

ANALISIS KLASIFIKASI KECELAKAAN LALU LINTAS DI POLRESTA BOGOR KOTA MENGGUNAKAN METODE K-NEAREST NEIGHBOR (KNN) DAN *NAÏVE BAYES* (NB)

Ziska Oktania, Jarot Prianggono

Sekolah Tinggi Ilmu Kepolisian, Indonesia

Email: ziskaoktania@gmail.com¹, komputerstik@gmail.com²

Abstrak

Kecelakaan lalu lintas di Kota Bogor meningkat seiring pertumbuhan kendaraan dan urbanisasi, memerlukan analisis berbasis data untuk mitigasi efektif. Penelitian ini bertujuan mengklasifikasikan kecelakaan menggunakan algoritma KNN dan NB guna mengidentifikasi pola berdasarkan lokasi, waktu, dan kondisi jalan. Pendekatan kuantitatif dengan teknik *data mining* diterapkan pada 502 data kecelakaan dari Polresta Bogor Kota (2020–2024). Data diproses melalui *preprocessing*, lalu diklasifikasikan dengan KNN dan NB, dievaluasi menggunakan *accuracy*, *precision*, *recall*, dan *Cohen's Kappa*. KNN lebih unggul (akurasi 91.89%) dibandingkan NB (37.70%), dengan kecelakaan ringan dominan dan kecelakaan berat lebih sering terjadi pada hari tertentu (Selasa, Kamis, Jumat, Minggu) di jalan utama. Temuan ini mendukung rekomendasi kebijakan seperti peningkatan patroli hari rawan dan optimasi infrastruktur jalan. Penelitian selanjutnya dapat mengembangkan model hybrid atau integrasi data real-time untuk analisis lebih mendalam.

Kata Kunci: Kecelakaan Lalu Lintas, K-Nearest Neighbor, Naïve Bayes, Kota Bogor.

Abstract

Traffic accidents in Bogor City are increasing along with the growth of vehicles and urbanization, requiring data-driven analysis for effective mitigation. This study aims to classify accidents using KNN and NB algorithms to identify patterns based on location, time, and road conditions. A quantitative approach with data mining techniques was applied to 502 accident data from the Bogor City Police (2020–2024). The data was processed through preprocessing, then classified with KNN and NB, evaluated using accuracy, precision, recall, and Cohen's Kappa. KNN is superior (91.89% accuracy) to NB (37.70%), with minor accidents predominant and heavy accidents occurring more frequently on certain days (Tuesday, Thursday, Friday, Sunday) on major roads. These findings support policy recommendations such as increased patrols on vulnerable days and optimization of road infrastructure. Further research can develop hybrid models or real-time data integration for more in-depth analysis.

Keywords: Traffic Accidents, K-Nearest Neighbor, Naïve Bayes, Bogor City

PENDAHULUAN

Manusia selalu melakukan mobilisasi demi memenuhi kebutuhannya, sehingga seiring berkembangnya zaman, inovasi dalam transportasi terus berkembang, termasuk dalam bentuk kendaraan bermotor. Kehadiran kendaraan dalam kehidupan sehari-hari kini telah menjadi bagian yang sangat penting dan tidak dapat dipisahkan dari aktivitas manusia. Seiring dengan meningkatnya kebutuhan mobilitas, jumlah kendaraan di Indonesia pun terus bertambah. Berdasarkan data dari Korlantas Polri, jumlah kendaraan di Indonesia telah mencapai 159.432.756 unit. Ketergantungan manusia terhadap kendaraan membawa dampak positif maupun negatif. Dari sisi positif, kendaraan mempermudah mobilisasi, memangkas waktu perjalanan, serta memungkinkan

perpindahan barang dan manusia secara lebih efisien. Namun, dampak negatif juga muncul, salah satunya adalah meningkatnya risiko kecelakaan lalu lintas yang dapat dipengaruhi oleh berbagai faktor, seperti pengemudi, kondisi kendaraan, infrastruktur jalan, serta faktor lingkungan lainnya.

Kecelakaan lalu lintas merupakan salah satu permasalahan serius yang dihadapi di berbagai kota besar di dunia, termasuk Kota Bogor. Sebagai salah satu daerah metropolitan dengan pertumbuhan urbanisasi yang pesat, Kota Bogor mengalami peningkatan angka kecelakaan lalu lintas yang cukup signifikan. Tingginya angka kecelakaan ini tidak hanya menyebabkan korban jiwa dan cedera, tetapi juga menimbulkan kerugian materiil yang besar serta mengganggu ketertiban dan keamanan lalu lintas (Yusuf et al., 2021). Berdasarkan data dari Polresta Bogor Kota, kecelakaan lalu lintas di wilayah ini menunjukkan tren yang mengkhawatirkan dengan berbagai faktor yang berperan, seperti kondisi jalan, perilaku pengemudi, jenis kendaraan, dan faktor cuaca. Penelitian sebelumnya yang dilakukan oleh Pratama et al. (2020) mengidentifikasi bahwa sebagian besar kecelakaan lalu lintas di kota besar disebabkan oleh faktor manusia, seperti kelelahan dan kelalaian pengemudi, yang diperburuk oleh kondisi jalan yang tidak ideal.

Kajian mengenai faktor-faktor penyebab kecelakaan lalu lintas telah banyak dilakukan di berbagai wilayah. Beberapa penelitian sebelumnya telah mengkaji berbagai pendekatan dalam memahami dan memitigasi risiko kecelakaan lalu lintas, termasuk melalui metode prediksi dan klasifikasi berbasis data (Zhang et al., 2022). Salah satu metode yang umum digunakan dalam analisis kecelakaan lalu lintas adalah teknik klasifikasi berbasis *machine learning*. Beberapa studi telah membandingkan berbagai metode klasifikasi, seperti *K-Nearest Neighbor* (KNN) dan *Naïve Bayes* (NB), dalam upaya mengidentifikasi pola kecelakaan serta menentukan faktor-faktor utama yang berkontribusi terhadap insiden kecelakaan lalu lintas (Kumar & Sinha, 2019). Dengan menggunakan pendekatan berbasis data mining dan *machine learning*, studi ini bertujuan untuk memberikan pemahaman yang lebih mendalam mengenai klasifikasi kecelakaan serta mendukung upaya mitigasi yang lebih efektif dan berbasis bukti.

Penelitian ini memiliki kebaruan ilmiah dalam penerapan metode *K-Nearest Neighbor* (KNN) dan *Naïve Bayes* (NB) dalam klasifikasi kecelakaan lalu lintas di Kota Bogor. Berbeda dari penelitian sebelumnya yang lebih banyak berfokus pada analisis deskriptif atau penggunaan metode statistik konvensional, penelitian ini menggunakan pendekatan klasifikasi berbasis *machine learning* untuk mengidentifikasi pola kecelakaan berdasarkan berbagai faktor, seperti lokasi kejadian, waktu kecelakaan, kondisi cuaca, serta karakteristik pengemudi dan kendaraan. Dengan demikian, hasil penelitian ini diharapkan dapat memberikan rekomendasi yang lebih berbasis data bagi pihak berwenang dalam meningkatkan keselamatan lalu lintas di Kota Bogor, serta mendukung upaya mitigasi kecelakaan sesuai dengan ketentuan yang diatur dalam Undang-Undang No. 22 Tahun 2009 tentang Lalu Lintas dan Angkutan Jalan serta Undang-Undang No. 2 Tahun 2002 tentang Kepolisian Negara Republik Indonesia.

Penelitian ini memberikan kontribusi baru dengan membandingkan kinerja

algoritma *K-Nearest Neighbor* (KNN) dan *Naïve Bayes* (NB) dalam klasifikasi kecelakaan lalu lintas di Kota Bogor, yang belum banyak dieksplorasi dalam studi sebelumnya (Kumar & Sinha, 2019; Aulia & Prasetyo, 2020). Berbeda dengan penelitian sebelumnya yang lebih berfokus pada analisis deskriptif atau metode statistik konvensional (Pratama et al., 2020; Yusuf et al., 2021), penelitian ini menggunakan pendekatan *machine learning* untuk mengidentifikasi pola kecelakaan berdasarkan faktor lokasi, waktu, dan kondisi jalan. Selain itu, penelitian ini mengungkap keunggulan KNN dengan akurasi di atas 91% dibandingkan NB yang hanya mencapai 37.70%, serta memberikan rekomendasi spesifik untuk mitigasi kecelakaan berbasis data. Kebaruan lain terletak pada analisis spasial dan temporal yang mendalam, termasuk identifikasi hari rawan (Selasa, Kamis, Jumat, Minggu) dan lokasi rawan (Jalan Provinsi dan Jalur Mudik), yang belum dibahas secara rinci dalam literatur sebelumnya (Zhang et al., 2022; Mamlook et al., 2019).

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif untuk menganalisis pola kecelakaan lalu lintas di wilayah hukum Polresta Bogor Kota. Pendekatan ini memungkinkan pemanfaatan data numerik secara sistematis guna mengidentifikasi hubungan antara variabel-variabel yang berkontribusi terhadap tingkat kecelakaan. Selain metode statistik konvensional, penelitian ini menerapkan teknik Data Mining untuk mengekstrak pola tersembunyi dalam dataset yang besar. Algoritma *K-Nearest Neighbor* (KNN) dan *Naïve Bayes* (NB) digunakan sebagai metode utama dalam proses klasifikasi, dengan tujuan memprediksi kategori kecelakaan berdasarkan berbagai faktor yang memengaruhinya.

Data dalam penelitian ini bersumber dari rekapan kasus kecelakaan lalu lintas yang tercatat di Polresta Bogor Kota selama periode 1 Januari 2020 hingga 30 Juli 2024. Dalam rentang waktu tersebut, terdapat 502 kejadian kecelakaan yang terdokumentasi dengan variabel-variabel utama meliputi tanggal dan waktu kejadian, kondisi cahaya, kondisi cuaca, fungsi jalan, kelas jalan, bentuk geometri jalan, kondisi permukaan jalan, status jalan, dan tingkat kecelakaan. Data yang dikumpulkan melalui catatan resmi ini diolah dan dianalisis untuk mengidentifikasi faktor-faktor yang berkontribusi terhadap kecelakaan serta pola-pola yang dapat digunakan dalam perumusan kebijakan pencegahan.

Sebelum analisis dilakukan, data terlebih dahulu melalui tahap *preprocessing* untuk memastikan keakuratan dan kelayakan penggunaannya dalam model prediksi. Proses ini dimulai dengan pembersihan data, yaitu menghapus entri duplikat dan menangani nilai yang hilang menggunakan teknik imputasi atau eliminasi entri yang tidak relevan. Selanjutnya, dilakukan normalisasi dan skalasi fitur numerik guna memastikan seluruh variabel memiliki skala yang seragam dan tidak menyebabkan bias dalam analisis. Data kategori yang tidak berbentuk numerik dikonversi menggunakan metode *one-hot encoding* atau *label encoding*, sehingga dapat diproses lebih lanjut oleh algoritma *machine learning*. Setelah itu, data dibagi menjadi dua kelompok, yaitu set pelatihan (70-

80%) yang digunakan untuk membangun model dan set pengujian (20-30%) untuk mengevaluasi performa model terhadap data yang belum pernah dilihat sebelumnya.

Analisis klasifikasi kecelakaan lalu lintas dilakukan menggunakan dua pendekatan utama. *K-Nearest Neighbor* (KNN) digunakan untuk mengelompokkan kategori kecelakaan berdasarkan kedekatan karakteristik antar data, dengan nilai K optimal yang ditentukan melalui validasi silang. Model ini dievaluasi berdasarkan metrik akurasi, *precision*, *recall*, dan *F1-score* guna memastikan kinerjanya dalam mengklasifikasikan kecelakaan secara tepat. Selain itu, penelitian ini juga menerapkan model *Naïve Bayes* (NB), yang didasarkan pada prinsip probabilitas dengan asumsi independensi antar variabel. Model ini menghitung probabilitas setiap kategori kecelakaan berdasarkan distribusi fitur yang diamati dalam dataset, kemudian diuji dan dibandingkan dengan model KNN untuk menentukan metode klasifikasi yang lebih akurat.

Hasil prediksi dari kedua model dibandingkan dengan data aktual guna mengukur tingkat keakuratan klasifikasi. Model yang memiliki akurasi tertinggi dipilih sebagai metode utama dalam analisis akhir. Dari hasil tersebut, dilakukan identifikasi faktor dominan yang berkontribusi terhadap kecelakaan, yang kemudian menjadi dasar dalam perumusan rekomendasi kebijakan. Rekomendasi ini mencakup perbaikan infrastruktur jalan, peningkatan sistem penerangan, serta revisi regulasi lalu lintas yang lebih adaptif terhadap pola kecelakaan yang terdeteksi. Dengan pendekatan berbasis data ini, diharapkan dapat dihasilkan strategi yang lebih efektif dalam menekan angka kecelakaan lalu lintas di Polresta Bogor Kota.

HASIL DAN PEMBAHASAN

Penelitian ini mengevaluasi performa dua algoritma machine learning, yaitu *K-Nearest Neighbors* (KNN) dan *Naïve Bayes*, dalam mengklasifikasikan tingkat keparahan kecelakaan lalu lintas berdasarkan data historis. Evaluasi dilakukan dengan menggunakan confusion matrix, akurasi, dan Cohen's Kappa sebagai metrik utama, serta dilakukan dengan dua skema pembagian data: 80% training - 20% test dan 75% training - 25% test. Hasil penelitian ini bertujuan untuk menilai seberapa baik model dapat mengidentifikasi kecelakaan dengan tingkat keparahan sedang, ringan, dan berat serta memahami tren performa model berdasarkan skema pembagian data yang berbeda.

Evaluasi Hasil Klasifikasi Kecelakaan Lalu Lintas Menggunakan KNN

Hasil klasifikasi menggunakan KNN ditampilkan dalam Tabel 1 yang menunjukkan nilai *precision* dan *recall* untuk masing-masing kelas tingkat keparahan kecelakaan.

a. Analisis Confusion Matrix

Pada skema 80% training - 20% test, model KNN menunjukkan performa yang cukup baik dengan nilai *precision* berkisar antara 76.71% hingga 88.07% dan nilai *recall* berkisar antara 77.24% hingga 89.04%. Model memiliki kinerja

terbaik dalam mengklasifikasikan kecelakaan ringan dengan nilai precision dan recall tertinggi (80.05% dan 89.04%). Namun, model sedikit kesulitan dalam mengklasifikasikan kecelakaan dengan tingkat keparahan sedang, yang ditunjukkan oleh recall yang lebih rendah (77.24%).

Pada skema 75% training - 25% test, performa model sedikit menurun dibandingkan skema 80/20. Precision dan recall masih menunjukkan pola yang serupa, tetapi terdapat sedikit penurunan terutama pada kelas "Sedang" (precision: 73.61% dan recall: 73.10%). Hal ini menunjukkan bahwa proporsi data latih yang lebih besar (80%) menghasilkan model yang lebih stabil dalam mengklasifikasikan kecelakaan lalu lintas.

Tabel 1. Confusion Matrix Hasil Klasifikasi Kecelakaan Lalu Lintas Menggunakan Algoritma KNN

Skema Pembagian Data 80% Training Set & 20% Test Set				
	True Sedang	True Ringan	True Berat	Class precision
Pred. Sedang	112	16	18	76.71%
Pred. Ringan	18	325	63	80.05%
Pred. Berat	15	24	288	88.07%
<i>Class Recall</i>	77.24%	89.04%	78.05%	
Skema Pembagian Data 75% Training Set & 25% Test Set				
	True Sedang	True Ringan	True Berat	Class precision
Pred. Sedang	106	18	20	73.61%
Pred. Ringan	20	320	68	78.43%
Pred. Berat	19	27	281	85.93%
<i>Class Recall</i>	73.10%	87.67%	76.15%	

b. Evaluasi dengan Metrik Akurasi dan Cohen's Kappa

Sebagaimana disajikan dalam Tabel 2, akurasi model KNN cukup tinggi, mencapai 91.89% untuk skema 80/20 dan 91.20% untuk skema 75/25. Selain itu, Cohen's Kappa menunjukkan nilai yang tinggi (0.87 untuk 80/20 dan 0.86 untuk 75/25), yang menunjukkan kesepakatan prediksi model dengan data aktual yang hampir sempurna.

Hasil ini konsisten dengan penelitian sebelumnya yang menunjukkan bahwa KNN cenderung memberikan hasil yang baik dalam klasifikasi berbasis kedekatan pola data, terutama ketika dataset memiliki distribusi yang seimbang (Zhang et al., 2020). Namun, tantangan yang dihadapi adalah penurunan performa pada kelas sedang, yang dapat mengindikasikan kebutuhan untuk melakukan tuning lebih lanjut, seperti pemilihan nilai K yang optimal atau penggunaan teknik feature selection untuk meningkatkan akurasi.

Tabel 2. Evaluasi Hasil Klasifikasi Kecelakaan Lalu Lintas Menggunakan Algoritma KNN dengan Metriks Akurasi dan Cohen's Kappa

Skema <i>Splitting Data</i>	Akurasi	Cohen's Kappa
80% Training & 20% Test Set	91.89%	0.87
75% Training & 25% Test Set	91.20%	0.86

Evaluasi Hasil Klasifikasi Kecelakaan Lalu Lintas Menggunakan Naïve Bayes

a. Analisis Confusion Matrix

Sebagaimana ditunjukkan dalam Tabel 3, model Naïve Bayes menunjukkan kinerja yang jauh lebih buruk dibandingkan KNN. Precision dan recall yang rendah pada hampir semua kelas menunjukkan bahwa model ini kurang mampu mengklasifikasikan tingkat keparahan kecelakaan dengan baik.

Pada skema 80% training - 20% test, nilai recall untuk kelas Berat hanya sebesar 6.50%, yang menunjukkan bahwa model hampir tidak mampu mendeteksi kecelakaan dengan tingkat keparahan berat. Bahkan untuk kelas "Sedang" dan "Ringan", model hanya memiliki recall sebesar 52.41% dan 61.37%, yang menunjukkan bahwa lebih dari 40% data aktual gagal diklasifikasikan dengan benar.

Ketika skema diubah menjadi 75% training - 25% test, kinerja model semakin menurun dengan recall kelas "Berat" turun menjadi 5.69%. Hal ini menunjukkan bahwa Naïve Bayes tidak mampu menangani data kecelakaan lalu lintas dengan baik, kemungkinan karena asumsi independensi antar fitur yang sering kali tidak sesuai dengan data dunia nyata.

Tabel 3. Confusion Matrix Hasil Klasifikasi Kecelakaan Lalu Lintas Menggunakan Algoritma Naïve Bayes

Skema Pembagian Data 80% Training Set & 20% Test Set				
	True Sedang	True Ringan	True Berat	Class precision
Pred. Sedang	76	128	152	21.35%
Pred. Ringan	63	224	193	46.67%
Pred. Berat	6	13	24	55.81%
<i>Class Recall</i>	52.41%	61.37%	6.50%	
Skema Pembagian Data 75% Training Set & 25% Test Set				
	True Sedang	True Ringan	True Berat	Class precision
Pred. Sedang	101	223	243	17.81%
Pred. Ringan	38	130	105	47.62%
Pred. Berat	6	12	21	53.85%
<i>Class Recall</i>	69.66%	35.62%	5.69%	

b. Evaluasi dengan Metrik Akurasi dan Cohen's Kappa

Sebagaimana disajikan dalam Tabel 4, akurasi Naïve Bayes hanya mencapai 37.70% untuk skema 80/20 dan 29.14% untuk skema 75/25. Cohen's Kappa sangat rendah (0.0093 untuk 80/20 dan 0.0058 untuk 75/25), yang

menunjukkan bahwa model hampir tidak memiliki kesepakatan dengan data aktual.

Penurunan kinerja ini dapat dijelaskan oleh karakteristik data kecelakaan lalu lintas yang kompleks, yang mungkin memiliki ketergantungan antar variabel yang tinggi, seperti cuaca, kondisi jalan, dan waktu kejadian, yang tidak dapat ditangkap dengan baik oleh model berbasis probabilistik sederhana seperti Naïve Bayes (Chen et al., 2021).

Tabel 4. Evaluasi Hasil Klasifikasi Kecelakaan Lalu Lintas Menggunakan Algoritma KNN dengan Metriks Akurasi dan Cohen's Kappa

<i>Skema Splitting Data</i>	Akurasi	Cohen's Kappa
80% Training & 20% Test Set	37.70%	0.0093
75% Training & 25% Test Set	29.14%	0.0058

Perbandingan KNN dan Naïve Bayes

Dari hasil evaluasi di atas, dapat disimpulkan bahwa KNN jauh lebih unggul dibandingkan Naïve Bayes dalam mengklasifikasikan tingkat keparahan kecelakaan lalu lintas. Model KNN menunjukkan akurasi yang tinggi (>91%) dan Cohen's Kappa yang mendekati 1, yang menunjukkan bahwa prediksi model sangat sesuai dengan data aktual. Sebaliknya, Naïve Bayes memiliki performa yang buruk dengan akurasi di bawah 40% dan Cohen's Kappa yang hampir nol, yang menunjukkan bahwa model gagal menangkap pola dalam data.

Hasil ini selaras dengan penelitian sebelumnya yang menunjukkan bahwa KNN sering kali lebih efektif dalam tugas klasifikasi ketika fitur data memiliki keterkaitan yang kompleks dan tidak dapat diasumsikan independen (Kumar & Singh, 2022). Sebaliknya, Naïve Bayes hanya efektif dalam kasus di mana fitur-fitur independen dan memiliki distribusi probabilitas yang jelas (García et al., 2020), yang tampaknya tidak berlaku dalam dataset kecelakaan lalu lintas ini.

Perbedaan performa kedua algoritma juga dapat disebabkan oleh ketidakseimbangan dalam distribusi kelas. Pada data kecelakaan, kelas "Berat" memiliki jumlah data yang lebih sedikit dibandingkan kelas lainnya. KNN cenderung lebih fleksibel dalam menangani ketidakseimbangan ini, sementara Naïve Bayes sangat bergantung pada distribusi probabilitas, yang menyebabkan kesalahan prediksi yang lebih tinggi.

KESIMPULAN

Penelitian ini menunjukkan bahwa kecelakaan lalu lintas di Kota Bogor didominasi oleh kecelakaan ringan, dengan pola kejadian yang dipengaruhi oleh faktor hari (Selasa, Kamis, Jumat, dan Minggu berisiko lebih tinggi) dan status jalan (Jalan Provinsi serta Jalan Kota/Kabupaten—Jalur Mudik sebagai lokasi rawan). Algoritma KNN terbukti lebih unggul daripada Naïve Bayes dengan akurasi di atas 91%, menjadikannya model efektif untuk klasifikasi kecelakaan. Untuk penelitian selanjutnya, dapat dikembangkan dengan mengeksplorasi faktor tambahan (cuaca, manusia, kendaraan, temporal), meningkatkan model prediktif (hybrid ML, deep learning, optimasi KNN), analisis spasial (hotspot dan network analysis), evaluasi kebijakan (simulasi dampak intervensi dan studi komparatif), integrasi data real-time (IoT dan sistem peringatan dini),

studi kualitatif (wawancara dan analisis psikologis), serta validasi model di wilayah lain. Dengan pendekatan ini, temuan penelitian dapat memberikan rekomendasi lebih komprehensif untuk pencegahan kecelakaan di Kota Bogor maupun daerah lainnya.

DAFTAR PUSTAKA

- Alpaydin, E. (2020). *Introduction to Machine Learning*. MIT Press.
- Aulia, D., & Prasetyo, A. (2020). Klasifikasi Kecelakaan Lalu Lintas Menggunakan *Naïve Bayes* (NB). *Jurnal Sistem Informasi dan Teknologi*, 14(3), 95-104.
- Almasri, A., et al. (2019). *Classification of Road Traffic Accident Data Using Machine Learning Algorithms*. *IEET (2019)*, 682-687.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1986). *Classification and Regression Trees*. CRC Press.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297
- Davenport, T., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, 96(1), 108–116.
- Domingos, P., & Pazzani, M. (1997). On the relationship between *Naïve Bayes* and logistic regression. *Machine Learning*, 30(3), 195–208.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of Data Mining*. MIT Press.
- Infante, P., et al. (2022). *Comparison of Statistical and Machine-Learning Models on Road Traffic Accident Severity Classification*. *Computers (2022)*, 1-12.
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys (CSUR)*, 31(3), 264–323.
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260
- Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Informatica*, 31(3), 249–268.
- Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer.
- Kumar, A., & Sinha, R. (2019). *K-Nearest Neighbors* algorithm for classification and its applications in traffic safety. *International Journal of Data Science*, 12(4), 35–48.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- Libnao, M., et al. (2023). *Traffic Incident Prediction and Classification System Using Naïve Bayes Algorithm*. *Procedia Computer Science*, 227(2023), 316-325.
- Mamlook, R., et al. (2019). *Comparison of Machine Learning Algorithm for Predicting Traffic Accident Severity*. *JEET (2019)*, 272-276.
- McCallum, A., & Nigam, K. (1998). A comparison of event models for *Naïve Bayes* text classification. *Proceedings of the AAAI-98 Workshop on Learning for Text Categorization*, 41–48.
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.

- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis*. Wiley.
- Pratama, R., et al. (2020). Human factors and road conditions in traffic accidents: Evidence from Indonesian cities. *Transport Research Journal*, 8(3), 45–59.
- Ramadhan, N., & Gunawan, I. (2020). Klasifikasi Kejadian Kecelakaan Lalu Lintas Berdasarkan Data Pengemudi. *Jurnal Ilmiah Komputer*, 16(3), 185-200.
- Rennie, J. D., Shih, L., Teevan, J., & Karger, D. R. (2003). Tackling the poor assumptions of *Naïve Bayes* text classifiers. *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, 616–623.
- Rish, I. (2001). An empirical study of the *Naïve Bayes* classifier. *Proceedings of the 2001 ICML Workshop on Text Classification*, 41–46.
- Santosa, R., & Firdaus, M. (2019). Analisis Faktor Penyebab Kecelakaan Lalu Lintas Menggunakan *Naïve Bayes*. *Jurnal Teknologi dan Komputer*, 13(1), 45-60.
- Satria, H., & Prasetyo, A. (2021). Penggunaan *K-Nearest Neighbor* untuk Prediksi Kecelakaan Lalu Lintas di Jakarta. *Jurnal Teknologi dan Informatika*, 18(2), 120-130.
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. MIT Press.
- Undang-Undang No. 2 Tahun 2002 tentang Kepolisian Negara Republik Indonesia.
- Undang-Undang Nomor 22 Tahun 2009. Tentang Lalu Lintas dan Angkutan Jalan.
- Wijaya, A., & Gunawan, I. (2018). Perbandingan Algoritma Machine Learning untuk Klasifikasi Kecelakaan Lalu Lintas. *Seminar Nasional Teknologi Informasi*, 11(2), 123-135.
- Yusuf, M., et al. (2021). Analysis of traffic accident factors in urban areas: A case study in Jakarta. *Journal of Urban Safety Studies*, 14(2), 123–134.
- Zhang, L., et al. (2022). Application of *K-Nearest Neighbors* in traffic accident prediction. *Journal of Machine Learning Research*, 23(1), 87–102.