

Penerapan Regresi Linier Berganda Untuk Menganalisis Jumlah Kecelakaan Lalu Lintas

Gita Aprilia Nurhasanah

Program Studi Teknik Informatika, fakultas Teknologi Informasi, Universitas Serang Raya

Jl. Raya Serang, Cilegon KM. 5 Taman Drangong Serang Banten

Email: gaprilia29@gmail.com

ABSTRAKSI

Kecelakaan lalu lintas merupakan permasalahan serius yang dapat menimbulkan dampak negatif terhadap masyarakat dan ekonomi suatu wilayah. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang berkontribusi terhadap kejadian kecelakaan lalu lintas serta mengidentifikasi upaya pencegahan yang dapat dilakukan. Metode penelitian ini melibatkan pengumpulan data kecelakaan lalu lintas dari berbagai sumber, termasuk laporan polisi, keterangan saksi, dan rekam medis. Analisis faktor penyebab kecelakaan melibatkan aspek-aspek seperti kecepatan, ketidakpatuhan aturan lalu lintas, kondisi jalan, dan kondisi kendaraan. Hasil penelitian menunjukkan bahwa kecepatan yang tidak sesuai, penggunaan ponsel selama berkendara, dan pelanggaran aturan lalu lintas menjadi faktor dominan dalam kecelakaan lalu lintas. Selain itu, kondisi infrastruktur jalan yang buruk dan kurangnya pemeliharaan kendaraan juga berkontribusi pada tingginya angka kecelakaan. Penelitian ini memberikan kontribusi pada pemahaman lebih lanjut tentang faktor-faktor yang menyebabkan kecelakaan lalu lintas dan memberikan pandangan tentang strategi pencegahan yang dapat diadopsi oleh pemerintah dan pihak terkait. Dengan menerapkan upaya-upaya ini, diharapkan dapat mengurangi angka kecelakaan lalu lintas dan meningkatkan keselamatan jalan bagi seluruh pengguna lalu lintas. Kendaraan liner berganda, seperti bus dan truk besar, memiliki peran penting dalam transportasi massal dan logistik. Namun, kecelakaan lalu lintas yang melibatkan kendaraan-kendaraan ini dapat memiliki dampak serius terhadap keselamatan jalan dan kehidupan manusia. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang berkontribusi terhadap kecelakaan lalu lintas dengan kendaraan liner berganda dan mengidentifikasi strategi pencegahan yang dapat diimplementasikan.

Kata Kunci: data mining, regresi linier berganda, kecelakaan lalu lintas

ABSTRACT

Traffic accidents are a serious problem that can have negative impacts on both society and the economy of a region. This study aims to analyze the factors contributing to traffic accidents and identify possible prevention measures. The research method involves collecting traffic accident data from various sources, including police reports, witness statements, and medical records. The analysis of accident causes covers aspects such as speeding, non-compliance with traffic regulations, road conditions, and vehicle conditions. The results show that inappropriate speed, the use of mobile phones while driving, and violations of traffic rules are the dominant factors in traffic accidents. In addition, poor road infrastructure conditions and lack of vehicle maintenance also contribute to the high rate of accidents. This study contributes to a deeper understanding of the factors causing traffic accidents and provides insights into prevention strategies that can be adopted by the government and relevant stakeholders. By implementing these measures, it is expected that the number of traffic accidents can be reduced, thereby improving road safety for all road users. Multiple-unit vehicles, such as buses and large trucks, play an important role in mass transportation and logistics. However, traffic accidents involving these vehicles can have serious impacts on road safety and human life. This study aims to analyze the factors contributing to traffic accidents involving multiple-unit vehicles and identify prevention strategies that can be implemented.

Keywords: data mining, multiple linear regression, traffic accidents

Penulis Korespondensi

Gita Aprilia Nurhasanah

Tanggal Submit : 26/01/2024

Tanggal Diterima : 17/07/2025

Tanggal Terbit : 15/08/2025

This is an open access article under the [CC-BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license



Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 (CC BY-NC-SA 4.0) International License (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

Publisher's Note: JPPM stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

I. PENDAHULUAN

Jalan merupakan kebutuhan yang sangat penting dalam sistem transportasi lalu lintas, baik transportasi darat, laut maupun udara. Kurangnya perhatian terhadap masalah keselamatan akan menyebabkan kecelakaan lalu lintas semakin meningkat [5]. Kecelakaan adalah tabrakan, overtuning atau slip yang terjadi di jalan terbuka pada lalu lintas umum yang menyebabkan luka, kematian/fatal, kerusakan pada kendaraan atau kerugian material[2]. Dengan meningkatnya jumlah kepemilikan masyarakat terhadap kendaraan transportasi, maka tingkat kemacetan lalu lintas yang terjadi akan semakin tinggi pula. Faktor hambatan samping adalah merupakan salah satu penyebab terjadinya kemacetan lalu lintas yang dapat mempengaruhi tingkat kinerja pelayanan suatu ruas jalan. Kemacetan merupakan suatu permasalahan arus lalu lintas yang disebabkan oleh peningkatan volume kendaraan di suatu ruas jalan dan oleh hambatan samping [1]. Dalam upaya memahami faktor-faktor yang berkontribusi pada kejadian kecelakaan lalu lintas, penggunaan metode analisis statistik, seperti regresi linier berganda, dapat memberikan wawasan yang mendalam.

Regresi linier berganda memungkinkan kita untuk mengevaluasi hubungan antara variabel-variabel independen yang mungkin mempengaruhi tingkat kecelakaan lalu lintas. Dengan mempertimbangkan sejumlah variabel seperti kecepatan, kepadatan lalu lintas, kondisi jalan, dan faktor-faktor lainnya, kita dapat mengidentifikasi sejauh mana setiap variabel tersebut berperan dalam memprediksi jumlah kecelakaan lalu lintas.

Penelitian ini bertujuan untuk menggunakan analisis regresi linier berganda guna menyelidiki korelasi antara variabel-variabel tersebut dan tingkat kecelakaan lalu lintas. Data yang dikumpulkan akan mencakup catatan kecelakaan, data lalu lintas, serta informasi tentang kondisi jalan. Dengan memahami faktor-faktor yang memiliki korelasi signifikan dengan kecelakaan lalu lintas, diharapkan kita dapat mengembangkan strategi pencegahan yang lebih efektif.

Pentingnya penelitian ini terletak pada kemampuan untuk memberikan landasan ilmiah bagi kebijakan keselamatan jalan. Dengan memahami variabel-variabel yang memiliki dampak signifikan, pemerintah dan pihak terkait dapat mengambil langkah-langkah proaktif untuk meningkatkan kondisi keselamatan lalu lintas. Oleh karena itu, penelitian ini tidak hanya akan memberikan wawasan mendalam tentang faktor-faktor penyebab kecelakaan lalu lintas, tetapi juga memberikan kontribusi nyata dalam upaya meningkatkan keselamatan dan kesejahteraan masyarakat pengguna jalan.

II. METODE PENELITIAN

2.1 Data Mining

Data mining merupakan metode untuk menemukan suatu pengetahuan dalam suatu database yang cukup besar [4]. Data Mining merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi pengetahuan yang berguna dan bermanfaat yang tersimpan di dalam database besar [3]. Patut diingat bahwa kata mining sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar. Karena itu Data Mining sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan

(artificial intelligent), machine learning, statistik dan database [3]. Data mining adalah proses menggali dan menganalisa sejumlah data yang sangat besar untuk memperoleh sesuatu yang benar, baru, sangat bermanfaat dan akhirnya dapat menemukan suatu corak atau pola dalam data tersebut [10]. Data mining perlu dilakukan oleh semua pihak yang telah melakukan berbagai transaksi dan ingin melakukan inovasi baik di bidang penjualan, pelayanan jasa dan lain-lain. Proses menambang data ini akan menghasilkan berbagai informasi yang sangat penting untuk di jadikan dasar pengambilan keputusan inovasi apa yang akan dilakukan. Semakin valid datanya semakin [6].

2.2 Regresi Linier berganda

Regresi linier berganda adalah regresi yang variabel terikat Y dihubungkan dengan lebih dari satu variabel bebas X [2]. Regresi linier berganda merupakan metode statistik yang digunakan untuk menganalisis hubungan antara satu variabel dependen (variabel yang ingin diprediksi) dan dua atau lebih variabel independen (variabel yang digunakan untuk memprediksi variabel dependen). Metode ini memungkinkan kita untuk memahami sejauh mana variabel-variabel independen tersebut berkontribusi terhadap variasi dalam variabel dependen. Algoritma Regresi Linear Berganda merupakan Analisis yang memiliki variabel bebas lebih dari satu disebut analisis regresi linier berganda. Teknik regresi linier berganda digunakan untuk mengetahui ada tidaknya pengaruh signifikan dua atau lebih variabel bebas ($X_1, X_2, X_3, \dots, X_k$) terhadap variabel terikat (Y).

III. HASIL DAN PEMBAHASAN

Pada penelitian ini, perhitungan regresi linier berganda dilakukan secara manual dan kemudian dibandingkan dengan hasil yang diperoleh menggunakan aplikasi Orange. Data yang digunakan meliputi variabel kepadatan lalu lintas (X_1), kecepatan kendaraan (X_2), dan jumlah kecelakaan lalu lintas (Y). Untuk keperluan analisis, dibuat tabel tambahan yang memuat variabel turunan seperti $X_1^2, X_2^2, Y^2, X_1X_2, X_1Y$, dan X_2Y . Data ini menjadi dasar dalam proses perhitungan nilai koefisien regresi serta koefisien determinasi.

Langkah pertama dalam analisis manual adalah menghitung jumlah dari setiap variabel turunan yang diperlukan. Nilai yang diperoleh adalah:

$$\begin{aligned}\sum X_1^2 &= 303260,2305, \\ \sum X_2^2 &= 282855,5, \\ \sum Y^2 &= 449194,8, \\ \sum X_1Y &= 337732,9, \\ \sum X_2Y &= 330392,5, \text{ dan} \\ \sum X_1X_2 &= 257458,7\end{aligned}$$

Nilai-nilai ini kemudian digunakan dalam formula regresi linier berganda untuk menghitung koefisien b_1, b_2 , dan konstanta a .

$$b_1 = \frac{\{(\sum x_2^2 \times \sum x_1 y) - (\sum x_1 y \times \sum x_1 x_2)\}}{\{(\sum x_1^2 \times \sum x_2^2) - (\sum x_1 x_2)^2\}}$$

$$b_1 = \frac{\{(53533,2 \times 42202,1) - (44067,7 \times 20763,2)\}}{\{(58954,4 \times 53533,2) - (20763,2)^2\}}$$

$$b_1 = \frac{1344226991,08}{2724907211,9}$$

$$b_1 = 0,493$$

$$b_2 = \frac{(\sum x_1^2 \times \sum x_2 y) - (\sum x_1 y \times \sum x_1 x_2)}{\{(\sum x_1^2 \times \sum x_2^2) - (\sum x_1 x_2)^2\}}$$

$$b_2 = \frac{(58954,4 \times 44067,7) - (44067,7 \times 20763,2)}{\{(58954,4 \times 53533,2) - (20763,2)^2\}}$$

$$b_2 = \frac{1682998344,2}{2724907211,8}$$

$$b_2 = 0,617$$

$$a = \frac{(\sum y) - (b_1 \times \sum x_1) - (b_2 \times \sum x_2)}{n}$$

$$a = \frac{(5979,1) - (0,493 \times 4942,73) - (0,617 \times 4788,76)}{100}$$

$$a = 5,87$$

Hasil perhitungan manual menunjukkan bahwa nilai b1 adalah 0,493, nilai b2 adalah 0,617, dan konstanta a sebesar

5,87. Nilai koefisien determinasi (r^2) diperoleh sebesar 0,52, yang berarti bahwa 52% variasi pada jumlah kecelakaan lalu lintas dapat dijelaskan oleh kombinasi variabel kepadatan lalu lintas dan kecepatan kendaraan. Sisanya sebesar 48% dipengaruhi oleh faktor lain di luar model yang digunakan.

$$r^2 = \frac{(b_1 \times \sum x_1 y) + (b_2 \times \sum x_2 y)}{\sum y^2}$$

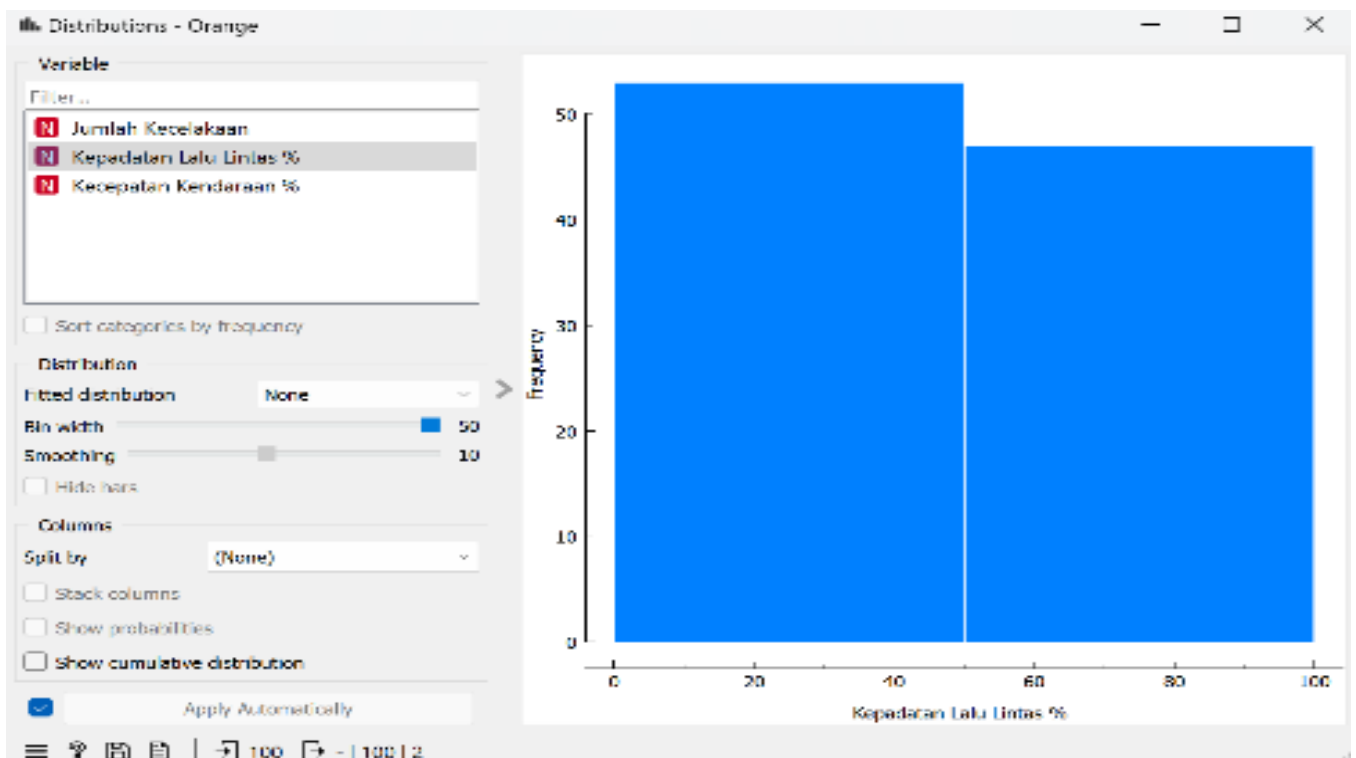
$$r^2 = \frac{(0,493 \times 42202,1) + (0,617 \times 44067,7)}{91698,4}$$

$$r^2 = \frac{239977031}{458492000}$$

$$r^2 = 0,52$$

Selanjutnya, perhitungan menggunakan aplikasi Orange dilakukan untuk memvalidasi hasil perhitungan manual. Data dari Microsoft Excel dimasukkan ke dalam Orange, kemudian variabel Y diatur sebagai target, sedangkan X1 dan X2 menjadi variabel prediktor. Modul yang digunakan meliputi File, Distributions, Data Table, Linear Regression, Correlations, Line Plot, dan Scatter Plot.

Modul Distributions (lihat Gambar 1) pada Orange menampilkan distribusi nilai dari masing-masing variabel. Jika terdapat variabel kelas, distribusi dapat dikondisikan sesuai kelas tersebut. Tampilan ini membantu untuk memahami persebaran data dan potensi adanya ketidakseimbangan atau bias pada dataset yang digunakan.



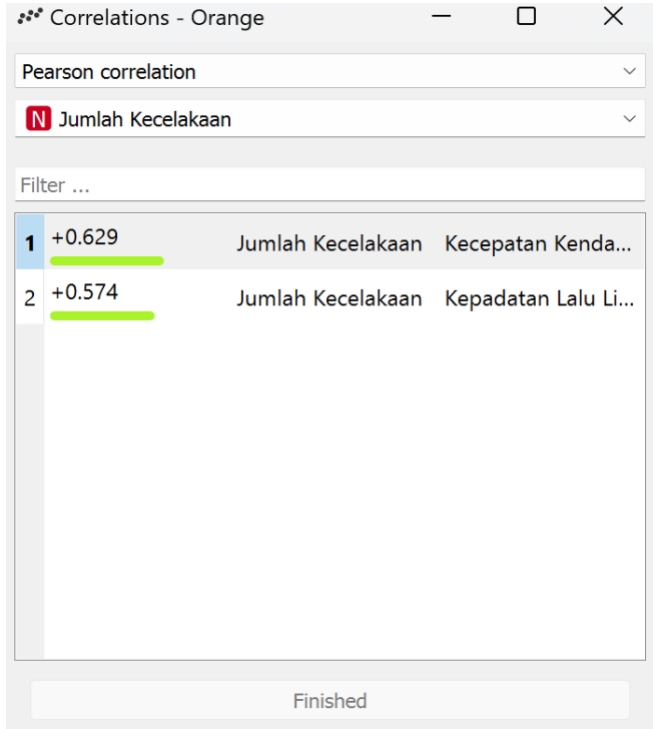
Gambar 1. Tampilan Distribusi Nilai

Modul Linear Regression di Orange memberikan hasil identifikasi hubungan antara variabel prediktor (X1 dan X2) terhadap variabel respon (Y). Nilai koefisien regresi yang dihasilkan mendekati hasil perhitungan manual, sehingga dapat disimpulkan bahwa kedua metode memberikan hasil yang konsisten (lihat Gambar 2).

	name	coef
1	intercept	5.15021
2	Kepadatan Lalu...	0.493311
3	Kecepatan Ken...	0.631849

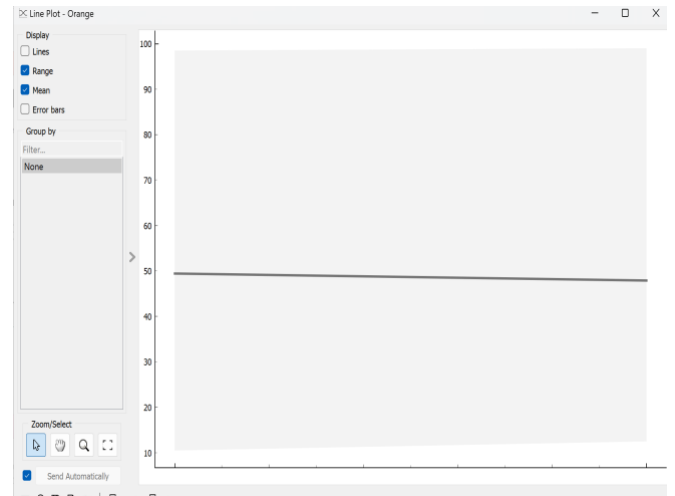
Gambar 2. Koefisien Regresi

Modul Correlation (lihat Gambar 3) digunakan untuk mengidentifikasi kekuatan hubungan antar variabel. Hasilnya menunjukkan adanya hubungan positif antara kepadatan lalu lintas dan jumlah kecelakaan, serta antara kecepatan kendaraan dan jumlah kecelakaan. Nilai korelasi yang diperoleh mendukung kesimpulan bahwa kedua variabel prediktor memiliki pengaruh signifikan terhadap jumlah kecelakaan lalu lintas.



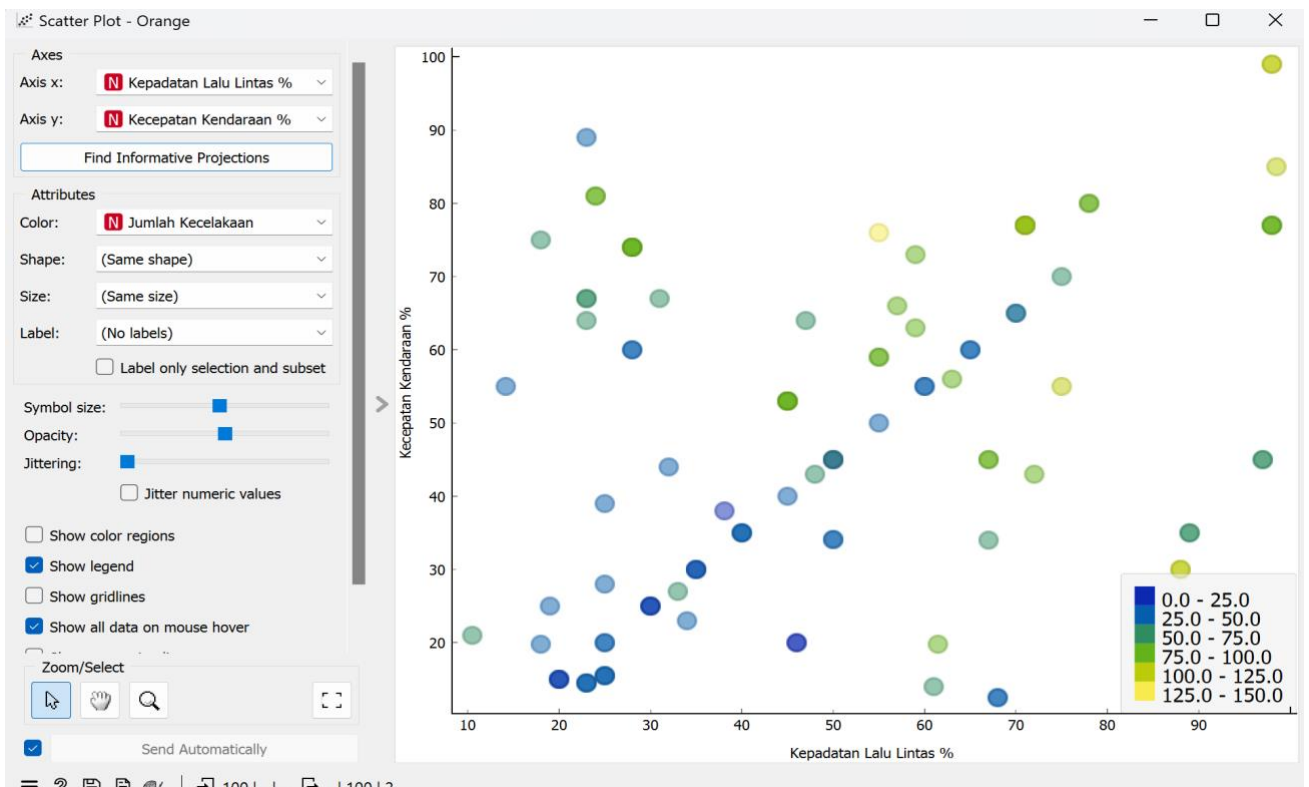
Gambar 3. Modul Korelasi Orange

Modul Line Plot (lihat Gambar 4) memvisualisasikan hubungan antara variabel prediktor dan respon dalam bentuk garis. Grafik ini membantu memahami arah dan bentuk hubungan antar variabel, sehingga dapat memperkuat interpretasi hasil perhitungan regresi yang telah dilakukan.



Gambar 4. Modul Line Plot

Modul Scatter Plot (lihat Gambar 5) menampilkan hubungan dua dimensi antara variabel numerik. Setiap titik pada grafik mewakili data observasi yang dipetakan pada sumbu X dan Y. Dari scatter plot, terlihat adanya pola hubungan positif di mana kenaikan nilai kepadatan lalu lintas dan kecepatan kendaraan diikuti oleh kenaikan jumlah kecelakaan. Hal ini konsisten dengan hasil analisis regresi dan korelasi.



Gambar 5. Visualisasi Scatter Plot

Secara keseluruhan, hasil perhitungan manual dan perhitungan menggunakan Orange menunjukkan konsistensi. Nilai koefisien determinasi sebesar 0,52 menunjukkan model mampu menjelaskan sebagian besar variasi jumlah kecelakaan berdasarkan kepadatan lalu lintas dan kecepatan kendaraan. Visualisasi yang dihasilkan oleh Orange juga membantu dalam memahami hubungan antar variabel, serta memberikan gambaran yang jelas mengenai pengaruh masing-masing faktor terhadap jumlah kecelakaan lalu lintas.

V. KESIMPULAN

Berdasarkan hasil analisis yang telah dilakukan, dapat disimpulkan bahwa penerapan metode data mining dengan pendekatan regresi linier berganda mampu mengidentifikasi faktor-faktor signifikan yang memengaruhi terjadinya kecelakaan lalu lintas. Faktor-faktor seperti kecepatan kendaraan, kepatuhan terhadap aturan lalu lintas, kondisi jalan, serta kelayakan kendaraan terbukti memiliki pengaruh yang cukup besar terhadap frekuensi kecelakaan. Dengan memanfaatkan teknik analisis ini, data yang semula bersifat mentah dapat diolah menjadi informasi yang relevan dan berguna sebagai dasar pengambilan keputusan. Hal ini membuktikan bahwa data mining bukan hanya menjadi alat analisis data, tetapi juga instrumen strategis dalam merancang kebijakan keselamatan lalu lintas yang lebih efektif.

Selanjutnya, temuan penelitian ini memberikan kontribusi penting dalam mendukung upaya pencegahan kecelakaan lalu lintas secara terukur dan berbasis data. Pemerintah, aparat penegak hukum, dan pihak terkait lainnya dapat memanfaatkan hasil ini untuk menyusun strategi yang lebih tepat sasaran, seperti perbaikan infrastruktur jalan, peningkatan pengawasan terhadap kepatuhan lalu lintas, serta program edukasi bagi pengendara. Dengan implementasi yang konsisten dan terintegrasi, diharapkan angka kecelakaan lalu lintas dapat ditekan secara signifikan, sehingga tercipta lingkungan berkendara yang lebih aman, nyaman, dan berkelanjutan bagi seluruh pengguna jalan.

DAFTAR PUSTAKA

- [1] M. Harmizi*,Azanuddin.,**Elfitriani.** 2019, Estimasi Data Mining Tingkat Kecelakaan Lalu Lintas Kabupaten Serdang Bedagai Menggunakan Algoritma Regresi Linier Berganda
- [2] "15912-ID-faktor-faktor-yang-mempengaruhi-keparahan-korban-kecelakaan-lalu-lintas-di-kota"
- [3] Purwadi, Puji Sari Ramadhan, Nurdianti Safitri 2019, Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang
- [4] Agus Sasmito Aribowo 2012, Analisa Asosiatif Data Mining untuk Mengetahui Pola Kecelakaan Lalu Lintas
- [5] I Kadek Agus Pande Purnanta, Ida Ayu Sri Mahatpani, I Made Harta Wijaya 2020, Analisis Kecelakaan Lalu Lintas di Ruas Jalan Nasional Kolektor Primer diKabupaten Gianyar dengan Metode Statistik Regresi Linier Berganda
- [6] upriyanta 1) Noor Hasan 2013, DATA MINING DENGAN METODE REGRESI LINIER UNTUK MELAKUKAN PREDIKSI
- [7] https://orange3-readthedocsio.translate.goog/projects/orange-visual-programming/en/latest/widgets/visualize/distributions.html?_x_tr_sl=en&_x_tr_tl=id&_x_tr_hl=id&_x_tr_pto=c
- [8] https://www-polymersearch-com.translate.goog/glossary/scatter-plot?_x_tr_sl=en&_x_tr_tl=id&_x_tr_hl=id&_x_tr_pto=tc
- [9] Tada, T., Nagashima, T., Okada, Y., 2010, Rule-based Classification for Audio Data Based on Closed Itemset Mining, Proceedings of the International MultiConference of Engineers and Computer Scientist 2010 Vol I, Hongkong