoder dari Scikit-Learn

```
from sklearn.preprocessing import LabelEncoder
print("=== TAHAP 5 & 6: TRANSFORMASI FITUR & ENCODING ===")
target_col = 'Penyebab'
X = df_clean.drop(columns=[target_col])
y = df_clean[target_col]

encoder_dict = {}

# Fitur Teks (Categorical Features) diubah menjadi Angka
for col in X.columns:
    if X[col].dtype == 'object':
        le = LabelEncoder()
        X[col] = le.fit_transform(X[col].astype(str))
        encoder_dict[col] = le

# Target Teks (Variabel Y) diubah menjadi Angka
le_target = LabelEncoder()
y = le_target.fit_transform(y.astype(str))
encoder_dict[target_col] = le_target
print("✓ Proses encoding seluruh kolom teks ke numerik berhasil dilakukan.\n")
```

Gambar 14. Source Code Transformasi Fitur dan Encoding

Untuk memvalidasi hasil dari dari proses Transformasi Fitur dan Encoding menggunakan codingan sebagai berikut :

```

baris_sampel = [0, 1, 2, 3, 4]

print("=====")
print(" 1. PERBANDINGAN FITUR / PARAMETER (X)")
print("=====")
print("\n [SEBELUM ENCODING - Teks Asli pada df_clean]")
# Menampilkan fitur hasil rekayasa waktu dan hari yang masih berbentuk teks
display(df_clean[['Kategori_Waktu', 'Tipe_Hari']].iloc[baris_sampel])

print("\n [SESUDAH ENCODING - Berubah Menjadi Angka pada X]")
# Menampilkan fitur yang sama namun sudah diproses oleh algoritma C4.5
display(X[['Kategori_Waktu', 'Tipe_Hari']].iloc[baris_sampel])

print("\n=====")
print(" 2. PERBANDINGAN VARIABEL TARGET / PENYEBAB KECELAKAAN (Y)")
print("=====")
print("\n [SEBELUM ENCODING - Kelas Target Asli]")
display(df_clean[['Penyebab']].iloc[baris_sampel])

print("\n [SESUDAH ENCODING - Kelas Target Menjadi Angka]")
# Karena 'y' berbentuk numpy array, kita bungkus ke DataFrame agar tampilannya rapi
df_y_sesudah = pd.DataFrame(y, columns=['Penyebab (Encoded)']).iloc[baris_sampel]
display(df_y_sesudah)

```

Gambar 15. Source Code Validasi Proses Transformasi Encoding

Training dan Testing

Dataset dibagi menjadi dua proporsi, yaitu 80% sebagai data latih (*training data*) dan 20% sebagai data uji (*testing data*). Pembagian ini menggunakan parameter stratify yang memastikan rasio kelas target pada data uji sebanding dengan data latih, sehingga pengujian berlangsung secara representatif dan seimbang.

```

print("==== TAHAP 7: TRAINING DAN TESTING SPLIT ===")
# Menggunakan stratify=y agar rasio target di data training & testing tetap seimbang
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42, stratify=y
)
print("✓ Pembagian Data Berhasil:")
print(f" - Jumlah Data Training (80%) : {X_train.shape[0]} baris")
print(f" - Jumlah Data Testing (20%) : {X_test.shape[0]} baris\n")

==== TAHAP 7: TRAINING DAN TESTING SPLIT ===
✓ Pembagian Data Berhasil:
- Jumlah Data Training (80%) : 552 baris
- Jumlah Data Testing (20%) : 138 baris

```

Gambar 16. Source Code Training dan Testing

Modelling Decision Tree C4.5

Pada permodelan merupakan bagian utama dari proses pencarian dimana data kecelakaan lalu lintas yang bersumber dari dataaset Polres Kota Bengkulu yang telah melalui proses pra processing diekstrak menjadi sebuah keputusan yang berbasis data. Adapun source codenya adalah sebagai berikut :

```

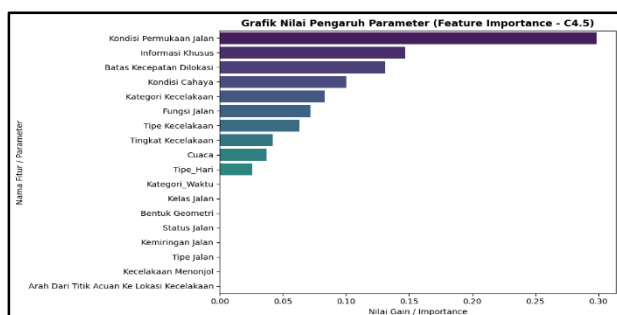
print("==== TAHAP 8: MODELING C4.5 & FEATURE IMPORTANCE ===")
# Menerapkan Regularisasi / Pruning (max_depth=5) agar model tidak Overfitting
model_c45 = DecisionTreeClassifier(
    criterion='entropy',
    max_depth=5,
    min_samples_split=10,
    min_samples_leaf=5,
    random_state=42
)
model_c45.fit(X_train, y_train)
print("✓ Model C4.5 berhasil dilatih.")

# Visualisasi Feature Importance (Days pembeda Information Gain)
importances = model_c45.feature_importances_
indices = np.argsort(importances)[::-1]

plt.figure(figsize=(10, 6))
sns.barplot(x=importances[indices], y=X.columns[indices], palette='viridis')
plt.title("Grafik Nilai Pengaruh Parameter (Feature Importance - C4.5)", fontsize=12, fontweight='bold')
plt.xlabel("Nilai Gain / Importance")
plt.ylabel("Nama Fitur / Parameter")
plt.tight_layout()
plt.show()

```

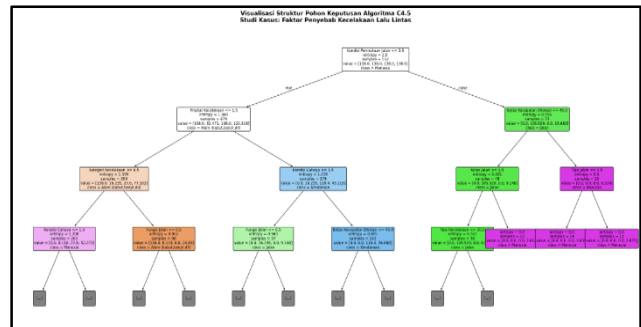
Gambar 17. Source Coding Modelling Decision Tree



Gambar 18. Grafik Hasil Model Decision Tree C4.5

Berdasarkan hasil analisis *feature importance* pada model Decision Tree C4.5, variabel *Kondisi*

Permukaan Jalan merupakan faktor yang memiliki pengaruh terbesar terhadap proses klasifikasi dengan nilai importance sebesar 0,30. Selanjutnya, variabel *Informasi Khusus* (0,15), *Batas Kecepatan di Lokasi* (0,13), dan *Kondisi Cahaya* (0,10) juga menunjukkan kontribusi yang signifikan.



Gambar 19. Pohon Keputusan Parameter Kecelakaan Lalu Lintas

Evaluasi

Tahapan pengujian dan evaluasi dilakukan untuk mengukur performa, reliabilitas, dan kemampuan generalisasi model *Decision Tree* (C4.5) terhadap data yang belum pernah di untuk memvalidasi apakah model yang telah dilatih mampu bekerja proses sebelumnya (*unseen data*).

```

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
print("\n==== TAHAP 9: EVALUASI MODEL ===")
y_pred = model_c45.predict(X_test)

print("✓ Akurasi Model Pasca-Pruning: {accuracy_score(y_test, y_pred) * 100:.2f}%\n")
print("Classification Report:")

# PERBAIKAN DI BARIS INI: Tambahkan parameter labels=range(len(le_target.classes_))
print(classification_report(
    y_test,
    y_pred,
    labels=range(len(le_target.classes_)), # <--- Ini kunci perbaikannya
    target_names=le_target.classes_,
    zero_division=0
))

```

Gambar 20. Source Code Evaluasi

```

==== TAHAP 9: EVALUASI MODEL ===
✓ Akurasi Model Pasca-Pruning: 65.22%

Classification Report:

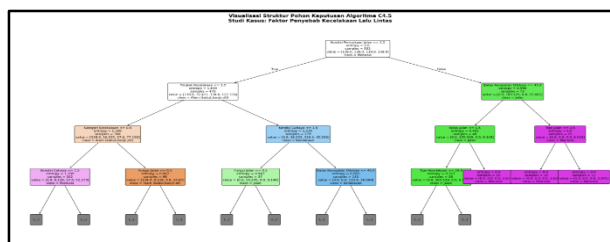
```

	precision	recall	f1-score	support
Alam (kabut,banjir,dll)	0.00	0.00	0.00	0
Jalan	0.20	0.60	0.30	5
Kendaraan	0.04	1.00	0.07	1
Manusia	0.98	0.65	0.78	132
accuracy			0.65	138
macro avg	0.30	0.56	0.29	138
weighted avg	0.94	0.65	0.76	138

Gambar 21. Hasil Evaluasi Decision Tree

Analisis Struktur Pohon Keputusan Algoritma C4.5

Pohon keputusan (*decision tree*) merupakan representasi visual dari proses pengambilan keputusan yang dihasilkan oleh algoritma klasifikasi. Dalam penelitian ini, algoritma **C4.5** digunakan untuk mengidentifikasi faktor-faktor yang memengaruhi penyebab kecelakaan lalu lintas berdasarkan sejumlah atribut yang terkait dengan kondisi jalan, lingkungan, karakteristik kecelakaan, serta pengaturan lalu lintas.



Gambar 22. Struktur Pohon Keputusan Algoritma C4.5

Visualisasi pohon keputusan, setiap node merepresentasikan atribut yang digunakan sebagai dasar pengambilan keputusan. Cabang-cabang yang terbentuk menunjukkan hasil pengujian terhadap nilai atribut tertentu, sedangkan node terminal (*leaf node*) menunjukkan kelas hasil klasifikasi. Informasi yang ditampilkan pada setiap node meliputi nilai *entropy*, jumlah *samples*, distribusi data pada masing-masing kelas (*value*), dan *class* yang menunjukkan kelas mayoritas pada node tersebut.

Nilai *entropy* digunakan untuk mengukur tingkat ketidakpastian atau heterogenitas data pada suatu node. Semakin tinggi nilai *entropy*, semakin beragam distribusi data antar kelas sehingga proses klasifikasi masih memiliki tingkat ketidakpastian yang tinggi. Sebaliknya, semakin rendah nilai *entropy* menunjukkan bahwa data pada node tersebut semakin homogen dan proses klasifikasi semakin baik.

Berdasarkan visualisasi pohon keputusan yang dihasilkan, atribut Kondisi Permukaan Jalan terpilih sebagai node akar (*root node*), yang menunjukkan bahwa atribut tersebut memiliki kemampuan terbaik dalam memisahkan data dibandingkan atribut lainnya. Selanjutnya, model melakukan percabangan menggunakan atribut seperti Tingkat Kecelakaan, Batas Kecepatan di Lokasi, Kondisi Cahaya, Kategori Kecelakaan, Fungsi Jalan, Kelas Jalan, dan Tipe Jalan. Struktur pohon yang terbentuk menggambarkan hubungan hierarkis antar atribut dalam menentukan penyebab kecelakaan lalu lintas, sehingga dapat digunakan untuk memahami faktor-faktor dominan yang berkontribusi terhadap terjadinya kecelakaan.

Analisa Rule Algoritma Decision Tree C4.5

Kemampuan algoritma *Decision Tree* C4.5 tidak hanya untuk mengekstraksi pengetahuan yang tersembunyi dari suatu data akan tetapi juga mampu memberikan hasil prediksi akhir. Rule yang terbentuk dari dataset kecelakaan akan disajikan dalam bentuk aturan kaidah yang logis (IF..Then). Adapun hasil Rule yang terbentuk dari proses *decision tree* sebanyak 20 rule yaitu :

Aturan 1 : JIKA Kondisi Permukaan Jalan adalah Baik, Banjir/ Tergenang Air atau Basah DAN Tingkat Kecelakaan adalah Berat atau Ringan DAN Kategori Kecelakaan adalah Kontra DAN Kondisi Cahaya adalah Gelap / Sulit Terlihat atau Redup / Samar (Tidak jelas terlihat) DAN Batas Kecepatan Dilokasi ≤ 25 MAKA Faktor Penyebab = Kendaraan

Aturan 2 : JIKA Kondisi Permukaan Jalan adalah Baik, Banjir/ Tergenang Air atau Basah DAN Tingkat Kecelakaan adalah Berat atau Ringan DAN Kategori Kecelakaan adalah Kontra DAN Kondisi Cahaya adalah Gelap / Sulit Terlihat atau Redup / Samar (Tidak jelas terlihat) DAN Batas Kecepatan Dilokasi > 25 MAKA Faktor Penyebab = Manusia

Aturan 3 : JIKA Kondisi Permukaan Jalan adalah Baik, Banjir/ Tergenang Air atau Basah DAN Tingkat Kecelakaan adalah Berat atau Ringan DAN Kategori Kecelakaan adalah Kontra DAN Kondisi Cahaya adalah Terang / Jelas DAN Informasi Khusus adalah Tabrak Lari MAKA Faktor Penyebab = Jalan

Aturan 4 : JIKA Kondisi Permukaan Jalan adalah Baik, Banjir/ Tergenang Air atau Basah DAN Tingkat Kecelakaan adalah Berat atau Ringan DAN Kategori Kecelakaan adalah Kontra DAN Kondisi Cahaya adalah Terang / Jelas DAN Informasi Khusus adalah Tabrak Lari, Tidak Ada Saksi, Tabrakan Beruntun, Tidak Ada Catatan Khusus, Tidak Ada Saksi, Tidak Ada Saksi, Tabrak Lari atau Tidak Ada Saksi, Tabrak Lari, Tabrakan Beruntun MAKA Faktor Penyebab = Manusia

Aturan 5 : JIKA Kondisi Permukaan Jalan adalah Baik, Banjir/ Tergenang Air atau Basah DAN Tingkat Kecelakaan adalah Berat atau Ringan DAN Kategori Kecelakaan adalah Tunggal DAN Fungsi Jalan adalah Arteri DAN Tipe Jalan adalah 1/1 (1 Lajur/1 Arah), 2/1 (2 Lajur/1 Arah) atau 2/2 TB (2 Lajur/2 Arah Tanpa Batas Median) MAKA Faktor Penyebab = Alam (kabut,banjir,dll)

Aturan 6 : JIKA Kondisi Permukaan Jalan adalah Baik, Banjir/ Tergenang Air atau Basah DAN Tingkat Kecelakaan adalah Berat atau Ringan DAN Kategori Kecelakaan adalah Tunggal DAN Fungsi Jalan adalah Arteri DAN Tipe Jalan adalah 4/2 B (4 Lajur/2 Arah dengan Batas Median), 4/2 TB (4 Lajur/2 Arah Tanpa Batas Median) atau 6/2 B (6 Lajur/2 Arah dengan Batas Median) MAKA Faktor Penyebab = Manusia.

Aturan 7 : JIKA Kondisi Permukaan Jalan adalah Baik, Banjir/ Tergenang Air atau Basah DAN Tingkat Kecelakaan adalah Berat atau Ringan DAN Kategori Kecelakaan adalah Tunggal DAN Fungsi Jalan adalah Kolektor atau Lokal / Lingkungan DAN Informasi Khusus adalah Tabrak Lari, Tabrak Lari, Tidak Ada Saksi, Tabrakan Beruntun atau Tidak Ada Catatan Khusus MAKA Faktor Penyebab = Jalan

Aturan 8 : JIKA Kondisi Permukaan Jalan adalah Baik, Banjir/ Tergenang Air atau Basah DAN Tingkat Kecelakaan adalah Berat atau Ringan DAN Kategori Kecelakaan adalah Tunggal DAN Fungsi Jalan adalah Kolektor atau Lokal / Lingkungan DAN Informasi Khusus adalah Tidak Ada Saksi, Tidak Ada Saksi, Tabrak Lari atau Tidak Ada Saksi, Tabrak Lari, Tabrakan Beruntun MAKA Faktor Penyebab = Manusia. Aturan 9 : JIKA Kondisi Permukaan Jalan adalah Baik, Banjir/ Tergenang Air atau Basah DAN

Tingkat Kecelakaan adalah Sedang DAN Kondisi Cahaya adalah Gelap / Sulit Terlihat atau Redup / Samar (Tidak jelas terlihat) DAN Fungsi Jalan adalah Arteri DAN Kategori Kecelakaan adalah Kontra MAKFA Faktor Penyebab = Manusia Aturan 10 : JIKA Kondisi Permukaan Jalan adalah Baik, Banjir/ Tergenang Air atau Basah DAN Tingkat Kecelakaan adalah Sedang DAN Kondisi Cahaya adalah Gelap / Sulit Terlihat atau Redup / Samar (Tidak jelas terlihat) DAN Fungsi Jalan adalah Arteri DAN Kategori Kecelakaan adalah Tunggal MAKFA Faktor Penyebab = Jalan Aturan 11 : JIKA Kondisi Permukaan Jalan adalah Baik, Banjir/ Tergenang Air atau Basah DAN Tingkat Kecelakaan adalah Sedang DAN Kondisi Cahaya adalah Gelap / Sulit Terlihat atau Redup / Samar (Tidak jelas terlihat) DAN Fungsi Jalan adalah Kolektor atau Lokal / Lingkungan MAKFA Faktor Penyebab = Manusia. Aturan 12 : JIKA Kondisi Permukaan Jalan adalah Baik, Banjir/ Tergenang Air atau Basah DAN Tingkat Kecelakaan adalah Sedang DAN Kondisi Cahaya adalah Terang / Jelas DAN Batas Kecepatan Dilokasi ≤ 45 DAN Kategori Kecelakaan adalah Kontra MAKFA Faktor Penyebab = Kendaraan. Aturan 13 : JIKA Kondisi Permukaan Jalan adalah Baik, Banjir/ Tergenang Air atau Basah DAN Tingkat Kecelakaan adalah Sedang DAN Kondisi Cahaya adalah Terang / Jelas DAN Batas Kecepatan Dilokasi ≤ 45 DAN Kategori Kecelakaan adalah Tunggal MAKFA Faktor Penyebab = Kendaraan Aturan 14 : JIKA Kondisi Permukaan Jalan adalah Baik, Banjir/ Tergenang Air atau Basah DAN Tingkat Kecelakaan adalah Sedang DAN Kondisi Cahaya adalah Terang / Jelas DAN Batas Kecepatan Dilokasi > 45 MAKFA Faktor Penyebab = Manusia Aturan 15 : JIKA Kondisi Permukaan Jalan adalah Berdebu, Berlubang, Berombak, Keriting atau Licin DAN Batas Kecepatan Dilokasi ≤ 45 DAN Kelas Jalan adalah I (Jalan Besar utk beban 10 ton & max 18 m Panjang Ran) atau II (Jalan Sedang utk beban s/d 10 ton & 12m Panjang Ran) DAN Tipe Kecelakaan adalah Di ruas jalan, pejalan kaki berdiri ragu-ragu di tengah jalan, Di ruas jalan, pejalan kaki di bahu jalan atau di trotoar, Di ruas jalan, pejalan kaki menyeberang dari kanan ke kiri, Di ruas jalan, pejalan kaki menyeberang dari kiri ke kanan, Di ruas jalan, pejalan kaki sejajar jalan di kiri atau di kanan, Di ruas, kendaraan lurus dengan pejalan kaki dari kanan di Zebra Cross, Di simpang, pejalan kaki menyeberang dari kanan ke kiri, Di simpang, pejalan kaki menyeberang dari kiri ke kanan, Di simpang, tabrakan dengan Kendaraan B yang datang dari arah kanan, Di simpang, tabrakan dengan Kendaraan B yang datang dari arah kiri, Kendaraan Out of Control keluar ke kanan jalan, Kendaraan Out of Control keluar ke kiri jalan, Kendaraan sedang melintas tertimpa Tiang atau Pohon, Kendaraan tidak terkendali di simpang, Penumpang jatuh dari

kendaraan (input 1 penumpang), Tabrak belakang kendaraan depan yang belok kanan, Tabrak belakang kendaraan depan yang belok kiri, Tabrak kendaraan belok kanan dengan kendaraan yang jalan lurus, Tabrakan antara dua kendaraan belok, Tabrakan dengan Kendaraan menyebrang dari sisi kiri jalan atau Tabrakan dengan atau kecelakaan akibat binatang melintas jalan DAN Informasi Khusus adalah Tabrak Lari, Tabrak Lari, Tidak Ada Saksi, Tabrakan Beruntun atau Tidak Ada Catatan Khusus MAKFA Faktor Penyebab = Jalan Aturan 16 : JIKA Kondisi Permukaan Jalan adalah Berdebu, Berlubang, Berombak, Keriting atau Licin DAN Batas Kecepatan Dilokasi ≤ 45 DAN Kelas Jalan adalah I (Jalan Besar utk beban 10 ton & max 18 m Panjang Ran) atau II (Jalan Sedang utk beban s/d 10 ton & 12m Panjang Ran) DAN Tipe Kecelakaan adalah Di ruas jalan, pejalan kaki berdiri ragu-ragu di tengah jalan, Di ruas jalan, pejalan kaki di bahu jalan atau di trotoar, Di ruas jalan, pejalan kaki menyeberang dari kanan ke kiri, Di ruas jalan, pejalan kaki menyeberang dari kiri ke kanan, Di ruas jalan, pejalan kaki sejajar jalan di kiri atau di kanan, Di ruas, kendaraan lurus dengan pejalan kaki dari kanan di Zebra Cross, Di simpang, pejalan kaki menyeberang dari kanan ke kiri, Di simpang, pejalan kaki menyeberang dari kiri ke kanan, Di simpang, tabrakan dengan Kendaraan B yang datang dari arah kanan, Di simpang, tabrakan dengan Kendaraan B yang datang dari arah kiri, Kendaraan Out of Control keluar ke kanan jalan, Kendaraan Out of Control keluar ke kiri jalan, Kendaraan sedang melintas tertimpa Tiang atau Pohon, Kendaraan tidak terkendali di simpang, Penumpang jatuh dari kendaraan (input 1 penumpang), Tabrak belakang kendaraan depan yang belok kanan, Tabrak belakang kendaraan depan yang belok kiri, Tabrak kendaraan belok kanan dengan kendaraan yang jalan lurus, Tabrakan antara dua kendaraan belok, Tabrakan dengan Kendaraan menyebrang dari sisi kiri jalan atau Tabrakan dengan atau kecelakaan akibat binatang melintas jalan DAN Informasi Khusus adalah Tidak Ada Saksi, Tidak Ada Saksi, Tabrak Lari atau Tidak Ada Saksi, Tabrak Lari, Tabrakan Beruntun MAKFA Faktor Penyebab = Jalan Aturan 17 : JIKA Kondisi Permukaan Jalan adalah Berdebu, Berlubang, Berombak, Keriting atau Licin DAN Batas Kecepatan Dilokasi ≤ 45 DAN Kelas Jalan adalah I (Jalan Besar utk beban 10 ton & max 18 m Panjang Ran) atau II (Jalan Sedang utk beban s/d 10 ton & 12m Panjang Ran) DAN Tipe Kecelakaan adalah Tabrakan dengan benda di jalan atau di atas jalan, Tabrakan dengan kendaraan menyebrang dari sisi kanan jalan, Tabrakan dengan kendaraan parkir di kanan, Tabrakan dengan kendaraan parkir di kiri, Tabrakan depan - belakang, Tabrakan depan - depan, Tabrakan kendaraan belok

kanan dengan kendaraan dari arah kanan, Tabrakan kendaraan belok kanan dengan kendaraan dari arah kiri, Tabrakan kendaraan belok kanan dengan kendaraan yang datang dari arah berlawanan, Tabrakan kendaraan belok kiri dengan kendaraan dari arah kanan, Tabrakan kendaraan belok kiri dengan kendaraan dari arah kiri, Tabrakan kendaraan belok kiri dengan kendaraan yang datang dari arah yang berlawanan, Tabrakan kendaraan belok kiri dengan kendaraan yang jalan lurus, Tabrakan saat belok ke sisi kanan jalan, Tabrakan saat gerakan putar balik, Tabrakan saat kendaraan A melakukan gerakan mundur / lawan arah, Tabrakan saat menyalip, Tabrakan saat menyalip dari kanan, Tabrakan saat menyalip dari kiri, Tabrakan saat pindah lajur ke kanan, Tabrakan saat pindah lajur ke kiri atau Tabrakan samping MAKA Faktor Penyebab = Manusia

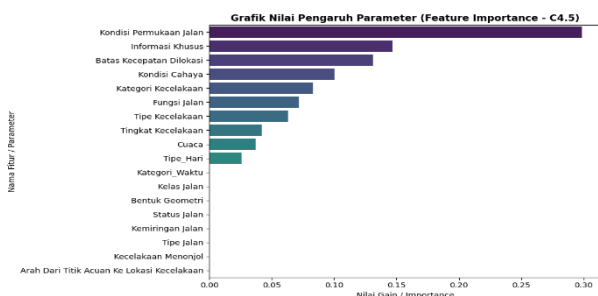
Aturan 18 : JIKA Kondisi Permukaan Jalan adalah Berdebu, Berlubang, Berombak, Keriting atau Licin DAN Batas Kecepatan Dilokasi ≤ 45 DAN Kelas Jalan adalah III (Jalan Kecil utk max beban 8 ton & 9 m Panjang Ran) MAKA Faktor Penyebab = Manusia

Aturan 19 : JIKA Kondisi Permukaan Jalan adalah Berdebu, Berlubang, Berombak, Keriting atau Licin DAN Batas Kecepatan Dilokasi > 45 DAN Tipe Jalan adalah 1/1 (1 Lajur/1 Arah), 2/1 (2 Lajur/1 Arah) atau 2/2 TB (2 Lajur/2 Arah Tanpa Batas Median) MAKA Faktor Penyebab = Manusia

Aturan 20 : JIKA Kondisi Permukaan Jalan adalah Berdebu, Berlubang, Berombak, Keriting atau Licin DAN Batas Kecepatan Dilokasi > 45 DAN Tipe Jalan adalah 4/2 B (4 Lajur/2 Arah dengan Batas Median), 4/2 TB (4 Lajur/2 Arah Tanpa Batas Median) atau 6/2 B (6 Lajur/2 Arah dengan Batas Median) MAKA Faktor Penyebab = Manusia

Analisa Grafik

Tahapan pemodelan merupakan inti dari proses pencarian pengetahuan (knowledge discovery), di mana data kecelakaan lalu lintas yang telah melalui proses prakondisi (preprocessing) diekstrak menjadi sebuah struktur matematis prediktif.



Gambar 23. Grafik Hasil Analisa Decissio Tree

Tahap evaluasi dilakukan untuk mengukur kemampuan model Decision Tree C4.5 dalam mengklasifikasikan faktor penyebab kecelakaan lalu lintas berdasarkan data pengujian yang digunakan.

Evaluasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* yang diperoleh dari *classification report* setelah proses *pruning* dilakukan pada pohon keputusan.

Secara lebih rinci, hasil evaluasi untuk masing-masing kelas menunjukkan adanya perbedaan performa yang cukup signifikan. Pada kelas Manusia, model menghasilkan nilai *precision* sebesar 0,98, *recall* sebesar 0,65, dan *F1-score* sebesar 0,78. Nilai *precision* yang sangat tinggi menunjukkan bahwa hampir seluruh prediksi yang diklasifikasikan sebagai faktor manusia merupakan prediksi yang benar. Selain itu, nilai *recall* sebesar 0,65 mengindikasikan bahwa model mampu mengenali sekitar 65% dari seluruh kejadian kecelakaan yang benar-benar disebabkan oleh faktor manusia. Hasil ini menunjukkan bahwa model memiliki kemampuan yang baik dalam mengidentifikasi faktor manusia sebagai penyebab kecelakaan.

V. PENUTUP

A.Kesimpulan

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan, dapat diambil beberapa kesimpulan sebagai berikut :

1. Algoritma C4.5 berhasil membentuk model klasifikasi penyebab kecelakaan lalu lintas. Model Decision Tree C4.5 mampu menghasilkan struktur pohon keputusan yang dapat digunakan untuk mengidentifikasi penyebab kecelakaan lalu lintas berdasarkan atribut-atribut yang terdapat dalam dataset. Struktur pohon yang terbentuk memberikan representasi aturan keputusan (*decision rules*) yang mudah dipahami dan diinterpretasikan sehingga dapat digunakan sebagai alat pendukung analisis faktor penyebab kecelakaan
2. Kondisi Permukaan Jalan merupakan faktor yang paling berpengaruh. Berdasarkan hasil analisis feature importance, atribut Kondisi Permukaan Jalan memiliki nilai pengaruh tertinggi dibandingkan atribut lainnya dalam proses klasifikasi. Temuan ini menunjukkan bahwa kondisi fisik jalan memiliki peranan dominan dalam membedakan penyebab kecelakaan lalu lintas. Selain itu, atribut lain yang turut memberikan kontribusi penting dalam pembentukan model adalah Tingkat Kecelakaan, Batas Kecepatan di Lokasi, Kondisi Cahaya, Kategori Kecelakaan, Fungsi Jalan, Hasil tersebut menunjukkan bahwa faktor infrastruktur jalan dan lingkungan lalu lintas memiliki pengaruh yang signifikan terhadap terjadinya kecelakaan

3. Model menghasilkan tingkat akurasi sebesar 65,22%. Hasil evaluasi menunjukkan bahwa model C4.5 yang dibangun memperoleh nilai akurasi sebesar 65,22% setelah dilakukan proses *pruning*. Nilai tersebut menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data pengujian dengan cukup baik. Kinerja terbaik diperoleh pada kelas Manusia dengan nilai *precision* sebesar 98% dan *F1-score* sebesar 78%, yang menunjukkan bahwa model memiliki kemampuan yang baik dalam mengenali kecelakaan yang disebabkan oleh faktor manusia
4. Ketidakseimbangan data memengaruhi performa model
Hasil evaluasi juga menunjukkan bahwa performa model pada kelas Jalan dan Kendaraan masih relatif rendah dibandingkan kelas Manusia. Kondisi ini disebabkan oleh distribusi data yang tidak seimbang (*imbalanced dataset*), di mana sebagian besar data berasal dari kelas Manusia, sedangkan jumlah data pada kelas Jalan dan Kendaraan jauh lebih sedikit. Akibatnya model cenderung lebih baik dalam mengenali kelas mayoritas dibandingkan kelas minoritas.

B. Saran

Berdasarkan hasil penelitian yang telah dilakukan, beberapa saran yang dapat diberikan untuk pengembangan penelitian selanjutnya adalah sebagai berikut. Menambah jumlah dan kualitas data penelitian. Penelitian selanjutnya disarankan menggunakan dataset dengan jumlah data yang lebih besar serta periode pengamatan yang lebih panjang. Penambahan data dapat meningkatkan kemampuan model dalam mempelajari pola kecelakaan secara lebih representatif dan menghasilkan tingkat akurasi yang lebih baik. Melakukan penyeimbangan data (*data balancing*) Karena penelitian ini menunjukkan adanya ketidakseimbangan jumlah data antar kelas, maka penelitian selanjutnya dapat menerapkan teknik *oversampling*, *undersampling*, atau metode seperti SMOTE (*Synthetic Minority Oversampling Technique*) untuk meningkatkan representasi kelas minoritas sehingga performa klasifikasi menjadi lebih seimbang. Membandingkan dengan algoritma klasifikasi lainnya. Untuk memperoleh model terbaik, penelitian selanjutnya dapat melakukan perbandingan kinerja antara algoritma C4.5 dengan metode klasifikasi lain seperti:

- a. Random Forest
- b. Naïve Bayes
- c. Support Vector Machine (SVM), dan lainnya

DAFTAR PUSTAKA

- [1] Abdullah, P., & Sipos, T. (2022). Drivers' Behavior and Traffic Accident Analysis Using Decision Tree Method. *Sustainability (Switzerland)*, 14(18), 1–11. <https://doi.org/10.3390/su141811339>
- [2] Alfandi, M. S., & Fatah, Z. (2024). Penerapan data Mining Menggunakan Metode K-Means Clustering Untuk Analisa Penjualan Toko Umama Hijab Kaliwates Jember. *Jurnal riset sistem informasI*, 1(4), 94–102.
- [3] Alfiansyah, G., Hendra, & Azhar, R. (2025). *Klasifikasi Keparahannya Kecelakaan Lalu Lintas Berbasis Decision Tree dan Teknik Smote*. *SEPTEMBER 2025*, 40–47.
- [4] Ananda, R., Sumarno, S., & Gunawan, I. (2022). Prediksi Promosi Jabatan Karyawan JNE Pematangsiantar Dengan Algoritma C4.5. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 3(3), 358. <https://doi.org/10.30865/json.v3i3.3763>
- [5] Atsir, E. M., Nurmalitasari, N., & Sari, A. A. (2025). Traffic Accident Severity Classification System Using Random Forest Algorithm. *J-Intech*, 13(02), 272–281. <https://doi.org/10.32664/j-intech.v13i02.2089>
- [6] Biyantoro, A. S., & Prasetyo, B. (2024). Penerapan Decision Tree untuk Klasifikasi Status Kesehatan dengan perbandingan KNN dan Naive Bayes. *Indonesian Journal of Informatic Research and Software Engineering*, 4(1), 47–55. <https://journal.irpi.or.id/index.php/ijirse/article/view/1342>
- [7] Febryana, S., Wulandari, I. A., & Prabowo, D. (2024). Rancang Bangun Aplikasi Pengolahan Data Inventory Barang Berbasis Desktop Pada Cv. Hamim Group. *Jurnal Mahasiswa Sistem Informasi (JMSI)*, 6(1), 24–35. <https://doi.org/10.24127/jmsi.v6i1.7553>
- [8] Hutagalung, W. M. S. N., Tony, & Perdana, N. J. (2023). *Analisis Sentimen Pada Opini Kenaikan Harga Bahan Bakar Minyak Pada Media Sosial Twitter*. 8(2), 280–284.
- [9] Indra, H., Yuliana, D., Ulia, U., & Adriant, R. (2023). *Strategi Promosi Penjualan Menggunakan C4.5 (Studi Kasus DI Restoran Pekanbaru, Plaza)*. 12, 1281–1292.

- [10] Irawan, B., Bahtiar, A., Studi, P., Informatika, T., Cirebon, K., & Adakami, A. (2023). *Penggunaan Algoritma Naïve Bayes Dalam Menganalisis*. 7(6), 3879–3885.
- [11] Kusumastutie, N. S., Patria, B., Kusrohmaniah, S., & Hastjarjo, T. D. (2024). A review of accident data for traffic safety studies in Indonesia. *IOP Conference Series: Earth and Environmental Science*, 1294(1). <https://doi.org/10.1088/1755-1315/1294/1/012012>
- [12] Naldy, E. T., & Andri, A. (2021). Penerapan Data Mining Untuk Analisis Daftar Pembelian Konsumen Dengan Menggunakan Algoritma Apriori Pada Transaksi Penjualan Toko Bangunan MDN. *Jurnal Nasional Ilmu Komputer*, 2(2), 89–101. <https://doi.org/10.47747/jurnalnik.v2i2.525>
- [13] Noiija, H. C., Wemaf, P. A., Nurdianty, O. A., Sohila, W., Haumahu, S., Yusuf, H., Tomagola, K., Wa Rindi La Ana, & Salaiswa, T. S. (2023). Perancangan Sistem Informasi Akuntansi Untuk Analisis Siklus Pendapatan Pada Orantata Celular Menggunakan DFD Dan Flowchart. *Jurnal Bisnis Dan Manajemen (JURBISMAN)*, 1(2), 577–592. <https://doi.org/10.61930/jurbisman.v1i2.188>
- [14] Octavia, K., Yulianto, F., & Siswa, T. A. Y. (2025). Penentuan jenis kecelakaan lalu lintas di kota samarinda menggunakan metode Decission Tree. *Didaktik: Jurnal Ilmiah PGSD FKIP Universitas Mandiri*, 11(1), 212–226.
- [15] Rahmat, A., Hardi, H., Syam, F. A., Zamzami, Z., Febriadi, B., & Windarto, A. P. (2021). Utilization of the field of data mining in mapping the area of the Human Development Index (HDI) in Indonesia. *Journal of Physics: Conference Series*, 1783(1). <https://doi.org/10.1088/1742-6596/1783/1/012035>
- [16] Ramadhanu, A., Defit, S., & Wahhab, S. (2021). *Hybrid Data Mining with the Combination of K-Means Algorithm and C4 . 5 to Predict Student Achievement*. 5(2), 180–189. <https://doi.org/10.29099/ijair.v6i1.225>
- [17] Ramadhon, R. N., Ogi, A., Agung, A. P., Putra, R., Febrihartina, S. S., & Firdaus, U. (2024). Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak Aktif pada Data Bank. *Karimah Tauhid*, 3(2), 1860–1874. <https://doi.org/10.30997/karimahtauhid.v3i2.11952>
- [18] Rozi, M., Pujiastutie, E. T., & Belladona, M. (2025). Peningkatan Keselamatan Jalan Dengan Metode Hazard Identification and Risk Assessment (Hira) Pada Ruas Jalan Pariwisata Pantai Panjang Kota Bengkulu. *RIGGS: Journal of Artificial Intelligence and Digital Business*, 4(3), 1559–1556. <https://doi.org/10.31004/riggs.v4i3.2185>
- [19] Salsabila, N. A., Goejantoro, R., & Syaripuddin, S. (2023). Klasifikasi Tingkat Keparahan Korban Kecelakaan Lalu Lintas Di Kota Samarinda Menggunakan Algoritma K-Nearest Neighbor dan Naive Bayes. *Jurnal EKSPONENSIAL*, 14(2), 99–106. <http://jurnal.fmipa.unmul.ac.id/index.php/expo-nensial99>
- [20] Saragih, R., Sihombing, M., Sihombing, A., & Kristianto Hondro, R. (2024). *Application of Decision Tree Algorithm Method To Analyze Traffic Accident Patterns*. 4(2), 200–202. <https://doi.org/10.46880/tamika.Vol4No2>
- [21] Sartika, D., & Jumadi, J. (2023). Implementasi Data Mining Untuk Pengelompokkan Wilayah Pelanggaran Lalu Lintas Menggunakan Metode K-Means Pada Polres Bengkulu. *Journal of Science and Social Research*, 4307(2), 345–349. <http://jurnal.goretanpena.com/index.php/JSSR>
- [22] Tourindo, W., & Lubis, M. R. A. (2025). 9 Bulan, 289 Kecelakaan Tewaskan 24 Orang, Terbaru, Warga Benteng Terlindas Dump Truk di Kota Bengkulu. *Rakyat Bengkulu News Online*. https://harianrakyatbengkulu.bacakoran.co/borgol/read/46841/9-bulan-289-kecelakaan-tewaskan-24-orang-terbaru-warga-benteng-terlindas-dump-truk-di-kota-bengkulu?utm_source=chatgpt.com
- [23] Wulandari, G., Sasmita, S., & Rahmadayanti, F. (2024). Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Obat. *Jurnal Teknik Informasi Dan Komputer (Tekinkom)*, 7(1), 144. <https://doi.org/10.37600/tekinkom.v7i1.1270>