CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 10 (2) (2025), Pages 519-531
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Comparison of Newton Raphson and Stochastic Gradient Descent for Traffic Accident Severity Modeling in Malang

Aldi Rahmad Nur Fauzi*, Sobri Abusini, and Corina Karim

Department of Mathematics, Faculty of Mathematics and Natural Sciences, Brawijaya University, Malang, Indonesia

Abstract

This study discusses a comparison between two optimization methods, Newton–Raphson and Stochastic Gradient Descent (SGD), in binary logistic regression modeling to analyze the severity of traffic accidents in Malang Regency. Parameter estimation was carried out using both methods to assess their effectiveness in achieving convergence and producing a well-fitted model. The results show that the Newton–Raphson method failed to achieve convergence despite its fast iteration speed, while the SGD method successfully converged, although it required a large number of iterations. Model evaluation was conducted by examining model fit through log-likelihood values and the Akaike Information Criterion (AIC). The results indicate that the SGD method produced a better-fitting model compared to Newton–Raphson. Additionally, the regression models from each method identified different predictor variables as significant, suggesting that the choice of optimization approach can influence analytical outcomes. These findings highlight the importance of selecting an appropriate optimization method in logistic regression analysis, particularly for complex and imbalanced accident data.

Keywords: Binary Logistics Regression, Maximum Likelihood, Newton-Raphson, Traffic Accident, and Stochastic Gradient Descent.

1 Introduction

Transportation plays a crucial role in interregional interaction and underpins regional development [1]. In Indonesia, advances in this sector have paralleled the expansion of road infrastructure and the rise in motor vehicle ownership. Malang Regency, in particular, has experienced rapid growth in its economy, tourism, and education sectors, with its population reaching 2,663,862 by 2023 [2]. Such growth has driven increases in mobility and traffic density, resulting in longer travel times, higher pollution levels, chronic congestion, and elevated accident risk [3]. High traffic density exacerbates congestion and accident risk, especially in single-vehicle crashes, which frequently result in fatalities [4]. According to the Directorate General of Land Transportation, traffic accidents are classified into four severity categories: fatal, serious injury, minor injury, and property damage only (PDO) [5]. In Malang Regency, the persistently high accident rate underscores the need for improvements in the transportation system. In 2022, 783 traffic accidents were recorded, and in 2023 there were 906 casualties, comprising 182 fatalities, 13 serious injuries, and 1,294 minor injuries [2].

*Corresponding author. E-mail: aldirahmadnufauzi@gmail.com

Submitted: April 04, 2025 Reviewed: May 03, 2025 Accepted: June 17, 2025
DOI: https://doi.org/10.18860/cauchy.v10i2.33177

Previous studies [6] highlight that accident severity is influenced by factors such as traffic density, driver behavior, and weather conditions that affect visibility and vehicle control. Moreover, [7] point out that novice drivers' lack of experience significantly increases crash risk, as they often have not fully mastered driving techniques or emergency decision-making. To uncover the factors that contribute to accident occurrence, precise analysis of accurate data is essential. [8] argue that police-reported crash data form the foundation for understanding accident patterns and injury severity distributions. This study analyzes traffic accident records from the Malang District Police covering the period 2020–2024. A widely used approach for modeling such discrete outcome data is the binary logit model, which has proven effective in crash severity analysis [9]–[11]. For instance, [12] demonstrated that a binary logit framework can successfully identify relationships between contributing factors and accident risk levels. Accordingly, this research employs a binary logistic regression model to examine the severity of traffic accidents, enabling the assessment of how various independent variables influence the dependent variable, accident severity [13].

The accuracy of any logistic model hinges on the parameter estimation technique. A commonly adopted method is maximum likelihood estimation (MLE), which seeks parameter values that maximize the likelihood function representing the probability of the observed data under those parameters [14]. Since the maximization of a likelihood function rarely yields closed-form solutions, numerical optimization techniques are required [15]. Two prevalent algorithms are Newton–Raphson and Stochastic Gradient Descent (SGD). Newton–Raphson leverages both first (gradient) and second (Hessian) derivatives to iteratively approach the likelihood maximum, but can fail to converge if initial parameters are poorly chosen or if the likelihood surface exhibits plateaus, extreme values, or Hessian singularities. SGD minimizes the negative log-likelihood through iterative, sample-based gradient updates, offering scalability at the expense of slower convergence. The convergence rate, the speed at which an algorithm attains an acceptably accurate solution, is influenced by problem convexity, smoothness, and imposed constraints [16], [17].

Given the pivotal role of estimation method selection in binary logistic modeling, this research provides novel contributions by systematically comparing the performance of Newton–Raphson and SGD algorithms in analyzing traffic accident severity. The study's originality lies in its empirical evaluation of these optimization techniques within the context of Indonesia's rapidly developing transportation landscape, particularly in Malang Regency where accident patterns remain understudied. The outcomes are expected to yield recommendations for the most efficient parameter estimation approach in crash severity studies and inform more effective traffic safety strategies, while also addressing the methodological gap in algorithm selection for transportation research in developing regions.

2 Methods

This study employs a binary logistic regression model to examine the factors influencing traffic accidents. The binary logistic regression framework is chosen because it is well-suited for handling a dichotomous dependent variable, here coded as (1) for serious injury/fatality and (0) for minor injury. The model links the probability of an event to the independent variables via the logit function. The general form of the binary logistic regression equation is given by [18], [12]:

$$\text{logit}(p_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}, \tag{1}$$

where:

- β_0: intercept.
- $\beta_1, \beta_2, \ldots, \beta_p$: regression coefficients for the independent variables.
- $X_{i1}, X_{i2}, X_{i3}, \ldots, X_{ip}$: variabel independen atau prediktor untuk observasi ke-i.

Next, before computing Eq. 1, all data obtained were encoded via one-hot encoding to transform categorical predictors into numerical form. From equation Eq. 1, the probability p of an accident event, assuming the parameters $\beta_0, \beta_1, \ldots, \beta_p = 0$, is given by:

$$\text{Prob}(p) = \frac{e^{\text{logit}(p_i)}}{1 + e^{\text{logit}(p_i)}}. \tag{2}$$

Once the probability p is obtained, the next step is to estimate the parameter vector $\beta_0, \beta_1, \ldots, \beta_p$ using the Maximum Likelihood Estimation (MLE) approach. The corresponding log-likelihood function for n observations is:

$$L(\beta) = \sum_{i=1}^{n} \left[Y_i \eta_i - \ln(1 + e^{\eta_i}) \right]. \tag{3}$$

With:

- Y_i : Binary response variable for observation i.
- η_i : $\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{ip}$.

Since Eq. 3 cannot be solved explicitly, optimization methods are used: Newton-Raphson and Stochastic Gradient Descent (SGD).

a. **Newton-Raphson.** This method utilizes the first and second derivatives of the log-likelihood function. The first and second derivatives are:

$$\frac{\partial L(\beta)}{\partial \beta_k} = \sum_{i=1}^{n} X_{ik}(Y_i - p_i), \quad \text{for } k = 0, 1, \ldots, p \tag{4}$$

$$\frac{\partial^2 L(\beta)}{\partial \beta_k \partial \beta_j} = -\sum_{i=1}^{n} X_{ik} X_{ij} p_i (1 - p_i), \quad \text{for } j = 0, 1, \ldots, p \tag{5}$$

in matrix form, Eq. 4 and Eq. 5 become:

$$\frac{\partial L(\beta)}{\partial \beta} = X^\top (Y - P). \tag{6}$$

$$\frac{\partial^2 L(\beta)}{\partial \beta_k \partial \beta_j} = -X^\top V X. \tag{7}$$

Therefore, the parameter update rule is given by:

$$\hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} + \left(-X^\top V X \right)^{-1} X^\top (Y - P). \tag{8}$$

b. **Stochastic Gradient Descent (SGD).** This method uses only the negative gradient of the log-likelihood function. Before applying this method, the dataset is shuffled randomly to avoid model bias from fixed sequences. The parameter update rule is:

$$\hat{\beta}^{(i+1)} = \hat{\beta}^{(i)} - \alpha \left(-\sum_{i=1}^{n} X_{ik}(Y_i - p_i) \right)_{\text{mini-batch}}. \tag{9}$$

Here, α is the learning rate (set to 0.0001) and the mini-batch size is 64. Iteration stops when convergence is achieved:

$$|\hat{\beta}^{(i+1)} - \hat{\beta}^{(i)}| < \varepsilon, \quad \text{where } \varepsilon = 10^{-6}$$

After the iteration is completed, several tests will be conducted, namely:

a. **Logistic Regression Model Fit Test.** The model fit test is used to evaluate whether the model is appropriate for the data, i.e., whether the observed values obtained are the same or close to the expected values from the model. In logistic regression, the method for testing model adequacy is measured using the chi-square value through the Hosmer and Lemeshow test. This test observes the Goodness of Fit (GoF), which is measured using the chi-square value at $\alpha = 5\%$ significance level. The statistical test used is the chi-square test based on the following equation: [18]

$$\hat{C} = \sum_{k=1}^{g} \frac{(O_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)}. \tag{10}$$

With:

O_k **:** Observed frequency in group k.

$\bar{\pi}_k$ **:** Average predicted probability in group k.

n'_k **:** Number of observations in group k.

g **:** Number of groups

Hypotheses:

- H_0: The model fits the data well.
- H_1: The model does not fit the data well.

Decision Rule:

- If p-value $> \alpha = 0.05$, accept H_0 (model fits well).
- If p-value $< \alpha = 0.05$, reject H_0 (poor fit).

b. **McFadden's Pseudo R^2 Test.** This test is conducted to determine and predict the extent or importance of the contribution made by the independent variables collectively toward the dependent variable. The value of the coefficient of determination ranges from 0 to 1. If the value approaches 1, it means the independent variables provide nearly all the information needed to predict the dependent variable. Conversely, if the R^2 value is low, it means the independent variables' ability to explain the dependent variable is quite limited. The statistical test used is: [19]

$$R^2 = 1 - \frac{\text{LL}_{\text{model}}}{\text{LL}_{\text{null}}}. \tag{11}$$

With:

$\mathbf{LL_{model}}$ **:** Log-likelihood with predictors.

$\mathbf{LL_{null}}$ **:** Log-likelihood without predictors.

c. **Partial Test for Parameter Significance.** Partial Test to Examine the Significance of Each Parameter in the Model (Wald Test). This test aims to determine whether each independent variable significantly affects the probability of an event occurring . The commonly used significance level is $\alpha = 5\%$. The statistical test used is: [19]

$$W^2 = \frac{\beta_i^2}{\text{SE}(\beta_i)^2}. \tag{12}$$

Hypotheses:

- H_0: Variable X_i has no significant effect on Y.
- H_1: Variable X_i has a significant effect on Y.

Decision Rule:

- If p-value < 0.05, reject H_0.
- If p-value ≥ 0.05, accept H_0.

d. **Coefficient Interpretation Test.** Interpretation of Parameter Coefficients. This test aims to understand the effect of predictor variables. If the odds ratio > 1, the likelihood of the event increases with an increase in variable x. If the odds ratio < 1, the likelihood of the event decreases with a change in x. The statistical test used is: [20]

$$\text{OR} = \frac{\text{odds}_A}{\text{odds}_B} = \frac{\pi_A/(1-\pi_A)}{\pi_B/(1-\pi_B)}. \tag{13}$$

With:

$\mathbf{odds}_A$ **:** Probability of occurrence when variable = 1.

$\mathbf{odds}_B$ **:** Probability of occurrence when variable = 0.

e. **AIC (Akaike Information Criterion) Test.** This comparison aims to evaluate model performance based on the Akaike Information Criterion (AIC) index using two different optimization methods: Newton-Raphson and Stochastic Gradient Descent (SGD). A lower AIC value indicates a more preferred model. In other words, the best model is the one with the smallest AIC value [21]. The AIC formula is as follows: [22]

$$AIC = 2k - 2\ln(L). \tag{14}$$

Description:

k = Number of parameters in the model.

L = Maximum value of the likelihood function for the model.

This study analyzes traffic accident records obtained from the Malang Regency Police Department spanning the years 2022 to 2024, encompassing a total of 3,977 documented cases. The investigation examines multiple factors potentially associated with accident severity, which are categorized and described in detail below:

Table 1: Researched variables

Variable Name	**Scale**	**Description**	**Variable Name**	**Scale**	**Description**
Accident Type (Y)	0	Minor injury	Gender (X_1)	1	Male
	1	Seriously injured or dead		2	Female
Age (X_2)	1	$<$ 18 years	Education (X_3)	1	Primary School
	2	18–30 years		2	Middle School
	3	31–50 years		3	High School
	4	$>$ 51 years		4	Higher Education
	5	Other		5	Other
Jobs (X_4)	1	Civil Servant	Vehicle Types (X_5)	1	2-wheel
	2	Soldier		2	3-wheel
	3	Police		3	4-wheel
	4	Employee		4	$>$ 4-wheel
	5	Student	Driving Licenses (X_6)	1	Do Not Have
	6	College Student		2	Have
	7	Driver	Weather (X_7)	1	Rainy
	8	Trader		2	Shine
	9	Farmer			
	10	Labor			
	11	Other			
Road Types (X_8)	1	Arterial	Region Level (X_{10})	1	National
	2	Collector		2	Province
	3	Local		3	Cities
	4	Toll		4	District
	5	Neighborhood		5	Village
Scene (X_9)	1	Settlement	Day of Incident (X_{12})	1	Monday
	2	Urban		2	Tuesday
	3	Market		3	Wednesday
	4	Tourist Attraction		4	Thursday
	5	Entertainment Venues		5	Friday
	6	Other		6	Saturday
				7	Sunday
Time of Incident (X_{11})	1	1 am–6 am	Month of Incident (X_{13})	1	January
	2	6.01 am–12 pm		2	February
	3	12.01 pm–6 pm		3	March
	4	6.01 pm–12 am		4	April
				5	May
				6	June
				7	July
				8	August
				9	September
				10	October
				11	November
				12	December

3 Results and Discussion

The traffic accident data were obtained from the Malang District Police (Polres Kabupaten Malang) covering the period from 2022 to 2024. The dataset consists of 13 predictor variables and 3,977 respondents. A portion of the raw data is presented below:

Table 2: Data of traffic accidents from 2022 to 2024

Respondent	Gender	Age	Education	Occupation	...	Accident Category
1	Female	Other	Junior High School	Student	...	Minor Injury
2	Male	31–50	Higher Education	Police Officer	...	Minor Injury
3	Male	> 51	Junior High School	Merchant	...	Serious Injury
⋮	⋮	⋮	⋮	⋮	⋱	⋮
3799	Male	18–30	Senior High School	Police Officer	...	Serious Injury

The above data were then transformed using one-hot encoding into matrix form. The resulting matrix **X** can be represented as:

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

Each row of matrix **X** corresponds to a respondent, and each column represents a binary indicator for a particular category of the predictor variables (e.g., gender, age group, education level, etc.). Next, parameter estimation is conducted using different optimization techniques.

3.1 Newton-Raphson Optimization Method

Although Eq. 3 cannot be solved explicitly, the Newton-Raphson method is applied by referring to Equation Eq. 8. The first iteration result is obtained using the Newton-Raphson optimization method, assuming that the initial values of the parameters $\beta_0, \beta_1, \beta_2, \ldots, \beta_{13}$ are all set to 0. The result is as follows:

$$\hat{\beta}_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -1.25 & 1.05 & -2.72 & 0.53 & 2.72 & 7.11 & -2.86 & \cdots & 5.71 \\ 8.28 & -1.01 & 1.34 & -0.53 & -1.38 & -2.45 & 2.48 & \cdots & 1.13 \\ 2.33 & -5.16 & 6.69 & -1.98 & -6.69 & -2.91 & 1.84 & \cdots & -1.09 \\ -1.85 & 2.37 & -3.79 & -4.00 & 5.24 & 2.22 & -9.08 & \cdots & 0.41 \\ -2.33 & 5.16 & -6.69 & 1.98 & 6.69 & 2.91 & -1.84 & \cdots & 1.09 \\ -5.19 & 6.71 & -3.27 & 2.16 & 3.27 & -1.25 & -6.12 & \cdots & -1.71 \\ 9.06 & -5.55 & 1.38 & 0.40 & 1.78 & 4.08 & 2.04 & \cdots & -2.84 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1.69 & -1.38 & -2.39 & 6.53 & -4.08 & -0.54 & -0.27 & \cdots & -5.71 \end{bmatrix} \begin{bmatrix} -1451.5 \\ 0 \\ -485 \\ 0 \\ -584 \\ -321 \\ -353 \\ \vdots \\ -78.5 \end{bmatrix} = \begin{bmatrix} -1.47363961 \\ -0.01103663 \\ -0.01906851 \\ 0.0040346 \\ 0.08204725 \\ -0.02644057 \\ 0.0508358 \\ \vdots \\ -0.01664946 \end{bmatrix}$$

The next step involves checking the convergence criterion: $\left|\log L(\hat{\beta}_{n+1}) - \log L(\hat{\beta}_n)\right| < 10^{-6}$. The resulting value is $2.297784 > 10^{-6}$, indicating that the iteration has not yet converged and further iterations are required. In the subsequent iterations, the parameter estimates were computed using **Python software**, and the results are as follows:

Table 3 shows that from the 16^{th} iteration onward, the result of $\left|\log L(\hat{\beta}_{n+1}) - \log L(\hat{\beta}_n)\right|$ remains constant at 2.236068, and since $2.236068 > 10^{-6}$, this indicates that the Newton-Raphson method did not achieve convergence.

3.2 Stochastic Gradient Descent (SGD) Optimization Method

Just like the Newton-Raphson method, Eq. 3 cannot be solved explicitly. Therefore, the Stochastic Gradient Descent (SGD) method is applied by referring to Eq. 9, with a learning rate of $\alpha = 0.0001$

Table 3: Newton-Raphson algorithm result

Iteration	$\left\vert \log L(\hat{\beta}_{n+1}) - \log L(\hat{\beta}_n) \right\vert$	Log likelihood
1	2,297784	1.596,043306
2	2,027119	1.546,439964
⋮	⋮	⋮
16	2,236068	1.544,302243
17	2,236068	1.544,302243

and a mini-batch size of 64. Prior to applying the SGD method, the design matrix X must be randomly shuffled. The shuffled matrix $X_{\text{random shuffle}}$ is as follows:

$$X_{\text{random shuffle}} = \begin{bmatrix} 1 & 0 & 1 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

The result of the first iteration using the SGD optimization method is as follows:

$$\hat{\beta}_2^{(1)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} - 0.001 \cdot \begin{bmatrix} 2161.7 \\ 0 \\ 7184.8 \\ 8690.8 \\ 4666.9 \\ \vdots \\ 2284.4 \end{bmatrix} = \begin{bmatrix} -2.1617 \\ 0 \\ -7.1848 \\ -8.6908 \\ -4.6669 \\ \vdots \\ -2.2844 \end{bmatrix}$$

Next, convergence is evaluated using the condition: $|\log L(\hat{\beta}_{n+1}) - \log L(\hat{\beta}_n)| < 10^{-6}$. The obtained value is $0.004627 > 10^{-6}$, indicating that the algorithm has not yet converged, and further iterations are required. The subsequent iteration results, obtained using Python software, are presented in the following table.

Table 4: Stochastic gradient descent algorithm result

Iteration	$\left\vert \log L(\hat{\beta}_{n+1}) - \log L(\hat{\beta}_n) \right\vert$	Log likelihood
1	0,004627	1.597,130753
51	0,001751	1.551,329936
101	0,001684	1.548,934941
151	0,002607	1.547,943000
⋮	⋮	⋮
83.951	0,000048	1.545,498095

The table above shows that convergence was achieved after 83,951 iterations—a relatively large number of iterations. The value of $\left\vert \log L(\hat{\beta}_{n+1}) - \log L(\hat{\beta}_n) \right\vert = 0.000048$ satisfies the convergence criterion, since $0.000048 < 10^{-6}$. This indicates that the Stochastic Gradient Descent (SGD) method has successfully achieved convergence.

3.3 Model Fit Test

The model fit can be evaluated using Eq. 10 , and the results are presented in the following table:

Table 5: Hosmer and lemeshow test

Method	Chi-square	df	Sig.
NR	15,51	54	0,2722
SGD	15,51	54	0,3488

The table above shows that the p-values of the Hosmer and Lemeshow goodness-of-fit test for both methods are greater than the significance level $\alpha = 0.05$. Therefore, the null hypothesis H_0 is accepted, indicating that the model is able to adequately predict the observed values, or in other words, the model is considered to have a good fit.

Table 5 shows that the p-values of the Hosmer and Lemeshow goodness-of-fit test for both methods are greater than the significance level $\alpha = 0.05$. Therefore, the null hypothesis (H_0) is accepted, indicating that the model is capable of adequately predicting the observed values. In other words, the model is considered to have a good fit.

3.4 Coefficient of Determination Test

The coefficient of determination test was conducted using Eq. 11, and the results are shown in the following table:

Table 6: Coefficient of determination Test

Method	LL_{Model}	LL_{null}	Pseudo R^2
NR	-1.544,302	-1.574,2	0,019005
SGD	-1.545,498	-1.574,2	0,018246

The Pseudo R^2 values in Table 6 indicate that the Newton-Raphson (NR) method yields a value of 0.019, while the Stochastic Gradient Descent (SGD) method produces a value of 0.018. These values suggest that approximately 1.9% and 1.8% of the variability in the occurrence of accidents (as the dependent variable) can be explained by the independent variables included in each model. In other words, the predictors used in the models contribute only modestly to the variation in accident occurrences. The remaining 98.1% to 98.2% of the variability is likely attributable to other factors not accounted for in the current models.

3.5 t-Test

The results of the significance test of independent variables on the dependent variable using the Newton-Raphson (NR) method are presented in Table 7.

Based on the results of the t-test in Table 7 , with the NR method, the independent variable that significantly affects the dependent variable is the occupation variable in the student category, represented by $X_4(5)$, with a significance value of $0,022 < 0,05$. Therefore, the null hypothesis H_0 is rejected and the alternative hypothesis H_1 is accepted, indicating that occupation significantly influences traffic accidents. From these results, a binary logistic regression model can be formed as follows:

$$\ln\left(\frac{P(y=1)}{1-P(y=1)}\right)_{NR} = \beta_0 + \beta_4 X_4(5) = -1,8960 - 0,5832 X_4.$$

Description:

- β_0: intercept,
- $\beta_4 X_4(5)$: variable representing occupation in the student category.

Table 7: Parameter estimation of the NR and SGD Method

Modifier	Category	Coefficient	Parameter Estimation		p-Value	
			NR	SGD	NR	SGD
Constant	-	β_0	-1.8960	-1.9991	0.000	0.001
Gender	1	-	Category as a comparison			
	2	$\beta_1 X_1(1)$	-0.0242	-0.0228	0.809	0.820
Age	1	-	Category as a comparison			
	2	$\beta_2 X_2(1)$	-0.0403	0.0098	0.824	0.957
	3	$\beta_2 X_2(2)$	0.0087	0.0531	0.964	0.785
	4	$\beta_2 X_2(3)$	0.1706	0.2128	0.359	0.258
	5	$\beta_2 X_2(4)$	-0.0632	0.0085	0.823	0.976
Education	1	-	Category as a comparison			
	2	$\beta_3 X_3(1)$	0.1132	0.1848	0.667	0.408
	3	$\beta_3 X_3(2)$	-0.0250	0.0346	0.910	0.878
	4	$\beta_3 X_3(3)$	-0.3416	-0.2890	0.246	0.332
	5	$\beta_3 X_3(4)$	0.1350	0.2181	0.712	0.552
Jobs	1	$\beta_3 X_3(1)$	-0.6888	-0.7982	0.262	0.215
	2	$\beta_3 X_3(2)$	-19.0213	-0.4683	0.999	0.774
	3	$\beta_3 X_3(3)$	1.2262	10.829	0.106	0.160
	4	$\beta_3 X_3(4)$	-0.0706	0.0665	0.536	0.561
	5	-	Category as a comparison			
	6	$\beta_4 X_4(5)$	-0.5832	-0.5406	0.022	0.032
	7	$\beta_4 X_4(6)$	-28.1361	-0.1003	1.000	0.975
	8	$\beta_4 X_4(7)$	-18.2557	-0.4009	0.999	0.815
	9	$\beta_4 X_4(8)$	-1.5170	-13.002	0.139	0.162
	10	$\beta_4 X_4(9)$	-28.5053	-0.1320	1.000	0.962
	11	$\beta_4 X_4(10)$	-0.0242	-0.0244	0.941	0.941
Vehicle Types	1	-	Category as a comparison			
	2	$\beta_5 X_5(1)$	0.0000	0.0000	0.000	0.000
	3	$\beta_5 X_5(2)$	0.0173	0.0227	0.901	0.870
	4	$\beta_5 X_5(3)$	0.1188	0.1223	0.532	0.520
Driving Licenses	1	$\beta_6 X_6(1)$	0.0376	0.0421	0.743	0.713
	2	-	Category as a comparison			
Weather	1	$\beta_7 X_7(1)$	-0.0274	-0.0257	0.801	0.814
	2	-	Category as a comparison			
Road Types	1	-	Category as a comparison			
	2	$\beta_8 X_8(1)$	-0.0274	0.0443	0.670	0.644
	3	$\beta_8 X_8(2)$	0.1751	0.2025	0.746	0.712
	4	$\beta_8 X_8(3)$	-0.4142	-0.4161	0.227	0.226
	5	$\beta_8 X_8(4)$	-17.8845	-0.2546	0.999	0.902
Scene	1	-	Category as a comparison			
	2	$\beta_9 X_9(1)$	-0.2329	-0.2483	0.148	0.124
	3	$\beta_9 X_9(2)$	0.4330	0.3874	0.123	0.171
	4	$\beta_9 X_9(3)$	0.2463	0.2063	0.753	0.796
	5	$\beta_9 X_9(4)$	0.0000	0.0000	0.000	0.000
	6	$\beta_9 X_9(5)$	0.2285	0.2076	0.148	0.322
Region Level	1	-	Category as a comparison			
	2	$\beta_{10} X_{10}(1)$	0.2367	0.2646	0.109	0.072
	3	$\beta_{10} X_{10}(2)$	0.0000	0.0000	0.000	0.000
	4	$\beta_{10} X_{10}(3)$	0.0412	0.0471	0.697	0.657
	5	$\beta_{10} X_{10}(4)$	0.0391	0.0518	0.943	0.925
	6	$\beta_{10} X_{10}(5)$	-0.0095	-0.0133	0.944	0.922

Modifier	Category	Coefficient	Parameter Estimation		p-Value	
			NR	SGD	NR	SGD
Time of Incident	1	$\beta_{11}X_{11}(1)$	-0.0095	-0.0133	0.944	0.922
	2	-	Category as a comparison			
	3	$\beta_{11}X_{11}(2)$	0.0225	0.0215	0.854	0.861
	4	$\beta_{11}X_{11}(3)$	-0.1811	-0.1769	0.164	0.174
Day of Incident	1	$\beta_{12}X_{12}(1)$	0.2275	0.2218	0.110	0.119
	2	$\beta_{12}X_{12}(2)$	-0.0155	-0.0583	0.744	0.712
	3	$\beta_{12}X_{12}(3)$	0.1680	0.1660	0.350	0.355
	4	$\beta_{12}X_{12}(4)$	0.1193	0.1153	0.480	0.495
	5	$\beta_{12}X_{12}(5)$	-0.0708	-0.0843	0.766	0.723
	6	-	Category as a comparison			
	7	$\beta_{12}X_{12}(6)$	0.1808	0.1724	0.253	0.275
Month of Incident	1	$\beta_{13}X_{13}(1)$	0.2000	0.1886	0.434	0.459
	2	$\beta_{13}X_{13}(2)$	0.0425	0.0320	0.872	0.903
	3	$\beta_{13}X_{13}(3)$	0.0556	0.0470	0.830	0.856
	4	$\beta_{13}X_{13}(4)$	0.1585	0.1523	0.529	0.544
	5	$\beta_{13}X_{13}(5)$	0.2342	0.2347	0.371	0.368
	6	$\beta_{13}X_{13}(6)$	-0.2679	-0.2981	0.312	0.261
	7	$\beta_{13}X_{13}(7)$	0.1793	0.1710	0.478	0.497
	8	$\beta_{13}X_{13}(8)$	-0.2950	-0.3101	0.250	0.226
	9	$\beta_{13}X_{13}(9)$	-0.0429	-0.0509	0.862	0.836
	10	$\beta_{13}X_{13}(10)$	-0.2875	-0.3098	0.267	0.231
	11	$\beta_{13}X_{13}(11)$	-0.1392	-0.1590	0.604	0.553
	12	-	Category as a comparison			

In the SGD (Stochastic Gradient Descent) method, the variable that significantly affects the dependent variable is the regional level with the province as its category, represented by $X_{10}(1)$, with a significance value of $0{,}072 < 0,05$. Thus, the null hypothesis H_0 is rejected and the alternative hypothesis H_1 is accepted, indicating that the regional level significantly influences traffic accidents. Based on these results, a binary logistic regression model can be constructed as follows:

$$\ln\left(\frac{P(y=1)}{1-P(y=1)}\right)_{SGD} = \beta_0 + \beta_{10}X_{10}(1) = -1,9991 + 0,2646X_{10}.$$

Description:

- β_0: intercept,
- $\beta_{10}X_{10}(1)$: variable representing the regional level with province as its category.

Parameter estimation in the binary logit model has been conducted using two optimization methods: Newton-Raphson and Stochastic Gradient Descent (SGD). The analysis indicates that the Newton-Raphson method failed to converge, although it converges faster in terms of iteration speed. On the other hand, the SGD method achieved convergence, albeit requiring many more iterations. Based on the AIC value, SGD provides a better model fit than Newton-Raphson in analyzing traffic accident data in Malang Regency, although the difference is not substantial.

Therefore, future studies aiming to develop multiclass classification models are encouraged to use larger and more balanced datasets. Techniques such as oversampling, undersampling, or other resampling methods are recommended to improve the representativeness and accuracy of the results. Additionally, to enhance the efficiency and accuracy of the SGD optimization method, it is recommended to incorporate convergence acceleration techniques such as momentum, Nesterov Accelerated Gradient, or Quasi-Hyperbolic Momentum to speed up the convergence process.

3.6 Interpretation Test of Parameter Coefficients

Based on the estimation results using the Newton-Raphson (NR) method, the occupation variable with the student category has a coefficient of $\beta_4 = -0.5832$, which is negative. This indicates that being a student reduces the likelihood of being involved in a traffic accident compared to the reference category, which is "school student". The odds ratio can be calculated as $e^{\beta_4} = e^{-0.5832} \approx 0.558$, meaning that students have a 55.8% chance of being involved in an accident compared to school students, or a 44.2% lower probability.

Conversely, the estimation result using the Stochastic Gradient Descent (SGD) method shows that the region level variable with the province category has a coefficient of $\beta_{10} = 0.2646$, which is positive. This indicates that the likelihood of accidents is higher in provincial regions compared to the national level as the reference category. The corresponding odds ratio is $e^{0.2646} \approx 1.303$, implying that the accident probability increases by approximately 30.3% in provincial regions compared to national regions, assuming other variables are held constant.

3.7 Akaike Information Criterion (AIC) Tests

The parameter estimation results show that the SGD method yields an AIC value of 3198.60 and a log-likelihood of -1544.30, while the NR method yields an AIC of 3201.93 and a log-likelihood of -1545.97. Since the AIC value from SGD is lower and the log-likelihood is higher than those from NR, it can be concluded that the SGD optimization method provides a better model fit in analyzing traffic accidents in Malang Regency.

4 Conclusion

Parameter estimation in the binary logit model has been conducted using two optimization methods: Newton-Raphson and Stochastic Gradient Descent (SGD). The analysis indicates that the Newton-Raphson method failed to converge, although it converges faster in terms of iteration speed. On the other hand, the SGD method achieved convergence, albeit requiring many more iterations. Based on the AIC value, SGD provides a better model fit than Newton-Raphson in analyzing traffic accident data in Malang Regency, although the difference is not substantial. Therefore, future studies aiming to develop multi-class classification models are encouraged to use larger and more balanced datasets. Techniques such as oversampling, undersampling, or other resampling methods are recommended to improve the representativeness and accuracy of the results. Additionally, to enhance the efficiency and accuracy of the SGD optimization method, it is recommended to incorporate convergence acceleration techniques such as momentum, Nesterov Accelerated Gradient, or Quasi-Hyperbolic Momentum to speed up the convergence process.

CRediT Authorship Contribution Statement

Aldi Rahmad Nur Fauzi: Conceptualization, Methodology, Writing–Original Draft. **Sobri Abusini:** Data Curation, Formal Analysis, Writing–Review & Editing. **Corina Karim:** Software, Validation, Visualization.

Declaration of Generative AI and AI-assisted technologies

No generative AI or AI-assisted technologies were used during the preparation of this manuscript.

Declaration of Competing Interest

The authors declare no competing interests.

Funding and Acknowledgments

TThis research received no external funding.

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request and subject to confidentiality agreements.

References

[1] D. Namiot, O. Pokusaev, and A. Chekmarev, "Transportation data as a source for identifying social clusters in the city," *Transportation Research Procedia*, vol. 664-669, 2022. DOI: 10.1016/j.trpro.2022.06.060.

[2] B. P. S. K. Malang, *Kabupaten Malang dalam Angka Malang Regency in Figure*. Badan Pusat Statistik Kabupaten Malang, 2024. Available online.

[3] J. L. A. Oliveira, A. L. Aquino, R. G. Pinheiro, and B. Nogueira, "Optimizing public transport system using biased random-key genetic algorithm," *Applied Soft Computing*, vol. 158, 2024. DOI: 10.1016/j.asoc.2024.111578.

[4] H. Yu, Z. Li, G. Zhang, P. Liu, and T. Ma, "Fusion convolutional neural network-based interpretation of unobserved heterogeneous factors in driver injury severity outcomes in single-vehicle crashes," *Analytic Methods in Accident Research*, vol. 30, 2021. DOI: 10.1016/j.amar.2021.100157.

[5] P. R. Indonesia, *Peraturan Pemerintah (PP) Nomor 34 Tahun 2006 tentang Jalan*. Badan Pemeriksa Keuangan, 2006. Available online.

[6] M. Deublein, M. Schubert, B. T. Adey, J. Köhler, and M. H. Faber, "Prediction of road accidents: A bayesian hierarchical approach," *Accident Analysis & Prevention*, vol. 51, pp. 274–291, 2013. DOI: 10.1016/j.aap.2012.11.019.

[7] N. A. Stanton and P. M. Salmon, "Keterkaitan jalur transportasi dan interaksi ekonomi kabupaten konawe utara dengan kabupaten/kota sekitarnya," *Safety Science*, vol. 47, pp. 227–237, 2009. DOI: doi.org/10.1016/j.ssci.2008.03.006.

[8] H. Ding, Y. Lu, N. Sze, C. Antoniou, and Y. Guo, "A crash feature-based allocation method for boundary crash problem in spatial analysis of bicycle crashes," *Analytic Methods in Accident Research*, vol. 37, 2023. DOI: 10.1016/j.amar.2022.100251.

[9] J. Nugraha, *Metode Maksimum Likelihood dalam Model Pemilihan Diskrit*. Universitas Islam Indonesia, 2017. Available online.

[10] Z. Li, Y. Ci, C. Chen, *et al.*, "Investigation of driver injury severities in rural single-vehicle crashes under rain conditions using mixed logit and latent class models," *Accident Analysis & Prevention*, vol. 124, pp. 219–229, 2019. DOI: 10.1016/j.aap.2018.12.020.

[11] A. Alogaili and F. Mannerin, "Differences between day and night pedestrian-injury severities: Accounting for temporal and unobserved effects in prediction," *Analytic Methods in Accident Research*, vol. 33, 2022. DOI: 10.1016/j.amar.2021.100201.

[12] W. Kriswardhana, S. Sulistyono, I. E. Muhammadiyah, *et al.*, "Modeling the probability of speeding behaviour and accident involvement using binary logistic regression in east java province," *Journal of Indonesia Road Safety*, vol. 2, pp. 149–158, 2019. DOI: 10.19184/korlantas-jirs.v2i3.15048.

[13] L. O. M. Saris, H. Pramoedyo, and A. A. R. Fernandes, "Modelling geographically weighted truncated spline regression using maximum likelihood estimation for human development disparities," *CAUCHY: Jurnal Matematika Murni dan Aplikasi*, vol. 10, pp. 290–300, 2025. DOI: 10.18860/ca.v10i1.31381.

[14] N. Fitriyati and M. Y. Wijaya, "A monte carlo simulation study to assess estimation methods in cfa on ordinal data," *CAUCHY: Jurnal Matematika Murni dan Aplikasi*, vol. 7, pp. 332–344, 2022. DOI: 10.18860/ca.v7i3.14434.

[15] A. Agresti, *Categorical Data Analysis*. John Wiley and Sons, 2013. Available online.

[16] M. A. A. Fawazdhia1 and Z. A. Rafsanjani, "Long short term memory using stochastic gradient descent and adam for stock prediction," *CAUCHY: Jurnal Matematika Murni dan Aplikasi*, vol. 8, pp. 16–29, 2023. DOI: 10.18860/ca.v8i2.17789.

[17] M. Ali and S. AbouRizk, "Updating simulation model parameters using stochastic gradient descent," *Automation in Construction*, vol. 166, 2024. DOI: 10.1016/j.autcon.2024.105676.

[18] M. F. Qudratullah, "Misklasifikasi mahasiswa baru fsaintek uin sunan kalijaga jalur tes dengan analisis regresi logistik," *CAUCHY: Jurnal Matematika Murni dan Aplikasi*, vol. 1, pp. 175–181, 2011. DOI: 10.18860/ca.v1i4.1799.

[19] I. Ghozali, *Aplikasi Analisis Multivariate dengan Program SPSS*. Universitas Diponegoro, 2016.

[20] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*. John Wiley & Sons, 2000. DOI: 10.1002/0471722146.

[21] S. Konishi and G. Kitagawa, *Information criteria and statistical modeling*. Springer series in statistics, 2007. DOI: 10.1007/978-0-387-71887-3.

[22] E. I. Harlyan, E. S. Yulianto, Y. Fitriani, and Sunardi, "Akaike information criterion (aic) in measurement of technical efficiency of purse seine fishery in tuban, east java," *Marine Fisheries: Journal of Marine Fisheries Technology and Managemen*, vol. 11, pp. 181–188, 2021. DOI: 10.29244/jmf.v11i2.38550.