p-ISSN 2338 – 3216
e-ISSN 2528 – 1070

JURNAL STATISTIKA
UNIVERSITAS MUHAMMADIYAH SEMARANG
14(1) 2026: 32-42

APPLICATION OF BINARY LOGISTIC REGRESSION USING TIME-BASED SAMPLING IN ANALYSIS OF RISK FACTORS FOR MOTORCYCLE TRAFFIC VIOLATIONS IN MEDAN CITY

Graceya Zagita Manik[1*], Irgie Attaurrazaq[2], Donni Ramadhan Siregar[3], Katrin Jenny Sirait[4]
[1,2,3,4] Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Sumatera Utara, Indonesia

***e-mail**: graceyazmanik@gmail.com

Article Info:
Received: March 31, 2026
Accepted: May 25, 2026
Available Online: June 2, 2026

Keywords:
traffic violations; binary logistic regression; Time-Based Sampling; motorcycle; Medan City

Abstract: Traffic violations by motorcyclists are a major contributor to accidents in Medan City. This research sought to examine the frequency of violations and the factors affecting the likelihood of such violations among riders at three intersections in Medan City: Dr. Mansyur, Setia Budi, and Fly Over Jamin Ginting. A cross-sectional quantitative design and time-based sampling was used. Observations were conducted over three days in two sessions (session 1 02:00-03:00 PM and session 2 04:00-05:00 PM WIB), with a sample of 540 motorcyclists. The dependent variable was violation status; independent variables included gender, motorcycle type, rider status, observation time, and location. Binary logistic regression was applied. Results showed a violation rate of 69.8%. Simultaneously, all independent variables had a significant effect ($p < 0.001$). Partially, only rider status was significant ($p < 0.001$; OR - 2.858), meaning riders with a passenger were 2.858 times more likely to violate than solo riders. Gender, motorcycle type, observation time, and location were not significant. The model fitted well (Hosmer Lemeshow test, p - 0.904). In conclusion, rider status is the main factor, so supervision should focus on riders with passengers.

1. INTRODUCTION

Motorcycles are a widely used mode of transportation in Indonesia due to their affordability and efficiency, especially in urban and rural areas [1]. The growth in transportation has led to a continuous increase in the number of motor vehicles. According to the Central Statistics Agency (BPS), the number of two-wheeled vehicles reached over 139 million units in 2024, a 4.81% increase from the previous year [2]. This growth has exacerbated traffic problems such as violations, congestion, and accidents [3].

According to data from the North Sumatra Regional Police Traffic Directorate's Online Traffic Ticket System [4], out of 68 traffic violation cases in the North Sumatra Regional Police jurisdiction, the Medan City Police recorded 32 cases (47%) , the highest number, with violators predominantly being males aged 17 or older and two-wheeled vehicles being the most commonly involved mode of transportation; the most common type of violation was a moderate violation (105 cases), followed by helmet violations (25 cases).

[DOI: 10.14710/JSUNIMUS.14.1.2026.32-42]

Factors contributing to ongoing accidents include a lack of driving ability and skills, as well as negligent behavior such as infrequent helmet use and reckless driving, which result in fatalities and injuries and impact quality of life [5].

The high accident rate is primarily due to human factors, particularly traffic violations and the prevalence of motorcyclists [6]. Although the ETLE system has been implemented at three intersections in Medan, specific and detailed violation data is not publicly available for scientific analysis; there are no periodicals or reports indicating a downward trend in violations, so there is currently no empirical information available for statistical analysis to identify violation patterns, times of occurrence, or characteristics of high-risk drivers. This research gap underscores the need for direct data collection and analysis using binary logistic regression based on Time-Based Sampling, where binary logistic regression was chosen for its ability to analyze the relationship between various factors and the binary decision of "violating" or "not violating" [7], while the binomial distribution is used to model the number of successes in the sample [8]. This combination allows for the quantitative identification of risk factors (gender, time, etc.) as well as periods of high violation risk. The research results are expected to serve as a basis for the police and local governments in developing more effective surveillance strategies and enhancing traffic safety awareness through targeted prevention.

The purpose of this study is:

1. To analyze the level of traffic violations at the Dr. Mansyur, Setia Budi, and Fly Over Jamin Ginting intersections in Medan City.
2. To analyze the factors influencing the probability of traffic violations among motorcyclists at these intersections.

2. LITERATURE REVIEW

2.1. Traffic Violations

Traffic is the process of movement or mobility that utilizes roadways to meet the basic needs of society [9]. A traffic violation is an act or activity committed by a driver or pedestrian that contravenes laws and regulations related to traffic [10]. The most common types of violations include not wearing an SNI-certified helmet, driving against traffic, underage driving, using a cell phone while driving, exceeding the speed limit, and driving under the influence of alcohol [11].

2.2. Binomial Distribution

The dependent variable in this study is binary, indicating whether a motorcyclist commits a traffic violation (1) or not (0). Such outcomes can be conceptually modeled using a binomial distribution, where each observation represents one of two possible outcomes[12]. This framework underlies the use of binary logistic regression, which models the probability of a violation as a function of the explanatory variables.

2.3. Binary Logistic Regression

The relationship that exists between a dichotomous outcome variable and one or multiple predictors (which can be continuous or categorical) is modeled using binary logistic regression[13]. The probability of success, $\pi(x)$, is given by:

[DOI: 10.14710/JSUNIMUS.14.1.2026.32-42]

$$\pi(x) = \frac{\exp\left(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p\right)}{1 + \exp\left(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p\right)} \tag{1}$$

The logit transformation linearizes the model:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \tag{2}$$

2.3.1. Parameter Estimation with MLE

Parameters are estimated by maximizing the likelihood function using the Maximum Likelihood Estimation (MLE) method. The likelihood function for n independent observations is:

$$L(\beta) = \prod_{i=1}^{n} \pi(x_i)^{y_i}[1 - \pi(x_i)]^{1-y_i} \tag{3}$$

The log-likelihood is:

$$\ln L(\beta) = \sum_{i=1}^{n} \{y_i \ln[\pi(x_i)] + (1 - y_i)\ln[1 - \pi(x_i)]\} \tag{4}$$

Maximization is done iteratively (e.g., Newton-Raphson).

2.3.2. Multicollinearity Test

Multicollinearity is assessed using Variance Inflation Factor (VIF) and Tolerance.

$$\text{Tolerance} = 1 - \text{R}_i^{\,2} \tag{5}$$

$$VIF = \frac{1}{1 - \text{R}_i^{\,2}} \tag{6}$$

A $VIF < 10$ and $Tolerance > 0.10$ indicate no multicollinearity.

2.3.3. Simultaneous Test (Likelihood Ratio)

The overall significance of the model is tested using the likelihood ratio statistic:

$$G = -2\log\left(\frac{L_0}{L_1}\right) \tag{7}$$

where L_0 is the log-likelihood of the null model (intercept only) and L_1 is the log-likelihood of the full model. Under the null hypothesis that all coefficients are zero, G follows a chi-square distribution with v degrees of freedom[14].

2.3.4. Partial Test (Wald Test)

The partial test is used to evaluate the significance of individual independent variables on the dependent variable. This test employs the Wald statistic to determine whether each independent variable is suitable for inclusion in the model. The Wald statistic is calculated as:

$$W_j = \left(\frac{\hat{\beta}_j}{\widehat{SE}\hat{\beta}_j}\right)^2 \tag{8}$$

[DOI: 10.14710/JSUNIMUS.14.1.2026.32-42]

where $\hat{\beta}_j$ is he estimated regression coefficient for the j-th independent variable and $\widehat{SE}\hat{\beta}_j$ is the standart error. The decision to reject the null hypothesis is based on comparing the Wald value to the critical value from the normal distribution or using the p-value[14].

2.3.5. Goodness-of-Fit Test (Hosmer–Lemeshow)

The Hosmer–Lemeshow test assesses model fit. The test statistic is:

$$\hat{C} = \sum_{m=1}^{g} \frac{(o_m - n'_m\bar{\pi}_m)^2}{n'_m\bar{\pi}_m(1-\bar{\pi}_m)} \tag{9}$$

where o_m is the observed number of successes in group m, n'_m is the total observations in group m, and $\bar{\pi}_m$ is the average predicted probability. A non-significant p-value (i.e., $p > 0.05$) indicates good model fit, as it fails to reject the null hypothesis that the model fits the data[14].

2.3.6. Odds Ratio

The odds ratio (OR) is used in binary regression models to interpret the parameter coefficients[13]. The odds ratio and the regression coefficient are related in the following way:

$$OR = exp(\beta_j) \tag{10}$$

2.4. Time Based Sampling

Time Based Sampling is an observational technique where data are collected at predetermined intervals to systematically capture variations in phenomena over time[15]. In this study, it is used to select motorcyclists at fixed intervals to ensure representative sampling across the observation period.

2.5. State of the Art

Binary logistic regression has been widely applied in various fields, including the analysis of determinants of stunting in children [16], customer satisfaction [17]; and risk factors for fatal traffic accidents in Indonesia [18]. This study applies binary logistic regression to motorcycle traffic violation data collected using a time-based sampling method at three busy intersections in Medan City. Unlike previous studies that often relied on administrative data or lacked temporal detail, this research directly observes riders during peak hours (02:00–03:00 PM and 04:00–05:00 PM) across three days. The contribution of this study is empirical: it provides quantitative estimates of violation probabilities and identifies rider status (solo vs. with passenger) as a significant risk factor, filling a gap in location-specific traffic violation analysis in Medan where systematic observational data have been limited.

3. METHODOLOGY

3.1. Research Design

The study adopts a quantitative approach combined with a cross-sectional *design*. Data were collected through direct observation using the time-based sampling method.

3.2. Location and Time

Observations were conducted at three intersections in Medan City: Dr. Mansyur, Setia Budi, and Fly Over Jamin Ginting. These locations were selected due to high traffic

[DOI: 10.14710/JSUNIMUS.14.1.2026.32-42]

density and diverse rider characteristics. Data collection took place over three days: Saturday, February 28, 2026; Sunday, March 1, 2026; and Tuesday, March 3, 2026. Two observation sessions were held each day: session 1 (02:00–03:00 PM) and session 2 (04:00–05:00 PM). These times represent peak traffic hours based on local traffic data[19].

3.3. Data Sources and Variables

Primary data were collected via direct observation. The data were collected through direct observations at the Dr. Mansyur Intersection, the Setia Budi Intersection, and the Jamin Ginting Flyover Intersection of motorcyclists passing through these intersections; these are quantitative data. The measurement scale used in this study is nominal data. The unit of analysis in this study is individual motorcycle riders observed at specific time intervals. The data collected includes: Violation Status (violating or not violating); Rider Gender (female/male); Observation Time (session 1/session 2); Motorcycle Type (automatic/geared); Rider Status (solo/with a passenger). The dependent variable was violation status (Y), coded as 1 = violation (not wearing a helmet or running a red light) and 0 = no violation. Independent variables were:

- X_1: Gender (0 = female, 1 = male)
- X_2: Motorcycle type (0 = automatic, 1 = geared)
- X_3: Rider status (0 = solo, 1 = with passenger)
- X_4: Observation time (0 = session 1, 1 = session 2)

Location was coded as two dummy variables with Fly Over Jamin Ginting as the reference category: Lok_Dr.Mansyur (1 = Dr. Mansyur, 0 = otherwise) and Lok_Setiabudi (1 = Setia Budi, 0 = otherwise).

3.4. Sampling

In logistic regression, sample size adequacy is often assessed using Events Per Variable (EPV), defined as the ratio of the number of events to the number of predictors. However, the conventional threshold of $EPV \geq 10$ does not necessarily reflect model quality[20]. With 540 observations and all independent variables, the sample size in this study is considered sufficient, leading to relatively stable parameter estimates. Using the time-based sampling method, the observation interval was set to 2 minutes, calculated as:

$$k = \frac{T}{n} = \frac{60\ minutes}{30\ samples} = 2\ minutes$$

Where T is the session duration (60 minutes) and n is the target sample per session (30). In each 2-minute interval, the first motorcyclist stopping at the red light was recorded. The total sample size was: 3 locations × 2 sessions × 3 days × 30 riders = 540 motorcyclists.

3.5. Data Collection Procedure

Observers positioned themselves safely near the intersections. At the start of each 2-minute interval, the first motorcycle in the queue at a designated lane was selected, alternating lanes each interval. Data recorded included: violation status, gender, motorcycle type, rider status, and time. If multiple violations occurred (e.g., no helmet and running red light), the status was recorded as “violating” ($Y = 1$). To reduce bias, observers rotated lanes and agreed on operational definitions before data collection.

[DOI: 10.14710/JSUNIMUS.14.1.2026.32-42]

3.6. Data Analysis

Data were analyzed using SPSS version 27 and Microsoft Excel. SPSS was chosen because it provides comprehensive features for binary logistic regression, including the likelihood ratio test, Wald test, and Hosmer-Lemeshow test. Excel was used in the initial stage for data entry and variable coding. The analysis was structured in sequential steps. First, multicollinearity was assessed using VIF and tolerance. Next, binary logistic regression analysis was performed, encompassing the simultaneous test (likelihood ratio), partial test (Wald), and goodness-of-fit test (Hosmer-Lemeshow). Model interpretation was then conducted using odds ratios.

4. RESULTS AND DISCUSSION

4.1. Descriptive Statistics

A total of 540 motorcyclists were observed. The overall violation rate was 69.8% (377 violators). By location, Setia Budi had the highest violation rate (75.0%), followed by Fly Over Jamin Ginting (70.0%) and Dr. Mansyur (64.4%).

Table 1. Distribution of Violations by Location

Location	Total Observations	Violators	Percentage
Dr. Mansyur	180	116	64.4%
Setia Budi	180	135	75.0%
Fly Over Jamin Ginting	180	126	70.0%
Total	540	377	69.8%

By time, violations were higher in the session 2 (71.9%) than in the session 1 (67.8%). Male riders had a higher proportion of violations (71.6%) than females (65.1%). Riders with passengers had a much higher violation rate (81.1%) compared to solo riders (61.9%). Interestingly, geared motorcycles had a higher proportion of violations (73.8%) than automatic motorcycles (68.8%), although the absolute number of violators was larger for automatics.

4.2. Binary Logistic Regression Analysis

4.2.1. Multicollinearity Test

Before conducting binary logistic regression, a multicollinearity test was performed. Table 2 presents the Tolerance and VIF values for all independent variables.

Table 2. Multicollinearity Test Results

Variable	Tolerance	VIF	Conclusion
Gender	0,955	1,047	No multicollinearity
Motorcycle type	0,950	1,053	No multicollinearity
Rider status	0,984	1,016	No multicollinearity
Observation time	0,996	1,004	No multicollinearity
Lok_Dr.Mansyur	0,745	1,343	No multicollinearity
Lok_Setiabudi	0,747	1,339	No multicollinearity

[DOI: 10.14710/JSUNIMUS.14.1.2026.32-42]

All independent variables had $Tolerance > 0.10$ and $VIF < 10$, indicating no multicollinearity.

4.2.2. Simultaneous Test (Likelihood Ratio)

The omnibus test was used to assess whether all independent variables jointly affect the dependent variable. Table 3 shows the results.

Table 3. Omnibus Test (Likelihood Ratio)

	Chi-square	df	Sig.
Step	35.380	6	<0.001
Block	35.380	6	<0.001
Model	35.380	6	<0.001

The significant result (p < 0.001) indicates that gender, motorcycle type, rider status, observation time, and location simultaneously have a significant effect on traffic violations.

4.2.3. Partial Test (Wald)

In order to examine the individual significance of each predictor on the outcome, the Wald test was utilized. Table 4 presents the results.

Table 4. Variables in the Equation (Wald Test)

Variable	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP (B)	
							Lower	Upper
Gender (1)	0,344	0,218	2,474	1	0,116	1,410	0,919	2,163
Motorcycle type (1)	-0,284	0,257	1,228	1	0,268	0,752	0,455	1,244
Rider status (1)	1,050	0,212	24,577	1	0,000	2,858	1,887	4,328
Observation time (1)	0,236	0,195	1,463	1	0,226	1,266	0,864	1,854
Lok_Dr.Mansyur (1)	-0,144	0,232	0,385	1	0,535	0,866		1,365
Lok_Setiabudi (1)	0,407	0,244	2,769	1	0,096	1,502		2,425
Constant	0,248	0,343	0,524	1	0,469	1,282		

Note: (1) indicates the category: Male, Geared motorcycle, With passenger, Session 2, Dr. Mansyur intersection, Setiabudi intersection.

Based on the Wald test, only rider status (p < 0.001) was statistically significant. Gender (p = 0.116), motorcycle type (p = 0.268), observation time (p = 0.226), and both location dummies (p = 0.535 for Dr. Mansyur; p = 0.096 for Setia Budi) were not significant at α = 0.05.

4.2.4. Goodness-of-Fit Test

The Hosmer–Lemeshow test was used to assess whether the logistic regression model fits the observed data. Table 5 presents the results.

[DOI: 10.14710/JSUNIMUS.14.1.2026.32-42]

Table 5. Hosmer–Lemeshow Test

Step	Chi-square	df	Sig.
1	3.432	8	0.904

The non-significant result ($p = 0.904 > 0.05$) indicates that the model fits the data well.

4.2.5. Model Equation

Based on the estimated coefficients in Table 4, the binary logistic regression model is:

$$\ln\left(\frac{p}{1-p}\right) = 0.248 + 0.344X_1 - 0.284X_2 + 1.050X_3 + 0.236X_4 - 0.144X_5 + 0.407X_6 \quad (11)$$

where:
X_1 = Rider gender,
X_2 = Motorcycle type,
X_3 = Rider status,
X_4 = Observation time,
X_5 = Lok Dr.Mansyur, and
X_6 = Lok Setiabudi.

4.2.6. Odds Ratios

The odds ratios ($Exp(B)$ in Table 4) are interpreted as follows:

1. Gender ($OR = 1.410$), male riders have 1.410 times higher odds of violating compared to female riders (41% higher). However, this effect is not statistically significant ($p = 0.116$).
2. Motorcycle type ($OR = 0.752$), riders of geared motorcycles have 0.752 times the odds of violating compared to automatic motorcycle riders (24.8% lower odds). This effect is not significant ($p = 0.268$).
3. Rider status ($OR = 2.858$), riders with a passenger have 2.858 times higher odds of violating compared to solo riders (approximately 185.8% higher). This effect is highly significant ($p < 0.001$).
4. Observation time ($OR = 1.266$) The odds of violation in the session 2 are 1.266 times higher than in the session 1 (26% higher). This effect is not significant ($p = 0.226$).
5. Location (Dr. Mansyur vs. Fly Over Jamin Ginting)
   The odds ratio for Lok_Dr.Mansyur is OR = 0.866. This indicates that riders at Dr. Mansyur intersection have 0.866 times the odds of committing a violation compared to those at Fly Over Jamin Ginting (reference location), or approximately 13.4% lower odds. However, with p = 0.535 (> 0.05), this effect is not statistically significant.
6. Location (Setia Budi vs. Fly Over)
   The odds ratio for Lok_Setiabudi is OR = 1.502. This means riders at Setia Budi intersection have 1.502 times higher odds of violating compared to those at Fly Over (approximately 50.2% higher). The p-value is 0.096, which is not significant at the conventional α = 0.05 level, though it approaches marginal significance.

[DOI: 10.14710/JSUNIMUS.14.1.2026.32-42]

4.3. Discussion

The high overall violation rate (69.8%) indicates that traffic rule compliance among motorcyclists at the studied intersections is low. This is consistent with previous findings that human factors dominate traffic accidents[6]. The significant effect of rider status suggests that carrying a passenger increases the likelihood of violations. This may be due to distraction, social pressure, or reduced attention. The odds ratio of 2.858 implies that riders with passengers are nearly three times more likely to commit a violation than solo riders. This finding has practical implications: enforcement and educational campaigns should target riders who carry passengers, especially at high-violation locations like Setia Budi intersection. After controlling for location by including dummy variables for Dr. Mansyur and Setia Budi (with Fly Over Jamin Ginting as reference), the model showed no statistically significant effect of location on violation probability ($p = 0.535$ for Dr. Mansyur; $p = 0.096$ for Setia Budi). Although Setia Budi exhibited a higher odds ratio (OR = 1.502) approaching marginal significance, it did not reach the conventional threshold of $p < 0.05$. This suggests that the observed descriptive differences in violation rates across intersections (64.4%–75.0%) are largely explained by differences in rider composition (e.g., proportion of riders with passengers) rather than location itself. Although gender, motorcycle type, and observation time were not individually significant, the simultaneous test showed they collectively contribute to the model. The lack of significance for gender differs from some studies where male riders were more prone to violations, but it aligns with the notion that other factors may be more influential. The model's excellent fit (Hosmer–Lemeshow test, $p = 0.904$) confirms that the selected variables adequately explain the variation in violation behavior.

However, several methodological limitations should be considered. First, regarding the sampling method, this study used a time-based sampling procedure in which, at every two-minute interval, the first motorcyclist stopping at the red light was selected. While this approach ensured systematic temporal coverage, it may have introduced selection bias. Riders who ran the red light without stopping were not observed at all, potentially underestimating the true violation rate, especially for red-light violations. Furthermore, the first rider in the queue might not represent all riders (e.g., they could be more cautious or more aggressive than those behind). Although lanes were rotated to mitigate this bias, the findings should be interpreted with caution when generalising to all motorcyclists at these intersections. Second, the dependent variable combined two distinct types of violations: not wearing a helmet and running a red light. These behaviours differ substantially in terms of risk (self-harm vs. danger to others) and potential determinants. By aggregating them into a single binary outcome, we may have masked meaningful differences in how independent variables affect each violation type. For instance, rider status (with passenger) might be more strongly associated with helmet non-use, while time of day could be more relevant for red-light running. Consequently, the current model does not allow separate identification of risk factors for each violation. Future research should analyse each infraction separately (or use a multinomial or multiple-binary outcome framework) to provide more targeted policy recommendations.

5. CONCLUSION

The study concludes that the level of traffic violations at the three intersections in Medan City is high, with 69.8% of observed motorcyclists committing a violation. Among the factors analyzed, only rider status significantly influences the probability of violations;

[DOI: 10.14710/JSUNIMUS.14.1.2026.32-42]

riders with passengers have 2.858 times higher odds of violating than solo riders ($p < 0.001$). Gender, motorcycle type, observation time, and location (both Dr. Mansyur and Setia Budi) do not show significant partial effects at $\alpha = 0.05$, although Setia Budi approached marginal significance ($p = 0.096$). The model, after controlling for location, demonstrates excellent fit (Hosmer-Lemeshow $p = 0.904$). This study included location as a control variable using dummy coding. The lack of statistically significant location effects (although Setia Budi approached marginal significance with $p = 0.096$) suggests that intersection-specific factors (e.g., road geometry, traffic light timing) were not major contributors after adjusting for rider characteristics. However, future research with larger samples might detect smaller spatial variations.

ACKNOWLEDGMENT

The authors thank the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Sumatera Utara, and all parties who supported this research.

REFERENCES

[1] A. Muh, R. Rajes, H. Kurniawan, and I. N. Fajri, “Implementasi metode prototype untuk perancangan sistem informasi penyedia jasa montir,” *JUTI J. Tek. Ind.*, vol. 2, no. 1, pp. 32–42, 2023, doi: 10.26798/juti.v2i1.959.

[2] Badan Pusat Statistik, “Jumlah Kendaraan Bermotor Menurut Jenis Kendaraan (Unit), 2022–2024.” Accessed: Feb. 04, 2026. [Online]. Available: https://www.bps.go.id/id/statistics-table/3/VjJ3NGRGa3dkRk5MTlU1bVNFOTVVbmQyVURSTVFUMDkjMw%3D%3D/jumlah-kendaraan-bermotor-menurut-jenis-kendaraan--unit---2024.html

[3] N. M. Muryatma, “Analisis program Yuk Pakai Helm sebagai media promosi kesehatan terhadap kepatuhan pemakaian helm SNI pada pelajar,” *J. Penelit. Kesehat.*, vol. 20, no. 2, pp. 28–39, 2022.

[4] Korps Lalu Lintas Kepolisian Republik Indonesia, “Dashboard penegakan hukum tilang online wilayah Sumatera Utara,” 2025.

[5] R. D. Lintang, J. F. Mallo, and D. Tomuka, “Gambaran luka pada kecelakaan lalu lintas khususnya pengendara kendaraan roda dua yang tidak memakai helm,” *e-CliniC*, vol. 9, no. 28, pp. 265–270, 2021.

[6] R. D. Anwar, “Kepribadian dan pelanggaran pada pengendara sepeda motor,” *Indones. J. Public Heal.*, vol. 12, no. August, pp. 179–189, 2017, doi: https://doi.org/10.20473/ijph.v12i2.2017.179-189.

[7] Z. R. Pangesti and C. Sormin, “Analisis churn pelanggan telekomunikasi menggunakan regresi logistik biner,” *J. Ris. Mhs. Mat.*, vol. 5, no. 2, pp. 142–158, 2025, doi: 10.18860/jrmm.v5i2.38982.

[8] N. A. A. Syifaa and A. Saputri, “Simulasi distribusi binomial untuk analisis Python,” *J. Kaji. Ilm. Interdisiplinier*, vol. 9, no. 4, pp. 44–49, 2025.

[9] I. G. Krisna, I. N. G. Sugiarta, and I. N. Subamiya, “Tindak pidana pelanggaran lalu lintas dan upaya penanggulangannya pada masa pandemi Covid-19,” *J. Konstr. Huk.*, vol. 2, no. 2, pp. 338–343, 2021, doi: 10.22225/jkh.2.2.3233.338-343.

[DOI: 10.14710/JSUNIMUS.14.1.2026.32-42]

[10] Rahayu Nurfauziah and Hetty Krisnan, "Perilaku pelanggaran lalu lintas oleh remaja ditinjau dari perspektif konstruksi sosial," *J. Kolaborasi Resolusi Konflik*, vol. 3, no. 1, pp. 77–84, 2021.

[11] B. Amyrulloh and Samuji, "Analisis penyebab pelanggaran lalu lintas oleh pengendara kendaraan," *J. Ilmu Sos. dan Hum.*, vol. 2, pp. 81–103, 2024.

[12] Nurhaliza, D. Rahmi, A. Kurniati, and S. Yuniati, "Model distribusi binomial dalam mengukur probabilitas keberhasilan uji coba kualitas layanan sistem informasi," *J. Teknol. dan Manaj. Ind. Terap.*, vol. 3, no. 4, pp. 405–410, 2024, doi: 10.55826/jtmit.v3i4.506.

[13] A. A. Putri, J. Titaley, and D. T. Salaki, "Kecenderungan gejala maag pada mahasiswa Jurusan Matematika FMIPA UNSRAT," *J. Mat. dan Apl.*, vol. 11 (1), pp. 38–43, 2022.

[14] B. L. Sihombing, M. S. Paendong, and Y. Langi, "Analisis regresi logistik untuk menentukan tingkat kepuasan masyarakat terhadap pelayanan pembuatan Kartu Tanda Penduduk elektronik (e-KTP) di Kota Manado," *J. Lppm Bid. Sains Dan Teknol.*, vol. 9, no. 2, pp. 102–111, 2024, doi: 10.35801/jlppmsains.9.2.2024.57794.

[15] L. Degroote, A. Desmet, I. De Bourdeaudhuij, D. Van Dyck, and G. Crombez, "Content validity and methodological considerations in ecological momentary assessment studies on physical activity and sedentary behaviour: A systematic review," *Int. J. Behav. Nutr. Phys. Act.*, vol. 17, no. 1, pp. 1–13, 2020, doi: 10.1186/s12966-020-00932-9.

[16] E. W. Pamungkas and S. Muchlisoh, "Penerapan regresi logistik biner untuk menganalisis determinan status stunting pada balita di Provinsi Sulawesi Tengah Tahun 2018," *Semin. Nas. Off. Stat. 2023*, vol. 2018, no. 2014, pp. 623–632, 2023.

[17] F. Sagai, D. Hatidja, and C. E. Mongi, "Penerapan Regresi Logistik untuk Menganalisis Tingkat Kepuasan Mahasiswa FMIPA UNSRAT Manado sebagai Pelanggan Indrive," *d'Cartesian*, vol. 13, no. 2, pp. 85–91, 2024, doi: 10.35799/dc.13.2.2024.55730.

[18] S. Amalia, A. Putri, M. D. Sianturi, R. O. Hutapea, A. S. Ndruru, and A. Arnita, "Binary logistic regression analysis of traffic accident risk factors in Indonesia," *Param. J. Mat. Stat. dan Ter.*, vol. 4, no. 1, pp. 63–72, 2025, doi: 10.30598/parameterv4i1pp63-72.

[19] TomTom, "TomTom Traffic Index – Medan, Indonesia." Accessed: Feb. 28, 2026. [Online]. Available: https://www.tomtom.com/traffic-index/city/medan

[20] M. van Smeden *et al.*, "Sample size for binary logistic prediction models: Beyond events per variable criteria," *Stat. Methods Med. Res.*, vol. 28, no. 8, pp. 2455–2474, 2019, doi: 10.1177/0962280218784726.

[DOI: 10.14710/JSUNIMUS.14.1.2026.32-42]