

BAB 1 PENDAHULUAN

1.1 Latar Belakang Masalah

Kecelakaan lalu lintas merupakan salah satu permasalahan utama dalam keselamatan publik di seluruh dunia. Walaupun tingkat mortalitas akibat kecelakaan lalu lintas menurun sebanyak 5% sejak tahun 2010, angka tersebut masih mencapai 1,19 juta nyawa [1]. Kecelakaan lalu lintas tidak hanya menyebabkan kerugian jiwa dan cedera fisik, tetapi juga menimbulkan dampak ekonomi yang signifikan, termasuk biaya perawatan medis dan kerusakan kendaraan. Infrastruktur jalan dan sistem transportasi yang ideal bersifat responsif terhadap kebutuhan pengguna jalan dan mampu mengakomodasikan *human error* [2]. Statistik dari *National Motor Vehicle Crash Causations Survey* (NMVCCS) menggambarkan bahwa sekitar 93% kecelakaan lalu lintas terjadi karena *human error* atau faktor manusia [3]. Hal ini dikarenakan ketidakmampuan manusia untuk bertindak benar secara konsisten dalam semua bentuk kondisi lalu lintas [4].

Human error adalah bagian dari sifat alami manusia yang tidak dapat dihindari, namun dapat diidentifikasi dan dicegah [5]. Parlemen Swedia memperkenalkan Rancangan Undang-Undang Lalu Lintas dan Aturan Jalan (RUU LLAJ) yang dinamakan *Vision Zero* pada tahun 1997 [6]. *Vision Zero* dikenal juga sebagai *the safe system* bertujuan untuk menciptakan sistem dan infrastruktur jalan raya yang mengkompensasi *human error* sehingga tidak mengakibatkan korban jiwa akibat kecelakaan lalu lintas [7]. Penerapan *Vision Zero* berkontribusi terhadap penurunan angka kematian akibat kecelakaan lalu lintas di Swedia sebesar 66% dalam periode 1995 hingga 2015 [6], kesuksesan ini mendorong berbagai negara di Eropa untuk mengadopsi sistem serupa [8] dan berhasil mencatat penurunan tingkat fatalitas kecelakaan lalu lintas yang signifikan [9].

Seiring dengan perkembangan teknologi, kendaraan otonom mulai dikembangkan untuk mengurangi ketergantungan pada pengemudi manusia. Sistem ini menggunakan berbagai sensor seperti LiDAR dan RADAR serta beroperasi dalam ekosistem kendaraan otonom yang memungkinkan pertukaran informasi untuk meningkatkan pengambilan keputusan selama berkendara [10]. Teknologi ini menawarkan solusi untuk mengurangi faktor *human error* dalam kecelakaan lalu lintas, namun kendaraan otonom juga memiliki beberapa tantangan seperti

cuaca ekstrim dan kondisi infrastruktur yang kurang baik sehingga berdampak pada efektivitas sistem navigasi serta mengurangi kemampuan kendaraan dalam mengambil keputusan yang aman [11]. Mempelajari pola yang mendasari terjadinya kecelakaan merupakan langkah awal untuk meningkatkan keamanan di jalan raya [12], namun ketersediaan data mengenai kecelakaan lalu lintas di berbagai negara tidak konsisten [13]. Inkonsistensi ini dikarenakan faktor-faktor seperti standar pengumpulan data terkait kecelakaan lalu lintas berbeda-beda pada tiap negara, *under-reporting*, dan kurangnya pengumpulan variabel yang berkontribusi pada kecelakaan lalu lintas [14]. Penggunaan pendekatan tradisional dalam memprediksi jenis kecelakaan lalu lintas berdasarkan faktor lingkungan memiliki keterbatasan, terutama dalam menangani interaksi kompleks antar variabel dan volume data yang besar. *Machine learning* menawarkan solusi yang lebih efektif dengan kemampuan untuk mengidentifikasi pola non-linear, melakukan generalisasi terhadap data baru, serta menghasilkan prediksi yang lebih objektif dan dapat direproduksi.

Terdapat penelitian sebelumnya yang melakukan komparasi algoritma *machine learning* antara lain, *logistic regression*(LR), *K-nearest neighbor*(KNN), *Naive Bayes*(NB), *random forest*(RF), *extreme gradient boosting* (XGB) dan *AdaBoost* dalam memprediksi tipe kecelakaan *fatal/serious accident* atau *minor/non-injury* kecelakaan lalu lintas pada *South Island, New Zealand*. Hasil penelitian tersebut menunjukkan algoritma RF memiliki nilai akurasi tertinggi pada 86.64% diikuti XGBoost dengan nilai 81.63%, AdaBoost 73.15%, NB 66.49%, LR dan KNN 64.79% [15]. Pada penelitian lainnya, komparasi akurasi algoritma LR, *classification and regression tree* (CART), dan RF dalam memprediksi tipe kecelakaan *serious accident* atau *minor accident* pada jalan raya di negara Taiwan menghasilkan nilai 73.07% untuk algoritma LR, 72.6% untuk CART dan 73.38% untuk RF [16]. Penelitian-penelitian tersebut mempertimbangkan atribut pengemudi seperti umur, jenis kelamin, dan pengaruh zat psikoaktif sebagai faktor utama dalam memprediksi jenis kecelakaan. Kondisi jalan yang tidak terawat, seperti jalan berlubang atau retak, permukaan yang tidak rata juga merupakan faktor signifikan dalam meningkatkan risiko kecelakaan fatal. Kualitas material jalan yang buruk, tingginya volume lalu lintas, penggunaan kendaraan berat dan kondisi cuaca ekstrem juga mempercepat kerusakan jalan, sehingga menambah potensi bahaya bagi pengguna jalan [17].

Penelitian menunjukkan bahwa performa algoritma *ensemble* secara konsisten lebih tinggi dibandingkan dengan algoritma *single classifier*. Algoritma

ensemble mampu mengurangi risiko *overfitting* berkat keragaman model dasarnya, sehingga model akhir memiliki performa generalisasi yang lebih baik [18][19]. Penelitian komparatif terhadap algoritma *ensemble*, yaitu RF, XGB, *AdaBoost*, *Bagging-LSVM*, serta kombinasi keempat algoritma tersebut (*stacking*), menunjukkan bahwa *stacking* memiliki nilai *mean squared error (MSE)* terendah (0,0032). Di luar *stacking*, RF memiliki performa terbaik (0,0084), diikuti oleh XGB (0,0089), sementara *AdaBoost* (0,0317) dan *Bagging-LSVM* (0,0531) [20]. Penelitian lain yang membandingkan algoritma SVM, RF, dan XGBoost menggunakan *confusion matrix* menunjukkan bahwa XGB memiliki akurasi tertinggi sebesar 82%, diikuti oleh RF dengan akurasi 71% dan SVM 69% [21].

Berdasarkan seluruh penjelasan dan penjabaran yang telah diuraikan, penelitian ini akan berfokus pada implementasi algoritma RF dan XGB serta evaluasi menggunakan *confusion matrix* pada variabel lingkungan seperti kondisi cuaca, waktu kejadian, kondisi jalan, dan faktor lainnya. Pendekatan ini bertujuan untuk mengeksplorasi bagaimana faktor lingkungan secara spesifik mempengaruhi jenis kecelakaan lalu lintas sekaligus memberikan perspektif baru dalam analisis data kecelakaan lalu lintas serta mendukung pengembangan sistem transportasi yang lebih aman, terutama pada sistem kendaraan otonom.

1.2 Rumusan Masalah

1. Bagaimana cara implementasi algoritma *random forest* dan *extreme gradient boosting* (XGB) untuk prediksi tipe kecelakaan lalu lintas?
2. Bagaimana mengevaluasi hasil prediksi algoritma yang diterapkan menggunakan *confusion matrix* dan *classification report*?

1.3 Batasan Permasalahan

1. Penelitian ini berfokus dalam analisis hasil prediksi dan evaluasi algoritma RF dan XGB,
2. Penelitian ini hanya menggunakan *dataset* yang diambil dari kaggle yang berisikan 24 kolom dan 209307 baris data berjudul *Traffic Accidents*.

1.4 Tujuan Penelitian

1. Mengimplementasi algoritma RF dan XGB dalam memprediksi tipe kecelakaan lalu lintas.
2. Mengevaluasi hasil implementasi menggunakan akurasi, presisi, *recall* dan *F1-score* dari *confusion matrix*.

1.5 Manfaat Penelitian

1. Memberi wawasan terhadap tipe kecelakaan lalu lintas yang memiliki variabel konsisten sehingga pengembangan sistem transportasi dan pertimbangan keselamatan lalu lintas bisa diterapkan secara *universal*.
2. Menunjukkan perbedaan performa prediksi algoritma dan XGB terhadap data kecelakaan lalu lintas.

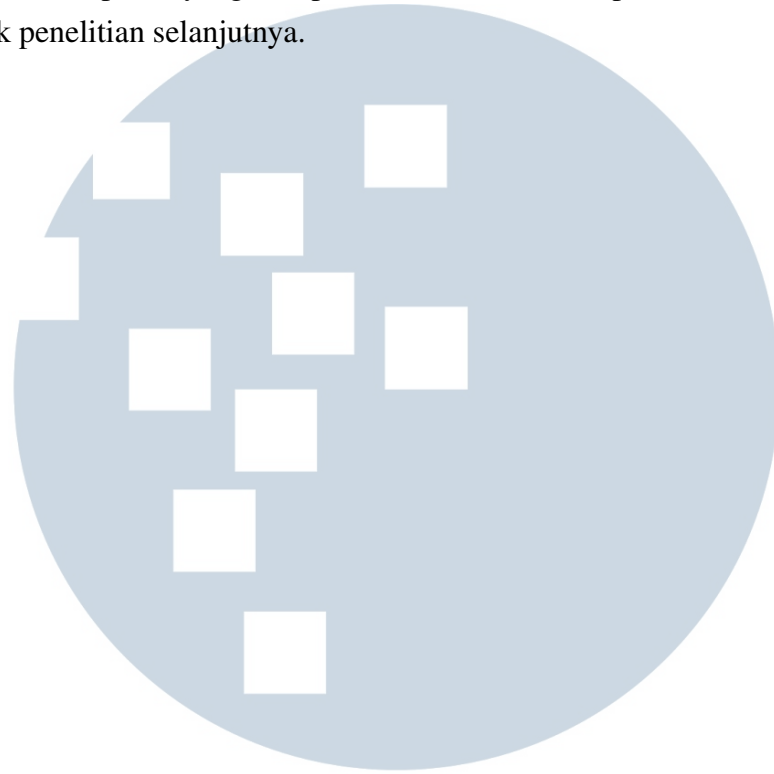
1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN
Bab ini akan menjabarkan latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan laporan.
- Bab 2 LANDASAN TEORI
Bab ini menjabarkan teori-teori yang terdapat pada penelitian ini seperti: *ensemble learning*, RF, XGB dan *confusion matrix*.
- Bab 3 METODOLOGI PENELITIAN
Bab ini menjelaskan alur metodologi yang akan dilakukan pada penelitian ini mulai dari pengumpulan data, *pre-processing data*, pembangunan model dan evaluasi hasil klasifikasi.
- Bab 4 HASIL DAN DISKUSI
Bab ini membahas hasil dari penelitian yang dilakukan juga menyertakan perbandingan akurasi antara algoritma RF dan XGB.

- Bab 5 KESIMPULAN DAN SARAN

Berisi kesimpulan yang didapat setelah melakukan penelitian ini dan saran untuk penelitian selanjutnya.



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA