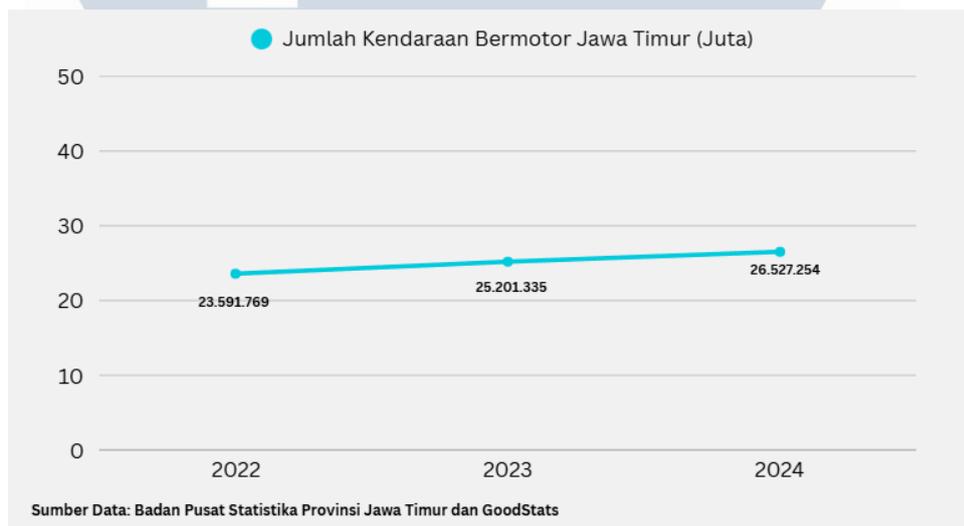


# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

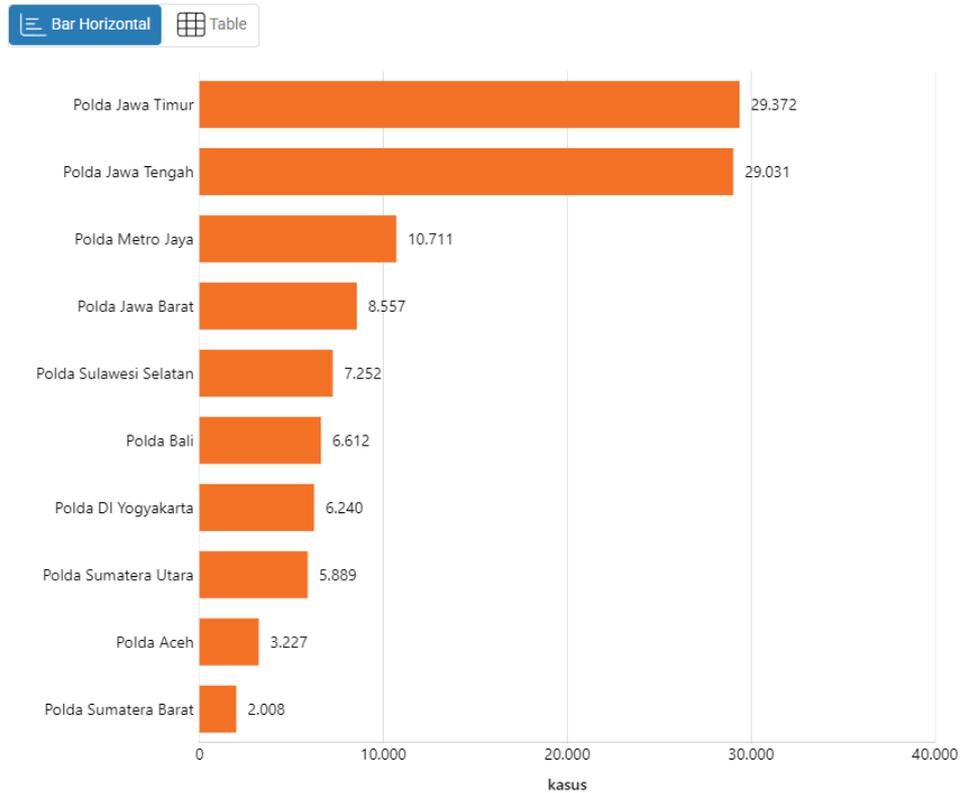
Kecelakaan lalu lintas merupakan peristiwa yang berisiko tinggi dan berpotensi menimbulkan kerugian jiwa maupun materi. Baik pengendara, pejalan kaki, hingga masyarakat sekitar jalan dapat menjadi korban ketika kecelakaan terjadi. Kesadaran terhadap pentingnya keselamatan berkendara sering kali diabaikan oleh pengguna jalan yang cenderung melanggar peraturan, berkendara secara tidak tertib, atau melakukan tindakan yang membahayakan pengguna lainnya.



Gambar 1.1 Peningkatan Pengguna Kendaraan Bermotor 2022-2024 [1], [2] Lalu lintas merupakan aspek vital dalam kehidupan sehari-hari yang berhubungan erat dengan mobilitas masyarakat. Seiring dengan pertumbuhan penduduk dan aktivitas ekonomi, jumlah kendaraan bermotor di Provinsi Jawa Timur meningkat signifikan dari tahun ke tahun. Berdasarkan data dari Badan Pusat Statistik Provinsi Jawa Timur dan GoodStats pada gambar 1.1, jumlah kendaraan bermotor di provinsi ini tercatat sebesar 23,5 juta unit pada tahun 2022, naik menjadi 25 juta unit pada 2023, dan mencapai 26,5 juta unit pada

2024 [1], [2]. Lonjakan ini menunjukkan tren kepadatan kendaraan yang terus meningkat, yang berpotensi memberikan tekanan terhadap sistem lalu lintas dan keselamatan jalan.

10 Satuan Kepolisian Daerah/Polda dengan Laporan Kasus Kecelakaan Lalu Lintas Terbanyak (Januari-November 2023)



Gambar 1.2 Data Laporan Kecelakaan Polda 2023[3]

Peningkatan jumlah kendaraan tersebut juga diiringi oleh naiknya jumlah kasus kecelakaan lalu lintas. Pada gambar 1.2, Menurut DataBoks (2023), selama periode Januari hingga November 2023, Polda Jawa Timur mencatat 29.372 kasus kecelakaan lalu lintas, menjadikannya sebagai satuan kepolisian daerah dengan jumlah kasus tertinggi di Indonesia, sedikit di atas Polda Jawa Tengah (29.031 kasus) dan jauh di atas wilayah lain seperti Polda Metro Jaya (10.711 kasus) dan Polda Jawa Barat (8.557 kasus). Data terbaru menunjukkan bahwa pada tahun 2024, jumlah kasus kecelakaan lalu lintas di Jawa Timur mencapai 28.272 kasus, menurut laporan dari Kepolisian Daerah Jawa Timur. Meski mengalami sedikit penurunan dibanding tahun sebelumnya, angka ini tetap

tergolong tinggi dan mencerminkan urgensi akan perlunya analisis serta pemanfaatan data untuk mengidentifikasi pola serta faktor-faktor penyebab kecelakaan secara lebih akurat dan sistematis.

Tingginya angka kecelakaan tersebut menjadi isu serius yang perlu ditindaklanjuti. Pemerintah, kepolisian dan pengguna jalan perlu memahami faktor-faktor penyebab kecelakaan, seperti waktu kejadian, kondisi cuaca, dan kondisi jalan, untuk melakukan pembinaan di wilayah rawan serta meningkatkan pengawasan terhadap pengendara yang berpotensi membahayakan dan dapat menjadi wawasan terhadap berkendara secara tertib.

Polda mencatat data terkait kasus kecelakaan, dengan mengkategorikan tingkat kecelakaan menjadi tingkat ringan, sedang, dan berat. Pengembangan model untuk analisis klasifikasi belum dilakukan agar dapat melakukan otomasi pendataan tingkat kecelakaan secara *real-time*. Data yang telah dikumpulkan belum dapat digunakan secara optimal yang dapat dikembangkan. Maka dari itu, penelitian ini memanfaatkan dataset yang telah dikumpulkan oleh Polda Jawa Timur untuk mengembangkan model klasifikasi serta pengaplikasian agar dapat membentuk aplikasi sederhana yang dapat meningkatkan efisiensi kerja dan efektifitas dari data.

Beberapa penelitian serupa telah dilakukan dalam mengklasifikasikan tingkat kecelakaan lalu lintas. Berdasarkan penelitian[4], menggunakan algoritma KNN, Naïve Bayes, dan Random Forest dalam klasifikasi kecelakaan di Kota Samarinda, dan memperoleh akurasi sebesar 75–79%. Sementara itu, berdasarkan penelitian [5], menggunakan dataset ISF Lebanon dan menerapkan Support Vector Machine (SVM), dengan hasil akurasi mencapai 86%. Hal ini menunjukkan adanya peluang untuk melakukan eksplorasi model klasifikasi lain dan strategi peningkatan performa, terutama dalam konteks data lokal seperti Polda Jatim.

Berdasarkan penelitian terkait klasifikasi tingkat kecelakaan, penelitian ini akan melakukan perbandingan terhadap beberapa model klasik, seperti SVM, KNN, Decision Tree, dan Naïve Bayes. Berdasarkan penelitian sebelumnya,

analisis klasifikasi tingkat kecelakaan belum ada yang mengangkat penggunaan model Decision Tree dan penggunaan metode variasi SMOTE (*Synthetic Minority Oversampling Technique*), ROS (Random Over Sampling), dan RUS (Random Under Sampling). Penelitian ini akan mengembangkan dan membandingkan seluruh model menggunakan atau tidak menggunakan teknik *sampling* untuk mencari tahu performa model terhadap klasifikasi tingkat kecelakaan berdasarkan faktor utama. Terhadap kasus kecelakaan yang tinggi di Polda Jatim, belum ada yang menerapkan analisis klasifikasi data yang mencakup wilayah Provinsi dengan angka kecelakaan tertinggi, juga pengembangan sederhana untuk mengklasifikasi kecelakaan berdasarkan faktor utama di Jawa Timur. Model yang telah dikembangkan akan diaplikasikan menjadi aplikasi nyata untuk optimasi pengumpulan data secara efisien, dengan melakukan otomatisasi terhadap prediksi terbaik.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka terbentuklah rumusan masalah terhadap penelitian ini berupa:

1. Apa saja faktor utama yang paling berpengaruh dalam menentukan tingkat kecelakaan lalu lintas di Jawa Timur?
2. Bagaimana performa algoritma KNN, SVM, Decision Tree, dan Naïve Bayes dalam mengklasifikasikan tingkat kecelakaan, baik dengan maupun tanpa teknik *sampling* (SMOTE, ROS, dan RUS)?
3. Bagaimana implementasi model terbaik dalam bentuk aplikasi web prediksi tingkat kecelakaan menggunakan Streamlit?

## 1.3 Batasan Masalah

Batasan masalah yang terdapat pada penelitian ini diantara lain adalah:

1. Model yang digunakan adalah algoritma KNN, SVM, Decision Tree, dan Naïve Bayes sebagai perbandingan model terbaik untuk klasifikasi tingkat kecelakaan lalu lintas.
2. Proses *sampling data* menggunakan variasi teknik SMOTE, ROS, dan RUS.

3. Klasifikasi menggunakan data bersumber dari Kepolisian Daerah Jawa Timur mencakup data kecelakaan lalu lintas tahun 2024.
4. Faktor-faktor yang dapat dibawa ke dalam analisis data merupakan faktor umum tercatat, tidak dapat menggunakan faktor spesifik, seperti lupa rem, rem blong, atau faktor personal lainnya.

#### **1.4 Tujuan dan Manfaat Penelitian**

##### **1.4.1 Tujuan Penelitian**

Tujuan Penelitian yang telah ditetapkan berdasarkan Batasan masalah adalah sebagai berikut:

1. Mengidentifikasi faktor utama yang memengaruhi tingkat kecelakaan lalu lintas di Jawa Timur.
2. Membandingkan performa algoritma KNN, SVM, Decision Tree, dan Naïve Bayes dengan dan tanpa metode sampling (SMOTE, ROS, dan RUS) dalam klasifikasi tingkat kecelakaan.
3. Membangun aplikasi web prediksi tingkat kecelakaan menggunakan model dengan akurasi terbaik.

##### **1.4.2 Manfaat Penelitian**

Penelitian ini diharapkan juga dapat memberikan manfaat berupa:

1. Mengidentifikasi dan menyajikan faktor-faktor utama yang berpengaruh terhadap peningkatan tingkat kecelakaan lalu lintas di Jawa Timur.
2. Memberikan informasi terkait performa algoritma KNN, SVM, Decision Tree, dan Naïve Bayes, baik dengan maupun tanpa penerapan metode SMOTE, dalam klasifikasi tingkat kecelakaan lalu lintas.
3. Menghasilkan aplikasi web sederhana dalam memprediksi tingkat kecelakaan menggunakan model dengan akurasi terbaik.

## **1.5 Sistematika Penulisan**

### **BAB I PENDAHULUAN**

Bab pertama pada penelitian ini berisikan latar belakang yang akan membentuk rumusan masalah, tujuan dari penelitian, juga manfaat penelitian.

### **BAB II LANDASAN TEORI**

Pada bab kedua berisikan kajian literatur yang menjelaskan konsep, teori, dan elemen terkait penelitian

### **BAB III METODOLOGI**

Pada bab tiga, dijelaskan detail dari metodologi yang digunakan, penjelasan terkait data yang digunakan dan alur dari penelitian ini.

### **BAB IV ANALISIS DAN HASIL PENELITIAN**

Pada bab keempat berisikan detail prosedur penelitian menggunakan metode CRISP-DM dalam melakukan analisis klasifikasi tingkat kecelakaan, penjelasan terstruktur hingga mendapatkan hasil dan dibandingkan dengan hasil dari penelitian terkait sebagai perbandingannya, juga mencakup seluruh percobaan yang didiskusikan pada akhir bab.

### **BAB V SIMPULAN DAN SARAN**

Pada bab kelima menampilkan hasil yang menjawab tujuan penelitian secara ringkas dan spesifik, juga membawakan saran untuk pengembangan topik atau metode pada penelitian dan hasil karya kedepannya.

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A