

BAB I

PENDAHULUAN

Bab ini berisi penjelasan tentang latar belakang dari penelitian, rumusan masalah yang ingin diselesaikan, tujuan yang ingin dicapai, manfaat yang diharapkan, serta batasan ruang lingkup penelitian. Semua bagian ini disusun untuk memberi gambaran awal tentang arah dan fokus dari penelitian yang dilakukan.

1.1 Latar Belakang

Transportasi memiliki peran yang krusial dalam mendukung perekonomian dan menunjang pertumbuhan ekonomi serta meningkatkan kualitas hidup, baik di wilayah perkotaan maupun di pedesaan. Namun, seiring dengan meningkatnya jumlah penduduk dan kendaraan pribadi, permasalahan baru turut muncul, seperti kemacetan, polusi udara, kebisingan, dan yang paling fatal, yakni meningkatnya angka kecelakaan lalu lintas [1]. Kecelakaan lalu lintas menjadi salah satu isu serius yang dihadapi negara berkembang, termasuk Indonesia. Setiap tahun, ribuan korban jiwa berjatuhan dan kerugian materiil terus meningkat akibat kecelakaan di jalan. Menurut data *World Health Organization* (WHO), sekitar 1,3 juta orang meninggal setiap tahunnya akibat kecelakaan lalu lintas, dan lebih dari 50 juta lainnya mengalami cedera non-fatal [2]. Hal ini menempatkan kecelakaan lalu lintas sebagai salah satu penyebab utama kematian global.

Di Indonesia, masalah kecelakaan lalu lintas masih mengkhawatirkan. Berdasarkan data dari Kepolisian Negara Republik Indonesia (Korlantas Polri), jumlah kecelakaan lalu lintas pada tahun 2022 mencapai 152.162 kasus, meningkat dari tahun sebelumnya yang berjumlah 103.645 kasus [3]. Di Kota Surabaya, sebagai salah satu kota metropolitan terbesar di Indonesia, angka kecelakaan lalu lintas tergolong tinggi. Data dari Badan Penaggulangan Bencana Daerah (BPDB) kota Surabaya mencatat bahwa pada tahun 2020 terjadi 1.741 kasus kecelakaan, menurun pada tahun 2021 menjadi 1.262, namun kembali meningkat pada tahun 2022 menjadi 2.132 kasus [4]. Data dari Suara Surabaya menunjukkan bahwa dalam periode 2019-2023, tercatat 1.978 korban kecelakaan lalu lintas di Surabaya, 815 di antaranya meninggal dunia. Tahun 2022 menjadi puncak tertinggi angka kecelakaan, sementara tahun 2021

mencatat jumlah terendah. Mayoritas korban adalah laki-laki dan melibatkan kendaraan bermotor roda dua [5].

Kecelakaan lalu lintas menimbulkan kerugian fisik, material, sosial, dan ekonomi. Di Surabaya, faktor utama penyebabnya adalah perilaku manusia, terutama pelanggaran lalu lintas dan rendahnya kesadaran keselamatan. Permasalahan ini bersifat multidimensi sehingga membutuhkan penanganan komprehensif dan berbasis data [6]. Kondisi ini menegaskan bahwa permasalahan kecelakaan lalu lintas bersifat multidimensi, mencakup aspek teknis, sosial, dan budaya, sehingga memerlukan pendekatan komprehensif dan berbasis data dalam upaya penanggulangannya.

Dalam menghadapi tantangan tersebut, Pendekatan berbasis data melalui *data mining* dan *machine learning* penting untuk merumuskan kebijakan keselamatan jalan. Metode klasifikasi dapat dimanfaatkan untuk memprediksi tingkat keparahan kecelakaan berdasarkan data historis dan mengidentifikasi pola serta faktor-faktor yang berkontribusi terhadap tingkat keparahan kecelakaan. Pada penelitian ini, tingkat keparahan kecelakaan ditentukan berdasarkan data historis kejadian kecelakaan lalu lintas yang telah dikategorikan ke dalam tiga kelas, yaitu ringan, sedang, dan berat. Kategori tersebut merujuk pada informasi yang tercatat dalam data mengenai kondisi korban dan dampak kecelakaan. Dengan klasifikasi berbasis data historis ini, analisis dapat dilakukan secara lebih langsung terhadap pola kejadian yang nyata, sekaligus memungkinkan penerapan model klasifikasi untuk mengidentifikasi faktor-faktor yang paling berkontribusi terhadap tingginya tingkat keparahan kecelakaan di Kota Surabaya [7].

Untuk mencapai tujuan tersebut, penelitian ini menggunakan Catboost (*Categorical Boosting*) sebagai algoritma utama karena mampu menangani data kategorikal tanpa encoding, mengurangi kompleksitas preprocessing, risiko informasi hilang pada data kategorikal, mencegah *target leakage* melalui *ordered boosting*, serta memberikan prediksi yang lebih stabil dibanding XGBoost dan LightGBM [8].

Hancock dan Khoshgoftaar (2020) dalam kajiannya menyatakan bahwa Catboost merupakan algoritma *Gradient Boosted Decision Tree* (GBDT) yang sangat efektif digunakan pada data heterogen dengan banyak fitur kategorikal. Studi tersebut menunjukkan bahwa Catboost memiliki performa yang kompetitif di berbagai bidang, termasuk kesehatan, keuangan, rekayasa lalu lintas, dan keamanan siber. Selain itu, Catboost juga menyediakan interpretasi model yang baik melalui analisis *feature*

importance dan mendukung penerapan pada skala *big data*. Oleh karena itu, Catboost dipandang sebagai algoritma yang tepat untuk digunakan dalam penelitian klasifikasi tingkat keparahan kecelakaan lalu lintas yang melibatkan berbagai fitur kategorikal [9] [10].

Untuk meningkatkan kualitas data yang digunakan dalam pelatihan model, penelitian ini menerapkan *feature engineering* berbasis *domain knowledge* untuk mengubah data mentah menjadi fitur yang lebih bermakna, seperti klasifikasi waktu kejadian, jenis kendaraan, jumlah kendaraan, dan kategori lokasi. Pendekatan ini terbukti dapat meningkatkan akurasi prediksi sekaligus memperbaiki interpretabilitas model. Mende et al. (2023) menunjukkan bahwa fitur yang direkayasa secara manual oleh ahli domain memberikan performa prediksi yang lebih baik sekaligus meningkatkan interpretabilitas model [11].

Masalah ketidakseimbangan kelas juga menjadi tantangan penting dalam klasifikasi kecelakaan lalu lintas. Untuk mengatasi hal ini, digunakan metode SMOTENC (*Synthetic Minority Over-sampling Technique for Nominal and Continuous*). SMOTENC merupakan perluasan SMOTE yang khusus menangani dataset berfitur campuran, dengan mensintesis sampel minoritas melalui interpolasi pada variabel numerik dan pemilihan acak pada variabel kategorikal. Sehingga meningkatkan sensitivitas model tanpa merusak representasi data asli. [12].

Model Catboost dioptimasi menggunakan Optuna, yang mengandalkan *Bayesian Optimization* berbasis *TPE* dan *pruning* dinamis untuk meningkatkan efisiensi pencarian hyperparameter. Studi Shekhar et al. (2022) menunjukkan bahwa Optuna memberikan keseimbangan terbaik antara akurasi (*f1-score*) dan waktu komputasi dibandingkan HyperOpt, Optunity, dan SMAC pada benchmark CASH [13].

Berdasarkan integrasi pendekatan tersebut, penelitian ini diharapkan Penelitian ini diharapkan menghasilkan model klasifikasi yang akurat sekaligus memberikan pemahaman mendalam tentang karakteristik kecelakaan lalu lintas di Surabaya, sehingga dapat menjadi dasar strategi keselamatan jalan berbasis data dan mendukung kebijakan penurunan angka kecelakaan.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, penelitian ini merumuskan beberapa permasalahan yang akan dibahas yakni sebagai berikut:

- 1) Bagaimana membangun model klasifikasi tingkat keparahan kecelakaan lalu lintas di Kota Surabaya berdasarkan data historis?
- 2) Bagaimana pengaruh penerapan model algoritma Catboost terhadap akurasi dan interpretabilitas model klasifikasi?
- 3) Bagaimana efektivitas metode *hyperparameter tuning* menggunakan Optuna dalam mengoptimalkan kinerja model klasifikasi tingkat keparahan kecelakaan lalu lintas?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disusun, maka tujuan yang ingin dicapai pada penelitian ini adalah sebagai berikut:

- 1) Untuk membangun model klasifikasi tingkat keparahan kecelakaan lalu lintas di Kota Surabaya dengan memanfaatkan data historis sebagai dasar kategorisasi tingkat keparahan.
- 2) Untuk menganalisis pengaruh penerapan model algoritma Catboost terhadap akurasi dan interpretabilitas model klasifikasi.
- 3) Untuk mengevaluasi efektivitas metode *hyperparameter tuning* menggunakan Optuna dalam mengoptimalkan kinerja model klasifikasi tingkat keparahan kecelakaan lalu lintas.

1.4 Manfaat Penelitian

Berdasarkan tujuan penelitian di atas, gambaran manfaat yang dapat diperoleh pada penelitian ini adalah sebagai berikut:

Manfaat Akademis:

- 1) Menjadi referensi ilmiah penerapan metode klasifikasi multikelas pada kasus data kecelakaan lalu lintas.
- 2) Memberikan kontribusi dalam pengembangan ilmu *data mining* dan *machine learning* yang menggunakan integrasi Catboost dengan *feature engineering*, dan *tuning hyperparameter* dalam konteks data tabular.

- 3) Menambahkan literatur lokal dalam bidang analisis kecelakaan lalu lintas di kawasan perkotaan padat seperti Surabaya yang masih relatif terbatas.

Manfaat Praktis:

- 1) Menyediakan model prediksi keparahan kecelakaan yang dapat membantu pemerintah daerah dalam pengambilan kebijakan keselamatan lalu lintas di Kota Surabaya.
- 2) Mendukung instansi seperti Dinas Perhubungan, Kepolisian, dan BPBD dalam memprioritaskan lokasi atau kondisi jalan yang membutuhkan penanganan khusus.

1.5 Batasan Masalah

Agar penelitian ini tetap terarah dan fokus, maka ditetapkan beberapa batasan masalah sebagai berikut:

- 1) Data yang digunakan merupakan data kecelakaan lalu lintas di wilayah Surabaya dengan rentang waktu tahun 2020 hingga 2024 yang telah tersedia dalam bentuk tabular dan bersumber dari instansi resmi terkait yaitu Polrestabes Surabaya.
- 2) Fokus pada klasifikasi keparahan kecelakaan menjadi menjadi tiga kategori yaitu ringan, sedang, berat yang didasarkan pada data historis.
- 3) Model klasifikasi menggunakan algoritma Catboost sebagai model utama. Untuk meningkatkan kinerja model, dilakukan proses optimasi *hyperparameter* secara otomatis menggunakan Optuna.
- 4) Parameter yang akan dioptimasi berupa parameter Catboost yaitu *learning_rate*, *iteration*, dan *depth*.