

Prediksi Klasifikasi Kecelakaan Lalu Lintas di Kota Surakarta dengan Menggunakan Metode Regresi Logistik Multinomial

Brilian Lucky Fauzan*, Tuti Agustin, Amirotul Musthofiah Hidayah Mahmudah

Program Studi Teknik Sipil, Fakultas Teknik, Universitas Sebelas Maret, Indonesia

Abstrak: Kecelakaan lalu lintas adalah sebuah permasalahan dalam bidang transportasi yang membutuhkan perhatian khusus. Untuk menurunkan frekuensi kecelakaan, khususnya di Kota Surakarta yang menjadi fokus penelitian ini, sangat penting untuk mengkaji klasifikasi kecelakaan lalu lintas. Tujuan dari penelitian ini adalah untuk membuat model klasifikasi kecelakaan di Kota Surakarta dengan menggunakan regresi logistik multinomial, sebuah metode dalam *machine learning*. Variabel yang digunakan dalam penelitian ini adalah klasifikasi kecelakaan, tipe tabrakan, kondisi cahaya, cuaca, fungsi jalan, kelas jalan, kondisi geometrik jalan, kondisi permukaan jalan, dan kemiringan jalan. Data kecelakaan yang digunakan dalam penelitian ini mencakup tahun 2018 hingga 2022. Berdasarkan temuan dari penelitian ini, terdapat enam variabel yang secara signifikan memengaruhi peningkatan klasifikasi kecelakaan menjadi kecelakaan berat dan empat variabel yang secara signifikan memengaruhi peningkatan klasifikasi kecelakaan menjadi kecelakaan sedang. Meskipun akurasi model mencapai 72%, model yang dihasilkan masih perlu ditingkatkan dalam hal klasifikasi pada kelas minoritas. Penelitian di masa depan diharapkan dapat menangani data yang tidak seimbang sebagaimana yang terjadi dalam penelitian ini agar dapat memperbaiki kinerja model yang dihasilkan.

Kata kunci: Keselamatan Transportasi, Klasifikasi Kecelakaan Lalu Lintas, *Machine Learning*, Regresi Logistik Multinomial

DOI:

<https://doi.org/10.47134/scbmej.v1i4.3159>

*Correspondence: Brilian Lucky Fauzan

Email: brilianlucky@student.uns.ac.id

Received: 17-08-2024

Accepted: 15-09-2024

Published: 30-10-2024



Copyright: © 2024 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Abstract: Traffic accidents are a problem in the transportation sector that requires special attention. To reduce the frequency of accidents, especially in Surakarta City, which is the focus of this research, it is important to study the classification of traffic accidents. The purpose of this study is to model the classification of accidents in Surakarta City using multinomial logistic regression, a method in machine learning. The variables used in this study are accident classification, collision type, light condition, weather condition, road function, road class, road geometric condition, road surface condition, and road slope. The accident data used in this study covers the years 2018 to 2022. Based on the findings of this study, there are six variables that significantly influence the increase in accident classification to severe accidents and four variables that significantly influence the increase in accident classification to moderate accidents. Although the accuracy of the model reached 72%, the resulting model still needs to be improved in terms of classification in minority classes. Future research is expected to handle imbalanced data as happened in this study in order to improve the performance of the resulting model.

Keywords: Machine Learning, Multinomial Logistic Regression, Traffic Accident Classification, Transportation Safety

Pendahuluan

Salah satu masalah yang perlu mendapatkan penanganan dalam bidang transportasi adalah kecelakaan lalu lintas. Menurut publikasi *Road Traffic Injuries* (2023) yang diterbitkan oleh Organisasi Kesehatan Dunia (WHO), 1,19 juta jiwa dilaporkan meninggal dunia setiap tahun akibat kecelakaan lalu lintas. Selain itu, salah satu penyebab utama kematian bagi anak-anak dan orang dewasa dengan usia antara 5 hingga 29 tahun adalah kecelakaan lalu lintas (Kadri, 2023; Marino, 2023).

Menurut data yang dikumpulkan dari Badan Pusat Statistik (BPS) (2022), terdapat 139.258 kejadian kecelakaan di Indonesia pada tahun 2022, yang mengakibatkan 28.131 korban jiwa, 13.364 korban luka berat, 160.449 korban luka ringan, dan kerugian material sebesar 280.009 juta rupiah. Kemudian, menurut data Polresta Surakarta mengenai kecelakaan di jalan raya (2023), di Kota Surakarta terjadi 1.315 kejadian kecelakaan pada tahun 2022 yang mengakibatkan 1.425 korban luka ringan dan 71 korban meninggal dunia. Oleh karena itu, kecelakaan merupakan masalah yang sangat merugikan dan memerlukan tindakan pencegahan. (Astarita, 2023; Sattar, 2023)

Analisis mengenai klasifikasi kecelakaan lalu lintas sangat penting untuk menurunkan jumlah kecelakaan dan meningkatkan keselamatan jalan (Dong, 2022; Hussein, 2022), khususnya di Kota Surakarta. Dengan mengklasifikasikan kecelakaan, faktor-faktor yang mempengaruhi kecelakaan lalu lintas dapat diidentifikasi (Chang & Wang, 2006). Analisis ini diperlukan karena sampai saat ini belum ada model yang tersedia untuk mengklasifikasikan kecelakaan lalu lintas di Kota Surakarta, dan belum diketahui faktor-faktor apa saja yang memengaruhi klasifikasi kecelakaan tersebut.

Dalam penelitian ini, *machine learning* digunakan untuk menganalisis data kecelakaan Kota Surakarta, dengan fokus khusus pada penerapan metode analisis regresi logistik multinomial. Metode ini dipilih karena mampu memodelkan variabel dependen yang memiliki lebih dari dua kategori (Abdulhafedh, 2017). Kemudian, regresi logistik multinomial mampu menginterpretasikan model yang telah dihasilkan melalui koefisien regresi untuk mengetahui faktor-faktor yang berpengaruh terhadap masing-masing tingkat keparahan kecelakaan (Shiran et al., 2021). Selain itu, regresi logistik multinomial memiliki tingkat akurasi yang cukup baik. Misalnya, studi yang dilakukan oleh Al Mamlook et al. (2019) menunjukkan bahwa regresi logistik multinomial memiliki akurasi sebesar 74,5%.

Penelitian ini menerapkan *machine learning* pada analisis data kecelakaan di Kota Surakarta, dengan penekanan khusus pada pendekatan analisis regresi logistik multinomial. Kemampuan metode ini untuk memodelkan variabel dependen dengan lebih dari dua kategori menjadi alasan pemilihan metode ini (Abdulhafedh, 2017). Selanjutnya, regresi logistik multinomial dapat mengidentifikasi variabel-variabel yang mempengaruhi setiap tingkat keparahan kecelakaan (Shiran et al., 2021). Selain itu, tingkat akurasi dari model regresi logistik multinomial yang dihasilkan relatif baik. Berdasarkan penelitian Al Mamlook et al. (2019), model yang dihasilkan dengan menggunakan metode regresi logistik multinomial memiliki tingkat akurasi sebesar 74,5%.

Hasil analisis yang dikembangkan dari penelitian ini dapat digunakan untuk merancang dan menerapkan kebijakan keselamatan lalu lintas yang lebih efektif. Tujuan dari penelitian ini adalah untuk: (1) Menghasilkan model prediksi klasifikasi kecelakaan lalu lintas di Kota Surakarta; (2) Mengidentifikasi faktor-faktor yang berpengaruh secara

signifikan terhadap klasifikasi kecelakaan; (3) Memperoleh kinerja model untuk analisis klasifikasi kecelakaan terhadap faktor-faktor yang memengaruhinya.

Metode

Penelitian ini dilakukan untuk mengklasifikasi kecelakaan lalu lintas di Kota Surakarta dan mengetahui faktor-faktor yang memengaruhi terjadinya kecelakaan lalu lintas. Data sekunder yang diperoleh berupa data kecelakaan Kota Surakarta dari tahun 2018 hingga tahun 2022. Kemudian data dianalisis dengan salah satu algoritma *machine learning* yakni regresi logistik multinomial untuk mengetahui akurasi model prediksi dan mengetahui variabel yang berpengaruh terhadap terjadinya kecelakaan lalu lintas (Aljasim, 2022; Sharma, 2022).

Penelitian ini dimulai dengan memperoleh data sekunder yang diperlukan dan menentukan variabel yang akan digunakan dalam penelitian. Variabel dependen yang digunakan dalam penelitian ini adalah klasifikasi kecelakaan. Sedangkan variabel independen yang digunakan dalam penelitian ini adalah tipe tabrakan, kondisi cahaya, cuaca, fungsi jalan, kelas jalan, karakteristik geometrik jalan, kondisi permukaan jalan, dan kemiringan jalan.

Selanjutnya, analisis dilakukan dengan salah satu metode *machine learning* berupa regresi logistik multinomial pada bahasa pemrograman Python (Sumathi et al., 2022). Tahapan pertama pada analisis data ini adalah melakukan *data pre-processing* dengan cara memisahkan variabel dependen dan independen; memisahkan data menjadi data *training* sejumlah 70% dan data *testing* sejumlah 30%; melakukan *random oversampling* pada kelas minoritas dengan rasio kelas minoritas terhadap kelas mayoritas sebesar 75% (Attenberg et al., 2013; Hasanin et al., 2019; Wongvorachan et al., 2023); serta memberikan regularisasi pada model regresi logistik multinomial (Kuhn & Johnson, 2013). Kemudian, data *training* dilatih dengan regresi logistik multinomial. Setelah itu, hasil analisis regresi logistik multinomial dapat ditampilkan, di mana hasilnya meliputi model regresi logistik multinomial berdasarkan koefisien yang dihasilkan; uji simultan yang digunakan untuk mengetahui apakah ada setidaknya satu variabel independen yang berpengaruh signifikan terhadap klasifikasi kecelakaan (Abdulhafedh, 2017); dan uji parsial yang digunakan untuk mengetahui apakah masing-masing variabel independen berpengaruh signifikan terhadap klasifikasi kecelakaan (Abdulhafedh, 2017). Tahapan terakhir dari penelitian ini adalah mengevaluasi data dengan membuat *confusion matrix* (Sammut & Webb, 2011).

Model yang dihasilkan dari regresi logistik multinomial terdiri dari dua fungsi, yakni fungsi logit dan fungsi probabilitas setiap kelasnya (Abdel-Salam, 2022; Yang, 2022). Pada kasus di mana variabel dependen memiliki tiga kelas, fungsi logit yang terbentuk untuk

kelas dua dan tiga dengan kelas satu sebagai kelas referensi adalah sebagai berikut (Hosmer & Lemeshow, 2000).

$$g_2(x) = \ln \left[\frac{P(Y = 2|x)}{P(Y = 1|x)} \right] = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \dots + \beta_{2p}x_p = x'\beta_2 \quad (1)$$

$$g_3(x) = \ln \left[\frac{P(Y = 3|x)}{P(Y = 1|x)} \right] = \beta_{30} + \beta_{31}x_1 + \beta_{32}x_2 + \dots + \beta_{3p}x_p = x'\beta_3 \quad (2)$$

Lalu, persamaan probabilitas kondisional setiap kategorinya adalah sebagai berikut (Hosmer & Lemeshow, 2000)

$$P(Y = 1|x) = \frac{1}{1+e^{g_2(x)}+e^{g_3(x)}} \quad (3)$$

$$P(Y = 2|x) = \frac{e^{g_2(x)}}{1+e^{g_2(x)}+e^{g_3(x)}} \quad (4)$$

$$P(Y = 3|x) = \frac{e^{g_3(x)}}{1+e^{g_2(x)}+e^{g_3(x)}} \quad (5)$$

Hasil dan Pembahasan

Rincian model regresi yang telah dilatih ditampilkan pada **Tabel 1** dan **Tabel 2**.

Pseudo R-square = 0,5439

Log-Likelihood Ratio p-value = 0,000

Tabel 1. Rincian Hasil Regresi Logistik Multinomial pada Kecelakaan Sedang

Variabel	Koefisien	Standard Error	Z	P > Z
Konstanta	-0,7764	9,704	-0,080	0,936
Tipe tabrakan	7,6519	1,302	5,875	0,000
Kondisi cahaya	-2,1318	5,563	-0,383	0,702
Cuaca	-1,9801	0,812	-2,440	0,015
Fungsi jalan	-6,5466	0,364	-17,978	0,000
Kelas jalan	-14,1089	1,262	-11,183	0,000
Kondisi geometrik jalan	-5,2143	1,241	-4,202	0,000
Kondisi permukaan jalan	-3,6907	0,692	-5,332	0,000
Kemiringan jalan	-0,9341	4,346	-0,215	0,830

Tabel 2. Rincian Hasil Regresi Logistik Multinomial pada Kecelakaan Berat

Variabel	Koefisien	Standard Error	Z	P > Z
Konstanta	-0,5515	0,359	-1,538	0,124
Tipe tabrakan	-0,2409	0,016	-14,940	0,000
Kondisi cahaya	0,4788	0,124	3,856	0,000
Cuaca	0,1736	0,055	3,157	0,002
Fungsi jalan	0,3537	0,049	7,259	0,000
Kelas jalan	-0,0552	0,048	-1,152	0,249
Kondisi geometrik jalan	0,0245	0,019	1,300	0,194
Kondisi permukaan jalan	0,0804	0,105	0,768	0,442
Kemiringan jalan	-0,5252	0,272	-1,929	0,054

Dari hasil di atas, diketahui bahwa *pseudo r-square* yang dihasilkan bernilai 0,5439. Ini artinya model yang dihasilkan mampu menjelaskan sekitar 54,39% variasi dalam data, sementara sisanya dijelaskan oleh variabel lain di luar model. Selanjutnya, nilai LLR *p-value* yang dihasilkan menunjukkan nilai 0,000 (di bawah 0,05). Hal ini menunjukkan bahwa model yang dihasilkan memiliki minimal satu variabel berpengaruh signifikan terhadap model.

Selanjutnya, dari uji parsial yang dihasilkan, dapat disimpulkan bahwa variabel independen yang berpengaruh signifikan secara statistik terhadap peluang terjadinya kecelakaan sedang dibandingkan dengan kecelakaan ringan adalah tipe tabrakan, cuaca, fungsi jalan, kelas jalan, kondisi geometrik jalan, dan kondisi permukaan jalan. Sedangkan, pada peluang terjadinya kecelakaan berat dibandingkan dengan kecelakaan ringan, variabel independen yang berpengaruh secara signifikan adalah tipe tabrakan, kondisi cahaya, cuaca, dan fungsi jalan. Variabel independen yang berpengaruh secara signifikan tersebut ditentukan dari nilai *p-value*, di mana jika nilainya berada di bawah 0,05, maka dapat disimpulkan variabel tersebut berpengaruh signifikan terhadap peluang terjadinya masing-masing klasifikasi kecelakaan.

Kemudian, berdasarkan nilai konstanta yang dihasilkan dari tabel, persamaan logit yang terbentuk adalah sebagai berikut.

$$g_2(x) = \ln \left[\frac{P(Y = 2|x)}{P(Y = 1|x)} \right] = -0,7764 + 7,6519x_1 - 2,1318x_2 - 1,9801x_3 - 6,5466x_4 - 14,1089x_5 - 5,2143x_6 - 3,6907x_7 - 0,9341x_8 \quad (6)$$

$$g_3(x) = \ln \left[\frac{P(Y = 3|x)}{P(Y = 1|x)} \right] = -0,5515 - 0,2409x_1 + 0,4788x_2 + 0,1736x_3 + 0,3537x_4 - 0,0552x_5 + 0,0245x_6 + 0,0804x_7 - 0,5252x_8 \quad (7)$$

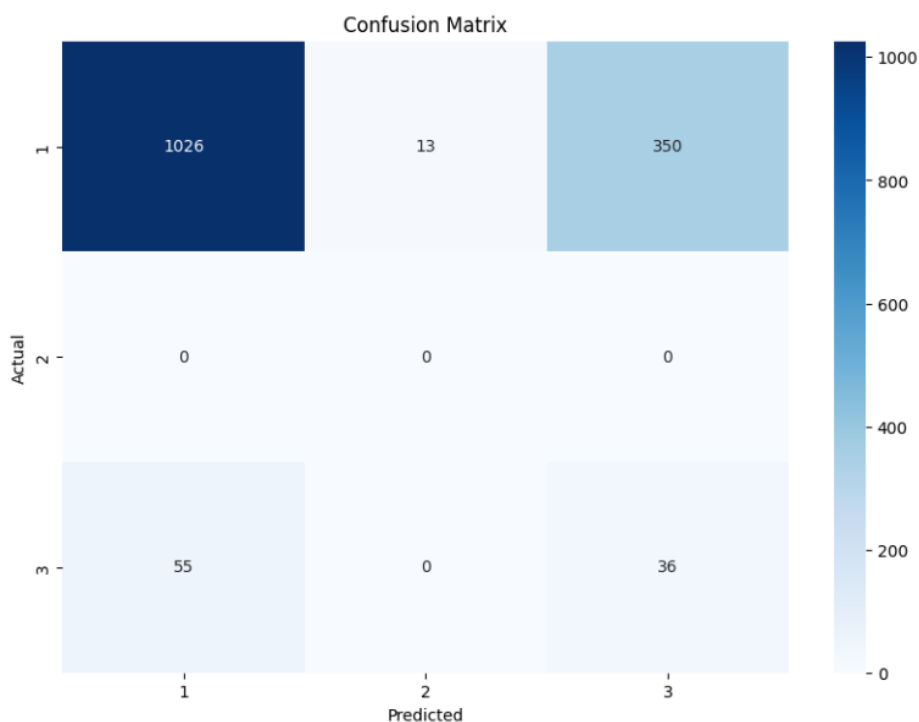
Sedangkan, persamaan probabilitas yang dihasilkan dari masing-masing klasifikasi adalah sebagai berikut.

$$P(Y = 1|x) = \frac{1}{1 + e^{g_2(x)} + e^{g_3(x)}} \quad (8)$$

$$P(Y = 2|x) = \frac{e^{g_2(x)}}{1 + e^{g_2(x)} + e^{g_3(x)}} \quad (9)$$

$$P(Y = 3|x) = \frac{e^{g_3(x)}}{1 + e^{g_2(x)} + e^{g_3(x)}} \quad (10)$$

Selanjutnya, *confusion matrix* ditampilkan untuk menyajikan kinerja model secara lebih mendalam.



Gambar 1. Confusion Matrix

Dari pemetaan *confusion matrix* yang ditampilkan pada **Gambar 1**, terlihat bahwa kelas kecelakaan rendah (kelas 1) mampu terprediksi dengan baik, dengan prediksi yang tepat sejumlah 1.026. Hanya saja, baik kecelakaan sedang (kelas 2) dan kecelakaan berat (kelas 3) tidak mampu terprediksi sebaik kecelakaan ringan. Model yang dihasilkan sama sekali tidak mampu meramalkan kecelakaan sedang, yang merupakan bagian terburuk dari penelitian ini.

Setelah *confusion matrix* ditampilkan, kinerja model dapat diperlihatkan melalui *precision*, *recall*, dan *F1 score*.

Tabel 3. Kinerja Model

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>	<i>Support</i>
1 (rendah)	0,95	0,74	0,83	1389
2 (sedang)	0,00	0,00	0,00	0
3 (tinggi)	0,09	0,40	0,15	91
Akurasi = 0,72				1480

Kesimpulan yang dapat ditarik dari kinerja model ini adalah akurasi model yang dihasilkan sebesar 72% menunjukkan bahwa model akurat secara keseluruhan. Kemudian, pada kecelakaan rendah, terdapat beberapa sampel aktual yang tidak terdeteksi sebagai kecelakaan rendah meskipun hampir seluruh sampel prediksi terprediksi dengan benar (*precision* = 0,95; *recall* = 0,74; *F1 score* = 0,82). Pada kecelakaan sedang, kinerja dari model tidak dapat dievaluasi karena tidak memiliki data (*precision* = 0,00; *recall* = 0,00; *F1 score* = 0,00). Selanjutnya, pada kecelakaan berat, banyak sampel yang salah diprediksi sebagai kecelakaan ringan, meskipun data aktualnya mampu mendeteksi sebagian kasus kecelakaan berat (*precision* = 0,09; *recall* = 0,40; *F1 score* = 0,15).

Simpulan

Model yang dihasilkan dari analisis regresi logistik multinomial terdiri dari dua persamaan utilitas yang ditampilkan pada persamaan berikut.

$$g_2(x) = \ln \left[\frac{P(Y = 2|x)}{P(Y = 1|x)} \right] = -0,7764 + 7,6519x_1 - 2,1318x_2 - 1,9801x_3 - 6,5466x_4 - 14,1089x_5 - 5,2143x_6 - 3,6907x_7 - 0,9341x_8 \quad (11)$$

$$g_3(x) = \ln \left[\frac{P(Y = 3|x)}{P(Y = 1|x)} \right] = -0,5515 - 0,2409x_1 + 0,4788x_2 + 0,1736x_3 + 0,3537x_4 - 0,0552x_5 + 0,0245x_6 + 0,0804x_7 - 0,5252x_8 \quad (12)$$

Hasil dari penelitian ini mengungkapkan bahwa terdapat enam variabel yang berpengaruh signifikan terhadap peningkatan klasifikasi menjadi kecelakaan sedang, yakni tipe tabrakan, cuaca, fungsi jalan, kelas jalan, kondisi geometrik, dan kondisi permukaan jalan. Sedangkan terdapat lima variabel yang berpengaruh signifikan terhadap peningkatan klasifikasi menjadi kecelakaan berat, yakni tipe tabrakan, kondisi cahaya, cuaca, dan fungsi jalan.

Akurasi yang dihasilkan dari model adalah 72%, tetapi klasifikasi pada kelas minoritas masih perlu ditingkatkan. Hal ini dikarenakan mayoritas sampel *testing* yang diberikan termasuk dalam kelas kecelakaan rendah, sangat sedikit yang masuk ke dalam kecelakaan berat, bahkan tidak ada yang masuk ke dalam kecelakaan sedang.

Penelitian selanjutnya perlu melakukan peningkatan kualitas dan kuantitas data yang akan digunakan dalam membentuk model. Sebagai tambahan, untuk meningkatkan kualitas data dapat digunakan metode *resampling* yang lebih efisien seperti SMOTE dan juga dapat menggunakan metode *machine learning* lainnya.

Daftar Pustaka

- Abdulhafedh, A. (2017). Incorporating the Multinomial Logistic Regression in Vehicle Crash Severity Modeling: A Detailed Overview. *Journal of Transportation Technologies*, 07(03), 279–303. <https://doi.org/10.4236/jtts.2017.73019>
- Abdel-Salam, R. (2022). RIECNN: real-time image enhanced CNN for traffic sign recognition. *Neural Computing and Applications*, 34(8), 6085–6096. <https://doi.org/10.1007/s00521-021-06762-5>
- Aljasim, M. (2022). E2DR: A Deep Learning Ensemble-Based Driver Distraction Detection with Recommendations Model. *Sensors*, 22(5). <https://doi.org/10.3390/s22051858>
- Al Mamlook, R. E., Kwayu, K. M., Alkasisbeh, M. R., & Frefer, A. A. (2019). Comparison of Machine Learning Algorithms for Predicting Traffic Accident Severity. *IEEE Jordan International Conference on Electrical Engineering and Information Technology (JEEIT)*, 272–276.
- Astarita, V. (2023). Developing new hybrid grey wolf optimization-based artificial neural network for predicting road crash severity. *Transportation Engineering*, 12. <https://doi.org/10.1016/j.treng.2023.100164>
- Attenberg, J., Batuwita, R., Chawla, N. V., Chen, Sheng, Ertekin, Ş., Japkowicz, N., He, H., Hoens, T. R., Liu, X.-Y., Palade, V., Weiss, G. M., & Zhou, Z.-H. (2013). Imbalanced

- Learning. In H. He & Y. Ma (Eds.), *Imbalanced Learning*. Wiley-IEEE Press. <https://doi.org/10.1002/9781118646106>
- Chang, L. Y., & Wang, H. W. (2006). Analysis of Traffic Injury Severity: An Application of Non-parametric Classification Tree Techniques. *Accident Analysis and Prevention*, 38(5), 1019–1027. <https://doi.org/10.1016/j.aap.2006.04.009>
- Data Kecelakaan Surakarta (2018-2022). (2023). Polres Kota Surakarta.
- Dong, S. (2022). Predicting and Analyzing Road Traffic Injury Severity Using Boosting-Based Ensemble Learning Models with SHAPley Additive exPlanations. *International Journal of Environmental Research and Public Health*, 19(5). <https://doi.org/10.3390/ijerph19052925>
- Hasanin, T., Khoshgoftaar, T. M., Leevy, J. L., & Bauder, R. A. (2019). Severely imbalanced Big Data challenges: investigating data sampling approaches. *Journal of Big Data*, 6(1). <https://doi.org/10.1186/s40537-019-0274-4>
- Hussein, N. H. (2022). A Comprehensive Survey on Vehicular Networking: Communications, Applications, Challenges, and Upcoming Research Directions. *IEEE Access*, 10, 86127–86180. <https://doi.org/10.1109/ACCESS.2022.3198656>
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression* (2nd ed.). John Wiley & Sons.
- Jumlah Kecelakaan, Korban Mati, Luka Berat, Luka Ringan, dan Kerugian Materi, 2022. (2022). Badan Pusat Statistik. <https://www.bps.go.id/id/statistics-table/2/NTEzIzI=/jumlah-kecelakaan--korban-mati--luka-berat--luka-ringan--dan-kerugian-materi.html>
- Kadri, N. (2023). New LSTM Deep Learning Algorithm for Driving Behavior Classification. *Cybernetics and Systems*, 54(4), 387–405. <https://doi.org/10.1080/01969722.2022.2059133>
- Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer Science & Business Media. <https://doi.org/10.1007/978-1-4614-6849-3>
- Marino, M. (2023). New frontiers in the risk assessment of ship collision. *Ocean Engineering*, 274. <https://doi.org/10.1016/j.oceaneng.2023.113999>
- Road Traffic Injuries. (2023, December 13). World Health Organization. <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>
- Sammur, C., & Webb, G. I. (2011). *Encyclopedia of Machine Learning*. Springer. <https://doi.org/10.1007/978-0-387-30164-8>
- Sattar, K. (2023). Transparent deep machine learning framework for predicting traffic crash severity. *Neural Computing and Applications*, 35(2), 1535–1547. <https://doi.org/10.1007/s00521-022-07769-2>
- Sharma, A. (2022). Machine Learning Based Misbehaviour Detection in VANET Using Consecutive BSM Approach. *IEEE Open Journal of Vehicular Technology*, 3, 1–14. <https://doi.org/10.1109/OJVT.2021.3138354>
- Shiran, G., Imaninasab, R., & Khayamim, R. (2021). Crash Severity Analysis of Highways Based on Multinomial Logistic Regression Model, Decision Tree Techniques and Artificial Neural Network: A Modeling Comparison. *Sustainability* (Switzerland), 13(10), 1–23. <https://doi.org/10.3390/su13105670>

-
- Sumathi, S., Rajappa, S. V, Kumar, L. A., & Paneerselvam, S. (2022). *Machine Learning for Decision Sciences with Case Studies in Python* (1st ed.). CRC Press.
- Wongvorachan, T., He, S., & Bulut, O. (2023). A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining. *Information (Switzerland)*, 14(1). <https://doi.org/10.3390/info14010054>
- Yang, K. (2022). Classification and Evaluation of Driving Behavior Safety Levels: A Driving Simulation Study. *IEEE Open Journal of Intelligent Transportation Systems*, 3, 111–125. <https://doi.org/10.1109/OJITS.2022.3149474>