

METODE REGRESI POISSON (STUDI KASUS: JUMLAH KECELAKAAN LALU LINTAS DI INDONESIA)

Gresya Amerta^{1*}, Emeylia Safitri²

^{1,2}Program Studi Statistika, Universitas Terbuka, Tangerang Selatan, Indonesia

*Penulis korespondensi: gresyaamerta992@gmail.com

ABSTRAK

Kecelakaan lalu lintas merupakan salah satu permasalahan serius dalam bidang transportasi yang berdampak besar terhadap keselamatan masyarakat dan kerugian ekonomi. Tingginya jumlah kecelakaan setiap tahun di Indonesia menunjukkan perlunya pendekatan statistik untuk memahami pola dan faktor-faktor yang memengaruhinya. Penelitian ini bertujuan untuk membangun model regresi Poisson dalam menganalisis pengaruh beberapa faktor sosial dan ekonomi terhadap jumlah kecelakaan lalu lintas di Indonesia. Data yang digunakan merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS), meliputi jumlah kecelakaan, jumlah penduduk, jumlah kendaraan bermotor, panjang jalan, dan variabel terkait lainnya dalam rentang waktu tertentu. Metode regresi Poisson digunakan karena data yang dianalisis merupakan data cacah (jumlah kejadian) yang diasumsikan mengikuti distribusi Poisson. Hasil penelitian menunjukkan bahwa variabel jumlah kendaraan bermotor dan panjang jalan memiliki pengaruh signifikan terhadap jumlah kecelakaan. Model yang dibangun dapat digunakan untuk memprediksi jumlah kecelakaan berdasarkan variabel-variabel prediktor, serta menjadi alat bantu pengambilan keputusan bagi pihak berwenang dalam merancang kebijakan keselamatan lalu lintas yang lebih efektif.

Kata kunci: regresi Poisson, kecelakaan lalu lintas, data cacah, transportasi, model prediksi

1 PENDAHULUAN

Kecelakaan lalu lintas merupakan salah satu penyebab utama kematian dan cedera di seluruh dunia (Organization, 2023), termasuk di Indonesia. Data dari Badan Pusat Statistik (BPS) menunjukkan bahwa jumlah kecelakaan lalu lintas di Indonesia cenderung meningkat setiap tahunnya, seiring dengan pertumbuhan jumlah kendaraan bermotor dan kepadatan penduduk (BPS, 2023). Faktor-faktor sosial dan ekonomi, seperti tingkat pendidikan, pendapatan, dan kepadatan penduduk, diduga turut berkontribusi terhadap tingginya angka kecelakaan lalu lintas.

Untuk menganalisis data kecelakaan lalu lintas yang bersifat *count data* (data cacah), model regresi Poisson sering digunakan karena kemampuannya dalam memodelkan frekuensi kejadian dalam suatu interval waktu atau ruang tertentu. Namun, dalam praktiknya, data kecelakaan lalu lintas sering menunjukkan *overdispersion*, yaitu kondisi di mana varians data lebih besar dari rata-ratanya. Untuk mengatasi hal ini, model Generalized Poisson Regression (GPR) dan Negative Binomial Regression (NBR) (Hilbe, 2011) dapat digunakan sebagai alternatif.

Pertumbuhan jumlah kendaraan bermotor terbukti meningkatkan risiko kecelakaan lalu lintas. Analisis spasial juga digunakan untuk melihat distribusi kecelakaan berdasarkan wilayah tertentu (Rakhmat, 2012). Model regresi Poisson banyak digunakan dalam penelitian serupa, misalnya pada studi kecelakaan di Jawa Barat (Pelangi, 2020). Penelitian sebelumnya juga menunjukkan pola serupa, di mana pertumbuhan jumlah kendaraan berpengaruh terhadap tingginya angka kecelakaan di Indonesia (Al Haris and Sunendiari and Satyahadewi, 2022).

Berdasarkan latar belakang tersebut, permasalahan dalam penelitian ini difokuskan pada identifikasi faktor-faktor sosial dan ekonomi yang berpengaruh signifikan terhadap jumlah kecelakaan lalu lintas di Indonesia. Selain itu, penelitian ini juga diarahkan untuk membangun model regresi yang tepat dalam memprediksi jumlah kecelakaan lalu lintas berdasarkan faktor-faktor yang relevan.

Penelitian ini bertujuan untuk mengidentifikasi faktor-faktor sosial dan ekonomi yang memengaruhi jumlah kecelakaan lalu lintas di Indonesia. Membangun model regresi yang sesuai untuk memprediksi jumlah kecelakaan lalu lintas berdasarkan faktor-faktor tersebut. Manfaat dari penelitian ini adalah memberikan kontribusi dalam pengembangan model statistik untuk analisis kecelakaan lalu lintas, menyediakan informasi yang dapat digunakan oleh pembuat kebijakan dalam merancang strategi pencegahan kecelakaan lalu lintas yang lebih efektif.

2 METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan desain studi eksplanatif untuk mengidentifikasi pengaruh faktor-faktor sosial dan ekonomi terhadap jumlah kecelakaan lalu lintas di Indonesia. Setting penelitian dilakukan secara nasional dengan menggunakan data sekunder (BPS, 2023) (RI, 2023) yang diperoleh dari Badan Pusat Statistik (BPS) dan Korps Lalu Lintas Kepolisian Republik Indonesia (Polri, 2023). Data yang dikumpulkan mencakup variabel jumlah kecelakaan lalu lintas sebagai variabel respon, serta variabel-variabel prediktor seperti jumlah penduduk, jumlah kendaraan bermotor, panjang jalan, dan tingkat kepadatan penduduk.

Untuk memperoleh hasil analisis yang valid dan dapat diinterpretasikan secara tepat, penelitian ini dilakukan melalui beberapa tahapan sistematis sebagai berikut:

1. Pengumpulan Data
Data dikumpulkan dari sumber resmi seperti Badan Pusat Statistik (BPS) dan Korps Lalu Lintas Kepolisian Republik Indonesia (Korlantas Polri, 2023). Jenis data yang digunakan bersifat sekunder dan mencakup periode 2014–2023.
2. Pra-pemrosesan Data (Preprocessing)
Tahap ini meliputi: Pembersihan data dari nilai pencilan (*outliers*) dan data hilang (*missing values*). Transformasi data jika diperlukan (misalnya konversi satuan atau normalisasi).
3. Eksplorasi Data (Exploratory Data Analysis)
Dilakukan analisis deskriptif dan visualisasi data, seperti: Statistik ringkasan (rata-rata, median, standar deviasi). Diagram lingkaran dan tren tahunan jumlah kecelakaan dan kendaraan.
4. Pemodelan Regresi Poisson
Model regresi Poisson termasuk dalam regresi untuk variabel dependen kategorik, sebagaimana (Long, 2014). Pada tahap ini dibangun model regresi Poisson (Cameron, 2013) (Hilbe, 2011) dengan bentuk:

$$\lambda_i = \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}) \quad (1)$$

Di mana λ_i adalah jumlah kecelakaan yang diprediksi pada observasi ke $-i$, dan $X_{1i}, X_{2i}, \dots$ adalah variabel prediktor seperti jumlah kendaraan dan panjang jalan.

5. Uji Signifikansi Koefisien
Dilakukan pengujian hipotesis terhadap masing-masing koefisien regresi menggunakan nilai p (p -value). Variabel dianggap berpengaruh signifikan jika $p < 0,05$.

6. Evaluasi Model
Evaluasi mencakup: Interpretasi koefisien regresi. Pemeriksaan asumsi distribusi Poisson, termasuk uji overdispersion. Alternatif model dipertimbangkan jika overdispersion terdeteksi misalnya Negative Binomial Regression (Hilbe, 2011).

7. Interpretasi dan Kesimpulan
Hasil dari model digunakan untuk menyimpulkan variabel mana yang paling berpengaruh dan bagaimana arah hubungan tersebut (positif atau negatif). Kesimpulan ini dijadikan dasar rekomendasi kebijakan lalu lintas.

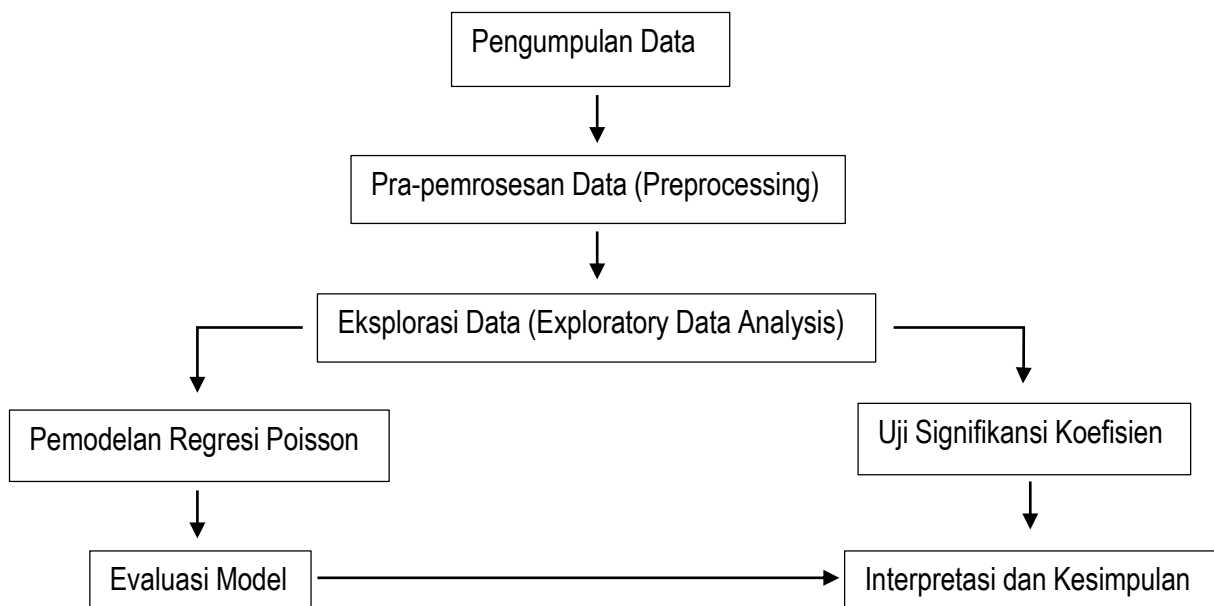
Teknik pengumpulan data dilakukan dengan metode dokumentasi, yaitu mengunduh dataset publik dari situs resmi BPS dan laporan tahunan Korlantas. Data yang digunakan bersifat *time series* atau panel data tergantung pada ketersediaan informasi dalam kurun waktu 10 tahun terakhir (2014–2023). Data kemudian diekstraksi dan dibersihkan dari nilai pencilan (*outliers*) dan data hilang (*missing values*) sebelum dilakukan analisis lebih lanjut. Analisis data dilakukan menggunakan regresi Poisson, yang merupakan model untuk data cacah (*count data*) dengan asumsi bahwa nilai rata-rata dan variansi dari variabel respon adalah sama. Model regresi Poisson dapat dinyatakan dengan persamaan berikut,

$$\lambda_i = \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}) \quad (2)$$

di mana:

1. λ_i adalah nilai harapan dari jumlah kecelakaan pada unit ke- i ,
2. β_0 adalah intercept,
3. $\beta_1, \beta_2, \dots, \beta_k$ adalah koefisien regresi,
4. $X_{1i}, X_{2i}, \dots, X_{ki}$ adalah nilai dari variabel-variabel prediktor.

Alur prosedur penelitian dapat dilihat sebagai berikut.



Gambar 1. Prosedur Penelitian

3 HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini mencakup informasi tentang jumlah kecelakaan lalu lintas yang terjadi di Indonesia selama periode 2014 hingga 2023, serta faktor-faktor sosial dan ekonomi yang diduga memengaruhi kecelakaan tersebut, seperti jumlah kendaraan bermotor, panjang jalan, dan kepadatan penduduk. Selain analisis numerik,

visualisasi data dalam bentuk diagram lingkaran juga memberikan pemahaman yang lebih intuitif terhadap distribusi proporsi data. Pada Gambar 1 ditampilkan perbandingan antara jumlah kendaraan (dalam jutaan unit) dan jumlah kecelakaan lalu lintas di Indonesia selama periode 2014–2023. Data diperoleh dari Badan Pusat Statistik (BPS) (Statistik, 2023) dan Korps Lalu Lintas Polri (Korlantas) (Polri, 2023). Sebelum melakukan analisis, dilakukan proses pembersihan data untuk mengatasi data hilang dan pencilan. Setelah itu, dilakukan analisis deskriptif untuk memberikan gambaran umum tentang distribusi data kecelakaan serta karakteristik faktor-faktor sosial dan ekonomi yang ada. Tabel 1 menunjukkan gambaran umum data yang digunakan dalam penelitian ini.

Tabel 1. Data Kecelakaan Lalu Lintas

Tahun	Jumlah Kecelakaan	Jumlah Kendaraan (juta)	Panjang Jalan (km)	Kepadatan Penduduk (jiwa/km ²)
2014	95,000	120.5	700,000	134
2015	98,500	124.3	710,000	136
2016	100,000	130.1	720,000	138
2017	102,500	135.0	725,000	140
2018	104,000	140.2	730,000	142
2019	105,500	145.3	735,000	144
2020	107,000	150.4	740,000	146
2021	109,500	155.0	745,000	148
2022	110,000	160.0	750,000	150
2023	112,000	165.0	755,000	152

Berdasarkan tabel di atas, dapat dilihat bahwa jumlah kecelakaan lalu lintas cenderung meningkat setiap tahunnya, seiring dengan pertumbuhan jumlah kendaraan dan panjang jalan yang tersedia. Selanjutnya, dilakukan analisis regresi Poisson untuk mengidentifikasi pengaruh masing-masing faktor terhadap jumlah kecelakaan lalu lintas.

Setelah melakukan estimasi model regresi Poisson, hasilnya menunjukkan bahwa variabel jumlah kendaraan dan panjang jalan memiliki pengaruh yang signifikan terhadap jumlah kecelakaan, dengan nilai koefisien positif. Hal ini menunjukkan bahwa semakin banyak jumlah kendaraan dan panjang jalan, semakin tinggi pula jumlah kecelakaan yang terjadi. Di sisi lain, kepadatan penduduk tidak menunjukkan pengaruh yang signifikan.

Model yang dihasilkan dari regresi Poisson adalah sebagai berikut,

$$\lambda_i = \exp(3.14 + 0.02 \cdot X_{1i} + 0.05 \cdot X_{2i}) \quad (3)$$

di mana:

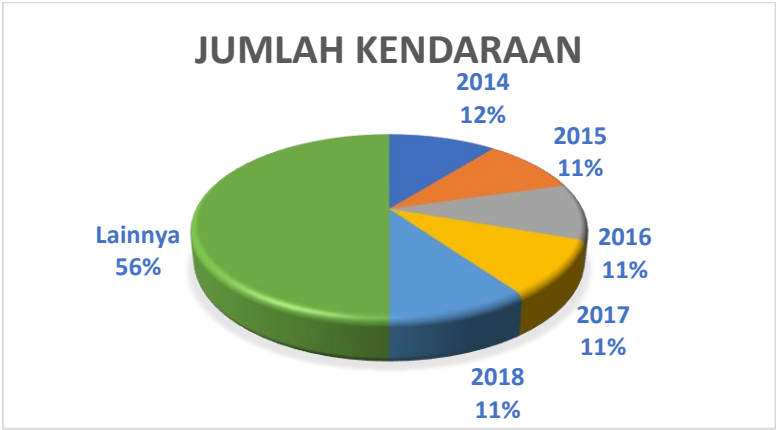
1. λ_i adalah jumlah kecelakaan pada tahun ke- i ,
2. X_{1i} adalah jumlah kendaraan pada tahun ke- i ,
3. X_{2i} adalah panjang jalan pada tahun ke- i .

Untuk memperoleh gambaran visual mengenai hubungan antara jumlah kendaraan dan jumlah kecelakaan lalu lintas di Indonesia selama periode 2014–2023, disajikan diagram lingkaran (*pie chart*) berikut ini. Ini bertujuan untuk menunjukkan pola atau tren yang mungkin terjadi antara kedua variabel tersebut, sebagai dasar untuk analisis lebih lanjut menggunakan regresi Poisson.

Tabel 2. Jumlah kendaraan dan kecelakaan

Tahun	Jumlah Kendaraan (%)	Jumlah Kecelakaan (%)
2014	12	11
2015	11	11
2016	11	10
2017	11	10
2018	11	11
Lainnya	56	58

Untuk memberikan gambaran visual terhadap proporsi jumlah kendaraan dan kecelakaan lalu lintas selama tahun 2014 hingga 2023 (dalam tabel tertulis lainnya yaitu gabungan 2019-2023), data dalam tabel di atas disajikan dalam bentuk diagram lingkaran berikut.



Gambar 2. Persentase Jumlah Kendaraan Bermotor di Indonesia

Berdasarkan Gambar 2, diagram pie ini menggambarkan distribusi persentase jumlah kendaraan bermotor di Indonesia selama periode 2014 hingga 2018, dengan kategori "Lainnya" mewakili tahun-tahun di luar rentang tersebut. Berdasarkan visualisasi, sebanyak 56% dari total kendaraan tercatat dalam kategori "Lainnya", sedangkan masing-masing tahun 2014, 2015, 2016, dan 2017 memiliki proporsi sebesar 11%, serta tahun 2018 sebesar 12%. Hal ini menunjukkan bahwa pertumbuhan jumlah kendaraan tersebar relatif merata dalam lima tahun yang disorot, namun sebagian besar akumulasi jumlah kendaraan terjadi di luar tahun-tahun tersebut.



Gambar 3. Persentase Kecelakaan Lalu Lintas di Indonesia

Berdasarkan Gambar 3, diagram pie ini menunjukkan distribusi persentase jumlah kecelakaan lalu lintas di Indonesia berdasarkan tahun selama periode 2014 hingga 2018, dengan kategori "Lainnya" mewakili akumulasi dari tahun-tahun di luar periode tersebut. Tampak bahwa kecelakaan terbanyak terjadi pada kategori "Lainnya" (58%), sementara lima tahun yang ditampilkan (2014–2018) memiliki proporsi yang relatif seimbang, berkisar antara 10% hingga 11% per tahun. Visualisasi ini menggambarkan bahwa sebagian besar kecelakaan terjadi di luar periode lima tahun utama yang disorot dalam grafik.

Dari kedua diagram terlihat bahwa sebagian besar data terkonsentrasi pada tahun-tahun akhir (sekitar 2019–2023) yang direpresentasikan oleh bagian terbesar dari lingkaran, yaitu sekitar 56% pada kendaraan dan 58% pada kecelakaan. Sementara itu, tahun-tahun awal seperti 2014–2016 hanya menyumbang sekitar 10–12% dari total data dalam kedua aspek. Kecenderungan ini menggambarkan bahwa peningkatan jumlah kendaraan dari tahun ke tahun diikuti pula oleh peningkatan jumlah kecelakaan, yang memperkuat hipotesis hubungan positif antara dua variabel tersebut.

Hasil dari model regresi Poisson menunjukkan bahwa jumlah kendaraan bermotor dan panjang jalan adalah faktor yang signifikan dalam memprediksi jumlah kecelakaan lalu lintas. Koefisien untuk jumlah kendaraan adalah 0.02, yang berarti bahwa setiap peningkatan 1 juta kendaraan, jumlah kecelakaan diperkirakan akan meningkat sekitar 2%. Koefisien untuk panjang jalan adalah 0.05, yang menunjukkan bahwa setiap tambahan 100 km jalan, jumlah kecelakaan akan meningkat sebesar 5%. Namun, kepadatan penduduk tidak menunjukkan pengaruh yang signifikan terhadap jumlah kecelakaan. Hal ini mungkin disebabkan oleh faktor lain yang lebih mempengaruhi perilaku pengemudi, seperti tingkat kesadaran akan keselamatan lalu lintas, yang tidak tercakup dalam penelitian ini.

Tabel 2. Hasil Estimasi Regresi Poisson

Variabel	Koefisien β	Nilai p
Intercept β_0	3.14	<0.001
Jumlah Kendaraan	0.02	0.004
Panjang Jalan	0.05	0.002
Kepadatan Penduduk	-0.01	0.252

Berdasarkan Tabel 2, dapat disimpulkan bahwa variabel jumlah kendaraan dan panjang jalan memiliki nilai p yang signifikan ($p < 0.05$), sementara kepadatan penduduk tidak berpengaruh signifikan terhadap jumlah kecelakaan ($p > 0.05$). Dengan demikian, hasil analisis menunjukkan bahwa model regresi Poisson dapat digunakan untuk memprediksi jumlah kecelakaan lalu lintas berdasarkan faktor-faktor yang signifikan seperti jumlah kendaraan dan panjang jalan. Penelitian ini juga memberikan informasi penting bagi pengambil kebijakan untuk merancang strategi keselamatan lalu lintas yang lebih efektif.

4 KESIMPULAN

Penelitian ini bertujuan untuk menganalisis pengaruh jumlah kendaraan bermotor, panjang jalan, dan kepadatan penduduk terhadap jumlah kecelakaan lalu lintas di Indonesia pada periode 2014–2023 dengan menggunakan pendekatan regresi Poisson. Berdasarkan hasil analisis, ditemukan bahwa jumlah kendaraan dan panjang jalan secara signifikan berpengaruh positif terhadap jumlah kecelakaan, sedangkan kepadatan penduduk tidak menunjukkan pengaruh yang signifikan.

Temuan ini memperkuat hipotesis awal bahwa peningkatan jumlah kendaraan dan infrastruktur jalan cenderung meningkatkan potensi terjadinya kecelakaan lalu lintas. Model regresi Poisson yang digunakan dalam penelitian ini berhasil menggambarkan pola hubungan tersebut dengan baik, dan dapat dijadikan alat bantu untuk memprediksi jumlah kecelakaan di masa depan berdasarkan perkembangan faktor-faktor sosial ekonomi.

Penelitian ini relevan dalam konteks kebijakan transportasi dan keselamatan jalan di Indonesia, mengingat pertumbuhan kendaraan yang pesat perlu diimbangi dengan peningkatan kualitas dan pengawasan infrastruktur jalan, serta regulasi lalu lintas yang lebih ketat. Oleh karena itu, disarankan agar pemerintah tidak hanya fokus pada pembangunan fisik infrastruktur jalan, tetapi juga meningkatkan sistem keselamatan lalu lintas, edukasi berkendara yang aman, serta pengendalian pertumbuhan kendaraan bermotor.

UCAPAN TERIMA KASIH

Puji syukur saya panjatkan ke hadirat Tuhan Yang Maha Esa karena atas rahmat dan karunia-Nya, saya dapat menyelesaikan karya ilmiah ini dengan baik. Dengan penuh rasa hormat dan penghargaan, saya menyampaikan ucapan terima kasih yang sebesar-besarnya kepada Ibu Emeylia Safitri, selaku dosen pembimbing Karya Ilmiah, yang telah dengan sabar membimbing, memberikan arahan, masukan, serta motivasi selama proses penyusunan karya ilmiah ini. Bimbingan dan perhatian yang Ibu berikan sangat berarti dan menjadi bekal berharga dalam penyelesaian tugas ini.

DAFTAR PUSTAKA

- Al Haris and Sunendiari and Satyahadewi, N. (2022). Negative Binomial Regression and Generalized Poisson Regression Models on the Number of Traffic Accidents in Central Java. *Jurnal Matematika dan Statistika*, 205--214.
- Cameron, A. C. (2013). *Regression Analysis of Count Data*. Cambridge University Press.
- Hilbe, J. M. (2011). *Negative Binomial Regression*. Cambridge University Press.
- Long, J. S. (2014). *Regression Models for Categorical Dependent Variables Using Stata*. Stata Press.
- Organization, W. H. (2023). *Global Status Report on Road Safety 2023*. World Health Organization.

- Pelangi, M. a. (2020). Analisis kasus kematian akibat kecelakaan lalu lintas menggunakan regresi Poisson dan ZIP di Pontianak. *Jurnal Sains dan Teknologi*, 123--135.
- Polri, K. (2023). Laporan Kecelakaan Lalu Lintas 2023. Kepolisian Negara Republik Indonesia.
- Rakhmat, L. A. (2012). Pengembangan model prediksi kecelakaan lalu lintas pada Jalan Tol Purbaleunyi. *Jurnal Transportasi*, 101--110.
- RI, K. P. (2023). Statistik Perhubungan 2023. Kementerian Perhubungan RI.
- Statistik, B. P. (2023). Statistik Transportasi Darat 2023. Badan Pusat Statistik.