

PREDIKSI JUMLAH KECELAKAAN LALU LINTAS DI SULAWESI TENGGARA MENGGUNAKAN METODE ALGORITMA REGRESI LINEAR

Zahra Vebryan Maharani, Anggun Lestari Arman Syah, Wa Ode Azzahra Astiani Hasiba

Teknik Informatika, Universitas Halu Oleo
Jalan H.E.A. Mokodompit Kendari, Indonesia
vebryanzahra@gmail.com

ABSTRAK

Kecelakaan lalu lintas merupakan salah satu masalah utama di Provinsi Sulawesi Tenggara, Indonesia, yang menyebabkan kerugian materiil dan non-materil yang signifikan, termasuk korban jiwa dan cedera. Penelitian ini bertujuan untuk menerapkan algoritma regresi linear dalam memprediksi jumlah kecelakaan tertinggi di berbagai Polres di Sulawesi Tenggara berdasarkan data kecelakaan dari tahun 2023 dan 2024. Variabel yang digunakan meliputi jumlah korban meninggal dunia, luka berat, dan luka ringan. Hasil penelitian menunjukkan bahwa model regresi linear mampu memprediksi jumlah kecelakaan dengan akurasi yang sangat tinggi, dengan nilai *Mean Squared Error* (MSE) mendekati nol dan nilai *R-squared* (R^2) sebesar 1.0. Penelitian ini memberikan kontribusi praktis bagi pihak kepolisian dan pemerintah daerah dalam merumuskan kebijakan dan tindakan preventif yang lebih efektif untuk meningkatkan keselamatan lalu lintas di Sulawesi Tenggara.

Kata kunci : Kecelakaan Lalu Lintas, Prediksi Jumlah Kecelakaan, Regresi Linear, Sulawesi Tenggara

1. PENDAHULUAN

Kecelakaan lalu lintas merupakan salah satu masalah utama yang dihadapi oleh banyak daerah di Indonesia, termasuk di Provinsi Sulawesi Tenggara. Setiap tahun, ribuan kasus kecelakaan terjadi, menyebabkan kerugian materiil dan non-materil yang signifikan, termasuk korban jiwa dan cedera. Pengelolaan dan analisis data kecelakaan lalu lintas menjadi krusial untuk memahami pola kejadian dan mengembangkan strategi yang efektif untuk mengurangi jumlah kecelakaan serta meningkatkan keselamatan di jalan raya [1].

Berdasarkan data yang diperoleh, kecelakaan lalu lintas di Sulawesi Tenggara menunjukkan variasi yang cukup besar antara berbagai wilayah dan satuan kerja kepolisian (Polres). Untuk dapat mengatasi masalah ini dengan lebih efektif, diperlukan analisis yang mendalam untuk mengidentifikasi pola dan faktor-faktor yang mempengaruhi terjadinya kecelakaan lalu lintas di berbagai wilayah tersebut.

Data *mining* adalah teknik yang efektif untuk menggali pola tersembunyi dalam data yang besar dan kompleks [2]. Salah satu metode data *mining* yang populer adalah regresi linear, yang dikenal karena kesederhanaan dan kemampuannya dalam memprediksi hubungan antara variabel [3]. Regresi linear dapat digunakan untuk mengidentifikasi dan memodelkan hubungan antara variabel-variabel kecelakaan lalu lintas seperti jumlah korban meninggal dunia, luka berat, dan luka ringan dengan total insiden kecelakaan. Dengan menggunakan metode ini, diharapkan dapat diidentifikasi Polres dengan tingkat korban kecelakaan tertinggi, sehingga dapat dijadikan dasar dalam merumuskan kebijakan dan tindakan preventif yang lebih tepat sasaran.

Penelitian ini bertujuan untuk menerapkan algoritma regresi linear dalam memprediksi Polres dengan tingkat korban kecelakaan tertinggi di

Sulawesi Tenggara berdasarkan data yang dikumpulkan dari berbagai Polres. Variabel yang digunakan meliputi tahun kejadian, nama Polres, jumlah korban meninggal dunia, luka berat, dan luka ringan. Melalui analisis ini, diharapkan dapat diperoleh gambaran yang lebih jelas tentang distribusi dan karakteristik kecelakaan lalu lintas di Sulawesi Tenggara, serta faktor-faktor yang berkontribusi terhadap tingkat keparahan kecelakaan.

Dengan demikian, penelitian ini tidak hanya berkontribusi dalam aspek akademis, tetapi juga memberikan manfaat praktis bagi pihak kepolisian dan pemerintah daerah dalam meningkatkan keselamatan lalu lintas di Sulawesi Tenggara. Analisis yang dihasilkan dapat menjadi dasar bagi pengambilan keputusan yang lebih efektif dalam upaya pencegahan dan penanganan kecelakaan lalu lintas di masa mendatang.

2. TINJAUAN PUSTAKA

Menurut Undang-Undang No. 22 Tahun 2009, kecelakaan di jalan adalah kejadian tak terduga dan tidak disengaja yang melibatkan kendaraan, baik dengan maupun tanpa melibatkan pengguna jalan lainnya, dan dapat menyebabkan kematian atau kerugian materiil. Berbeda dengan pandangan Kartika, kecelakaan lalu lintas merupakan peristiwa kompleks yang sulit diprediksi kapan dan di mana akan terjadi. Dampak kecelakaan tidak hanya berupa trauma, cedera, atau cacat, tetapi juga dapat berujung pada kematian. Selain itu, kecelakaan lalu lintas sulit ditekan dan cenderung meningkat seiring dengan bertambahnya panjang jalan dan volume pergerakan kendaraan [4].

Secara umum, kecelakaan lalu lintas didefinisikan sebagai kegagalan dalam interaksi antara sistem jalan, kendaraan, dan pengguna jalan. Dari perspektif makroekonomi, kecelakaan menimbulkan

inefisiensi dalam penyelenggaraan transportasi karena mengurangi kuantitas dan kualitas orang maupun barang yang diangkut, serta meningkatkan biaya operasional secara keseluruhan [5].

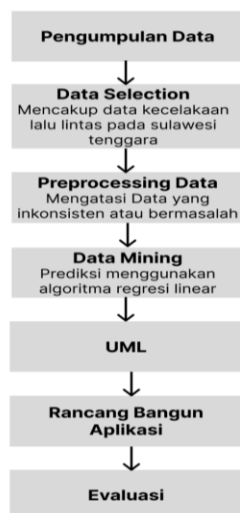
Regresi linear sederhana adalah metode statistik yang berfungsi untuk menguji sejauh mana hubungan sebab akibat antar variabel faktor penyebab (X) terhadap variabel akibatnya. Faktor penyebab pada umumnya dilambangkan dengan x atau disebut juga dengan prediktor, sedangkan variabel akibat dilambangkan dengan y atau disebut juga dengan respon [6].

Selain itu, Regresi Linear merupakan metode prediksi yang mempelajari ikatan antara variabel dependent dan variabel independent. Algoritma regresi linear berfungsi untuk mengukur kekuatan hubungan dua variabel, dan menunjukkan alur hubungan antara variabel dependent dengan variabel independent [7]. Kemudian regresi linear mampu mempelajari hubungan antara satu variabel tak bebas dan satu variabel bebas [8].

Analisis regresi linear digunakan untuk mendapatkan hubungan matematis dalam bentuk suatu persamaan antara variabel tak bebas dengan variabel bebas Tunggal [9].

3. METODE PENELITIAN

Metodologi penelitian yang digunakan adalah *Knowledge Discovery in Database* (KDD). Tahapan pada metodologi KDD meliputi pemilihan data, prapemrosesan data, transformasi data, data mining, UML, perancangan aplikasi, dan evaluasi.



Gambar 1. Alur Penelitian

3.1. Pengumpulan Data

Data mengenai kecelakaan lalu lintas dari berbagai Polres di Sulawesi Tenggara dikumpulkan melalui Polda Sulawesi Tenggara. Data tersebut meliputi jumlah korban meninggal dunia, luka berat, dan luka ringan akibat kecelakaan lalu lintas.

3.2. Data Selection

Pada tahap ini, dilakukan seleksi data. Data yang telah diperoleh dari Polda akan dipilih atribut-atributnya untuk diproses lebih lanjut dalam tahap *preprocessing*.

3.3. Data Preprocessing

Tujuan dari *preprocessing* data adalah mengubah data mentah menjadi data berkualitas, sehingga siap untuk diolah pada tahap berikutnya [10]. Tahap ini melibatkan pembersihan data dari masalah atau inkonsistensi, seperti data yang mengandung *noise* atau *error* [11].

3.4. Transformation

Transformasi adalah tahap di mana data diubah menjadi format yang sesuai dengan kebutuhan [12]. Teknik yang digunakan adalah *discretization*, yang bertujuan mengubah atribut numerik menjadi atribut kategorikal, sehingga data dapat dikelompokkan dalam beberapa kategori [13].

3.5. Data Mining Regresi Linear

Pada tahapan ini data sudah siap untuk proses deteksi. Metode yang digunakan adalah metode algoritma Regresi Linear. Metode algoritma Regresi linear ini adalah salah satu metode statistik paling sederhana yang digunakan untuk memprediksi nilai *output* berdasarkan nilai *input* [14]. Model ini berusaha menemukan hubungan linear antara variabel *input* (*predictor*) dan *output* (*respons*).

Konsep dasar dari regresi linear mencoba memodelkan hubungan antara dua variabel dengan memasang garis lurus ke dalam data [15]. Persamaan dari regresi linear sederhana adalah:

$$y = \beta_0 + \beta_1 x + \epsilon \quad (1)$$

Keterangan:

y : variabel respons (dependent).

x : variabel prediktor (independent).

β_0 : intersep (nilai y ketika $x=0$).

β_1 : koefisien kemiringan garis (mengindikasikan seberapa besar perubahan y untuk setiap unit perubahan x).

ϵ : error (deviasi antara nilai observasi dan prediksi).

Setelah pembentukan model, dilanjutkan dengan proses training model yang diawali dengan pembagian dataset menjadi dua bagian, yaitu 80% data training untuk melatih model dan 20% data testing untuk menguji akurasi model.

Implementasi prediksi dilakukan setelah model selesai dilatih. Pada tahap ini, model regresi linear yang telah dibentuk digunakan untuk memprediksi jumlah korban kecelakaan dengan memasukkan nilai-nilai variabel independen ke dalam persamaan. Hasil prediksi kemudian dianalisis untuk mengidentifikasi Polres dengan tingkat korban kecelakaan tertinggi dengan cara membandingkan dan mengurutkan nilai prediksi jumlah korban antar Polres. Selanjutnya

dilakukan analisis pola kecelakaan berdasarkan hasil prediksi, yang mencakup pengamatan tren kecelakaan di setiap Polres, identifikasi faktor-faktor yang mempengaruhi tingginya angka korban, serta analisis hubungan antar variabel yang diteliti.

3.6. Unified Modelling Language

Digunakan untuk memvisualisasikan dan merancang sistem perangkat lunak dengan menggunakan *use case* dan *activity diagram*.

3.7. Rancang Bangun Aplikasi

Berdasarkan data yang telah diproses, akan dirancang aplikasi untuk mempermudah analisis dan prediksi kecelakaan lalu lintas. Aplikasi ini akan mengimplementasikan algoritma regresi linear untuk memprediksi jumlah korban berdasarkan variabel-variabel yang ada. Dengan antarmuka yang *user-friendly*, aplikasi ini diharapkan dapat memberikan informasi yang akurat dan membantu dalam pengambilan keputusan terkait keselamatan lalu lintas.

3.8. Evaluasi

Pada tahap ini dilakukan evaluasi untuk mengukur validitas model prediksi dengan menggunakan dua metrik utama yaitu *Mean Squared Error* (MSE) untuk mengukur rata-rata kuadrat perbedaan antara nilai prediksi dan nilai aktual. Semakin kecil nilai MSE, semakin baik model dalam memprediksi data dan *R-squared* (R^2) untuk mengukur seberapa baik variabel *input* menjelaskan variasi dalam variabel *output*. Nilai R^2 berkisar antara 0 dan 1, dengan nilai mendekati 1 menunjukkan model yang sangat baik.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Tahapan pengumpulan data ini dilakukan untuk mengumpulkan data kecelakaan lalu lintas diberbagai Polres Sulawesi Tenggara yang diperoleh dari Polda Sulawesi Tenggara.

Tabel 1. Data Kecelakaan Lalu Lintas

Tahun	Nama Polres	Meninggal Dunia	Luka Berat	Luka Ringan
2023	Polresta Kendari	57	36	445
2023	Polres Kolaka	32	9	155
2023	Polres BauBau	23	1	287
2023	Polres Muna	38	0	128
2023	Polres Konawe	66	20	441
2024	Polresta Kendari	14	0	169
2024	Polres Kolaka	18	2	73
2024	Polres BauBau	7	0	74
2024	Polres Muna	7	0	24
2024	Polres Konawe	14	4	126
...	...	...	...	...

Setelah data didapatkan, selanjutnya mendeskripsikan data tersebut yang bertujuan untuk memahami informasi lebih lanjut mengenai atribut dari data yang akan diolah dalam memprediksi kecelakaan lalu lintas.

4.2. Data Selection

Data kecelakaan lalu lintas yang didapatkan di Polda Sultra selanjutnya memasuki tahap *data selection*. Tahapan *data selection* ini bertujuan memilih fitur yang dianggap sebagai fitur yang berpengaruh terhadap prediksi kecelakaan lalu lintas.

Tabel 2. Fitur

No	Fitur
1	Meninggal Dunia
2	Luka Berat
3	Luka Ringan

4.3. Data Preprocessing

a. Data Collection

Data yang digunakan dalam penelitian ini merupakan data kecelakaan lalu lintas yang terjadi di beberapa Polres di Sulawesi Tenggara pada tahun 2023 dan 2024. Data tersebut mencakup jumlah korban meninggal dunia, luka berat, dan luka ringan.

b. Feature Engineering

Dari data yang telah dikumpulkan, langkah selanjutnya adalah melakukan rekayasa fitur untuk menghitung total insiden kecelakaan di setiap Polres. Total insiden dihitung dengan menjumlahkan jumlah korban meninggal dunia, luka berat, dan luka ringan. Dengan rumus:

$$\text{total_incidents} = \text{meninggal_dunia} + \text{luka_berat} + \text{luka_ringan}$$

Hasil dari perhitungan ini ditambahkan sebagai kolom baru dataset.

c. Data Spilling

Dataset dibagi menjadi data pelatihan (tahun 2023) dan data pengujian (tahun 2024). Pembagian ini dilakukan untuk memastikan model prediksi yang dibangun dapat diuji dengan data yang tidak digunakan dalam proses pelatihan.

4.4. Data Mining

Dalam penelitian ini, dilakukan pembentukan model prediksi jumlah kecelakaan lalu lintas di Sulawesi Tenggara menggunakan regresi linier. Langkah-langkah yang dilakukan dalam pembentukan model adalah sebagai berikut:

a. Pemisahan Data

Data yang telah di preprocessing dibagi menjadi dua subset: data pelatihan (tahun 2023) dan data pengujian (tahun 2024). Pemisahan ini bertujuan untuk melatih model dengan data historis dan menguji performa model dengan data yang belum pernah dilihat oleh model sebelumnya.

b. Pemilihan Fitur

Fitur-fitur yang digunakan dalam model adalah: Jumlah korban meninggal dunia (meninggal_dunia), Jumlah korban luka berat (luka_berat) dan Jumlah korban luka ringan (luka_ringan).

Fitur-fitur ini dipilih karena secara langsung berhubungan dengan total insiden kecelakaan lalu lintas.

c. Pelatihan Model

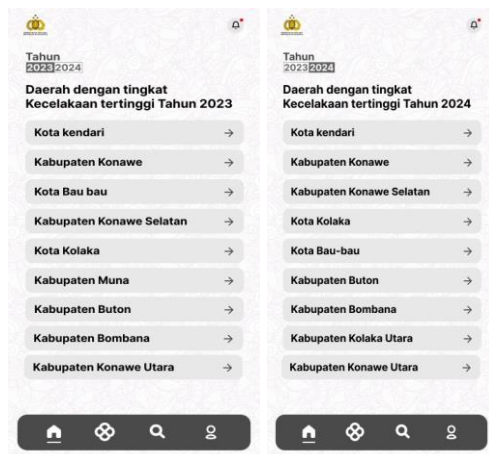
Model regresi linier digunakan untuk memprediksi total insiden kecelakaan lalu lintas. Data pelatihan yang terdiri dari tahun 2023 digunakan untuk melatih model ini. Model dilatih dengan menggunakan fitur-fitur yang telah dipilih untuk memprediksi target total *incidents*.

d. Prediksi

Setelah model dilatih, dilakukan prediksi terhadap data pengujian (tahun 2024) untuk mengevaluasi performa model. Hasil prediksi dibandingkan dengan data aktual untuk mengukur seberapa baik model dapat memprediksi total insiden kecelakaan.

4.5. Rancang Bangun Aplikasi

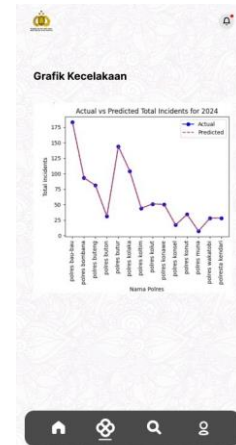
Berikut ini adalah implementasi tampilan aplikasi *mobile*.



Gambar 1. Halaman Home



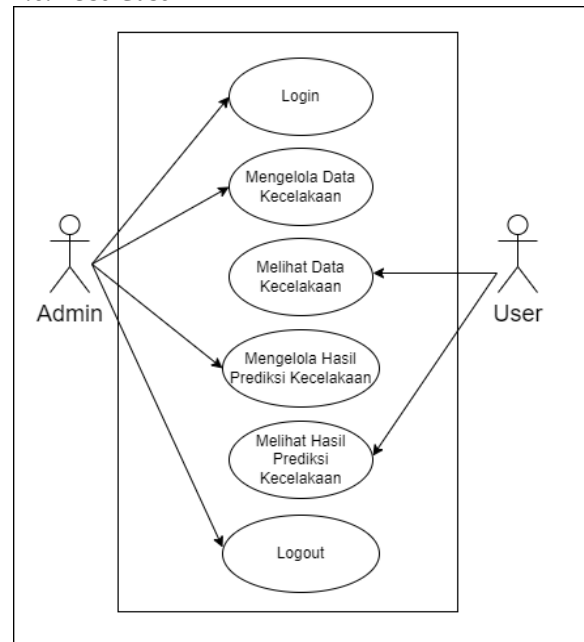
Gambar 2. Detail Kecelakaan



Gambar 3. Grafik Kecelakaan

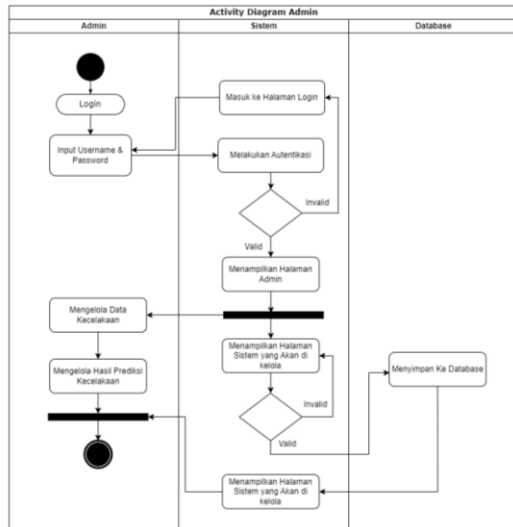
Implementasi tampilan aplikasi *mobile* ini mencakup beberapa halaman utama untuk memberikan pengalaman yang informatif dan intuitif bagi pengguna. Halaman *Home* (Gambar 1) yang menampilkan informasi penting untuk tahun 2023 dengan desain yang diperbarui. Halaman *Details* Kecelakaan (Gambar 2) memberikan rincian lengkap tentang kecelakaan tertentu, termasuk lokasi, waktu, jumlah korban, dan penyebab kecelakaan. Selain itu, Halaman Grafik Kecelakaan (Gambar 3) menampilkan visualisasi data mengenai tren dan pola kecelakaan, membantu pengguna memahami distribusi dan perubahan jumlah kecelakaan dari waktu ke waktu.

4.6. Use Case

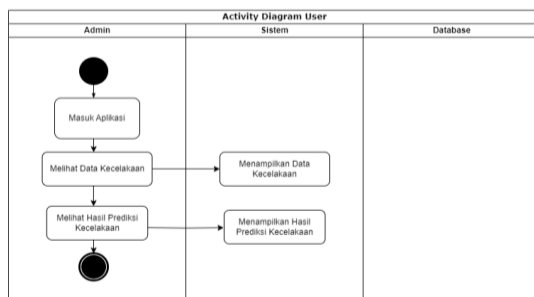


Gambar 2. Use Case

4.7. Activity Diagram



Gambar 3. Activity Diagram Admin



Gambar 4. Activity Diagram User

4.8. Evaluasi Model

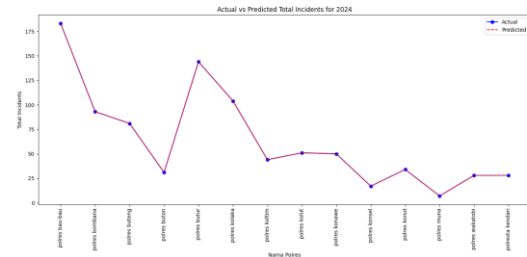
Penelitian ini bertujuan untuk memprediksi jumlah kecelakaan tertinggi di berbagai polres di Sulawesi Tenggara berdasarkan variabel meninggal dunia, luka berat dan luka ringan menggunakan algoritma regresi linear. Data pada tahun 2023 digunakan untuk melatih model dan data pada tahun 2024 digunakan untuk mengevaluasi kinerja model. Penelitian ini menggunakan algoritma regresi linear. Regresi linear adalah metode statistika yang digunakan untuk memodelkan hubungan antara satu atau lebih variabel independen (prediktor) dengan variabel dependen (respons) [12]. Tujuan utama dari regresi linear adalah untuk memprediksi nilai variabel dependen berdasarkan nilai variabel independen. Serta menggunakan Pustaka *scikit-learn* untuk membangun model regresi linear yang merupakan salah satu algoritma dasar dalam data *mining* untuk memprediksi nilai numerik berdasarkan hubungan linear antara variabel *input* dan *output*. Pada algoritma regresi linear membutuhkan metrik evaluasi yang digunakan termasuk *Mean Squared Error* (MSE) dan *R-squared* (R^2).

Mean Squared Error (MSE) mengukur rata-rata kuadrat perbedaan antara nilai yang diprediksi oleh model dan nilai yang diamati atau sebenarnya. Selain itu, *R-squared* (R^2) adalah metrik yang mengukur

seberapa baik model regresi linear cocok dengan data yang ada.

Untuk mendapatkan *output* pada algoritma regresi linear, diperlukan 4 tahapan:

Berikut hasil perbandingan antara total kecelakaan lalu lintas yang sebenarnya dan yang di prediksi untuk tahun 2024 di berbagai polres di Sulawesi Tenggara.



Gambar 5. Hasil Perdiiksi Kecelakaan Lalu Lintas

Nilai MSE yang mendekati nol membuktikan bahwa model prediksi yang dipakai memiliki kesalahan yang kecil. Ini berarti perbedaan antara nilai prediksi dan nilai *actual* sangat kecil, membuktikan bahwa model linear regresi sangat cocok dengan data yang digunakan. Selain itu, Nilai *R-squared* 1.0 yang artinya model mampu menjelaskan variasi total dari data tes secara sempurna. Model linear regresi yang dibangun mampu memprediksi total kecelakaan dengan akurasi yang sangat tinggi berdasarkan variabel *input* seperti meninggal dunia, luka berat dan luka ringan. Model ini dapat memberikan prediksi yang sangat akurat untuk tahun 2025 berdasarkan data yang dilatih pada tahun 2023 dan data yang dievaluasi pada tahun 2024. Hasil ini menunjukkan bahwa model memiliki performa yang sangat baik dan dapat diandalkan untuk prediksi jumlah kecelakaan lalu lintas di masa mendatang.

Dengan hasil MSE yang mendekati nol dan *R-squared* sebesar 1.0 dapat disimpulkan bahwa model linear *regression* yang dibangun sangat baik dalam memprediksi jumlah kecelakaan lalu lintas berdasarkan data yang di gunakan.

Pembahasan terhadap hasil penelitian dan pengujian yang diperoleh disajikan dalam bentuk uraian teoritik, baik secara kualitatif maupun kuantitatif. Hasil percobaan sebaiknya ditampilkan dalam berupa grafik ataupun tabel.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil menunjukkan bahwa algoritma regresi linear dapat digunakan dengan sangat efektif untuk memprediksi jumlah kecelakaan lalu lintas di Provinsi Sulawesi Tenggara berdasarkan data historis. Dengan nilai MSE yang mendekati nol dan nilai *R-squared* yang sempurna, model yang dibangun menunjukkan tingkat akurasi yang sangat tinggi. Hasil analisis ini memberikan gambaran yang jelas tentang distribusi dan karakteristik kecelakaan lalu lintas di berbagai wilayah Polres, serta faktor-faktor yang berkontribusi terhadap tingkat keparahan

kecelakaan. Oleh karena itu, model ini dapat diandalkan untuk digunakan dalam pengambilan keputusan terkait kebijakan dan tindakan preventif untuk mengurangi jumlah kecelakaan lalu lintas di masa mendatang. Sebagai saran, perlu dilakukan pemutakhiran data secara berkala dan integrasi dengan variable lain, seperti kondisi jalan dan cuaca, untuk meningkatkan akurasi model dan memberikan wawasan yang lebih komprehensif dalam upaya menurunkan tingkat kecelakaan.

DAFTAR PUSTAKA

- [1] S. B. Utami and Y. Al Irsyadi, "Sistem Klasifikasi Kecelakaan Lalu Lintas Jalan Raya Di Kota Boyolali Menggunakan Metode Naïve Bayes," *J. Mitra Manaj.*, vol. 2, no. 3, pp. 190–203, 2018, doi: 10.52160/ejmm.v2i3.91.
- [2] R. Saragih and J. N. Sitompul, "Analisis perbandingan data mining mengidentifikasi pola keterkaitan variabel kecelakaan lalu lintas di Polresta Kota Medan," *J. Inf. Syst. Dev.*, vol. 4, no. 1, pp. 39–45, 2019, [Online]. Available: <https://ejournal.medan.uph.edu/index.php/isd/article/view/281>
- [3] N. Kusumawati, F. Marisa, and I. D. Wijaya, "Prediksi Kurs Rupiah Terhadap Dollar Amerika Dengan Menggunakan Metode Regresi Linear," *J I M P - J. Inform. Merdeka Pasuruan*, vol. 2, no. 3, pp. 45–56, 2017, doi: 10.37438/jimp.v2i3.79.
- [4] D. Umur and M. Restorative, "Penyelesaian Kasus Kecelakaan Lalu Lintas Oleh Anak Dibawah Umur Melalui Restorative Justice," *J. Prointegrita*, vol. 6, no. 1, pp. 203–208, 2018, doi: 10.46930/jurnalprointegrita.v6i1.1507.
- [5] A. M. Rianti and I. Farida, "Analisis Pengujian Kendaraan Untuk Meminimalisir Risiko Kecelakaan Lalu Lintas," *J. Konstr.*, vol. 20, no. 1, pp. 151–160, 2022, doi: 10.33364/konstruksi.v.20-1.1043.
- [6] D. A. Trianggana, "a Peramalan Jumlah Siswa-Siswi Melalui Pendekatan Metode Regresi Linear," *J. Media Infotama*, vol. 16, no. 2, pp. 115–120, 2020, doi: 10.37676/jmi.v16i2.1149.
- [7] Y. Aqsho Ramadhan, A. Faqih, and G. Dwilestari, "Prediksi Penjualan Handphone di Toko X menggunakan Algoritma Regresi Linear," *J. Inform. Terpadu*, vol. 9, no. 1, pp. 40–44, 2023, doi: 10.54914/jit.v9i1.692.
- [8] W. Andriani, Gunawan, and A. E. Prayoga, "Prediksi Nilai Emas Menggunakan Algoritma Regresi Linear," *J. Ilm. Inform. Komput.*, vol. 28, no. 1, pp. 27–35, 2023, doi: 10.35760/ik.2023.v28i1.8096.
- [9] N. Suhandi, E. A. K. Putri, and S. Agnisa, "Analisis Pengaruh Jumlah Penduduk terhadap Jumlah Kemiskinan Menggunakan Metode Regresi Linear di Kota Palembang," *J. Ilm. Inform. Glob.*, vol. 9, no. 2, pp. 77–82, 2018, doi: 10.36982/jiig.v9i2.543.
- [10] D. Gunawan, "Evaluasi Performa Pemecahan Database dengan Metode Klasifikasi Pada Data Preprocessing Data mining," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 2, no. 1, pp. 10–13, 2016, doi: 10.23917/khif.v2i1.1749.
- [11] R. Bahtiar, "Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor," *J. Inform. MULTI*, vol. 1, no. 3, pp. 203–214, 2023, [Online]. Available: <https://jurnal.publikasitecno.id/index.php/jim203>
- [12] F. Yunita, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru," *Sistemasi*, vol. 7, no. 3, p. 238, 2018, doi: 10.32520/stmsi.v7i3.388.
- [13] P. A. R. Putri, S. S. Prasetyowati, and Y. Sibaroni, "The Performance of the Equal-Width and Equal-Frequency Discretization Methods on Data Features in Classification Process," *Sinkron*, vol. 8, no. 4, pp. 2082–2098, 2023, doi: 10.33395/sinkron.v8i4.12730.
- [14] A. Hijriani, K. Muludi, and E. A. Andini, "Penyajian Hasil Prediksi Pemakaian Air Bersih Pdam Informasi Geografis," *J. Inform.*, vol. 11, no. 2, pp. 37–42, 2016.
- [15] Harsiti, Z. Muttaqin, and E. Srihartini, "Penerapan Metode Regresi Linier Sederhana Untuk Prediksi Persediaan Obat Jenis Tablet," *JSiI (Jurnal Sist. Informasi)*, vol. 9, no. 1, pp. 12–16, 2022, doi: 10.30656/jsii.v9i1.4426.