



## Predicting the Segment-Based Effects of Heterogeneous Traffic and Road Geometric Features on Fatal Accidents

Martha Leni Siregar<sup>1</sup>, Tri Tjahjono<sup>1\*</sup>, Nahry Yusuf<sup>1</sup>

<sup>1</sup>*Department of Civil and Environmental Engineering, Faculty of Engineering, Universitas Indonesia, Kampus UI Depok, Depok 16424, Indonesia*

**Abstract.** Inter-urban roads in Indonesia are characterized mainly by distinct road geometry and heterogeneous traffic features. The accident database from the Republic of Indonesia National Traffic Police recorded a substantial number of fatal accidents and fatalities along inter-urban roads. This study aimed to analyze the effects of traffic heterogeneity and road geometry features on fatal accidents along inter-urban roads in South Sulawesi, Indonesia. Segment-based accident analysis was adopted to minimize bias due to the large standard deviations of road lengths. Vehicle-specific speeds, speed standard deviations, and volumes of six vehicle categories, road surface condition, and road geometry were the classified predicting factors. A machine learning technique was adopted to produce predictions of the classification problem. A total of 1,068 road segment observations from 2013–2016 were used to build and validate the model. Model generalization was carried out using the out-of-sample 2019 data. With 26 potential predictors, three machine learning techniques based on the ensembles of regression trees were used to avoid removing potential predictors altogether. The results indicate that road-related features show the greatest importance in predicting the number of fatal accidents. Among the speed features, the average speed of angkots and speed standard deviation of motorcycles showed the greatest importance. The average daily traffic (ADT) of pickups had the greatest importance among other vehicle-specific ADTs.

**Keywords:** Fatal accidents; Heterogeneous traffic; Machine learning; Segment-based effects; Speed standard deviation

### 1. Introduction

Traffic accidents on inter-urban roads in Indonesia are still considered a serious problem, with high rates of fatalities. As an illustration, the fatality rates on Bantaeng–Bulukumba, Jeneponto–Bantaeng, and Bulukumba–Tondong in South Sulawesi were 23.6, 6.5, and 5.1 deaths per 100 million vehicle-km in 2015 ([Australia–Indonesia Partnership, 2017](#)) and were 7.03, 3.35, and 8.19 deaths per 100 million vehicle-km in 2019, based on the present study.

Various factors related to traffic safety have been studied using different approaches. Studies on speed variations have revealed that both the standard deviation (SD) of speed and the coefficient of speed variation (CSV) were significantly related to traffic collisions ([Choudhary et al., 2018](#); [Wang et al., 2018](#)). The effects of road geometry on traffic safety were studied by [Chen et al. \(2019\)](#) and [Papadimitriou et al. \(2019\)](#), and the effects on

---

\*Corresponding author's email: [tjahjono@eng.ui.ac.id](mailto:tjahjono@eng.ui.ac.id), Tel.: +62-81311467022  
doi: [10.14716/ijtech.v13i1.4450](https://doi.org/10.14716/ijtech.v13i1.4450)

pedestrian safety were studied by Siregar et al. (2019). In these studies, the traffic factors were considered homogeneous, and vehicle-specific traffic speeds and volumes were not considered. The heterogeneity of traffic in speeds and volumes was adopted in the study on fatality rates and accident rates (Siregar et al., 2020) on inter-urban roads. Similarly, various approaches have been applied to analyze traffic safety indices. Some of the most widely used regression analyses include the Poisson regression (Mitra et al., 2017) and negative binomial regression (Poisson-gamma; Tjahjono, 2010; Gitelman et al., 2017). An artificial neural network was also used in a study by Alkheder et al. (2017).

Because the distribution of accidents is substantially zero inflated and most road segments under study (86% in 2013–2016 data) did not have fatal accidents, the distribution is instrumental in predicting fatal accidents on a segment base that can be applied to fatal accident preventive measures. Therefore, the present study aimed to analyze the segment-based effects of traffic and road factors on predicting the number of fatal accidents on inter-urban roads by considering the heterogeneity of traffic. These factors, referred to as “features” in machine learning, include the average daily traffic, average speeds, and speed standard deviations of the different categories of vehicles, road geometry, road surface condition, and road section length. Machine learning was adopted for the analysis, given that the technique can overcome the nonlinear relationship problem between fatal accidents and predictive features. Combinations of the different features that are hypothesized to determine the number of accidents were explored. The findings can be expected to contribute to the development of traffic safety improvement programs.

## 2. Methods

The present study adopted machine learning to predict the dependent variables, as shown in Figure 1. As found by Lee et al. (2020), machine learning has become a preferred method for analyzing traffic accidents. A comparison between statistical techniques and machine learning that was conducted by Zhang et al. (2018) found that, despite the overfitting issues, machine learning outperformed the classical statistical techniques and showed better prediction accuracy. Machine learning algorithms are preferred when the goal is to predict the nonlinear relationship of a number of predictor variables with the predicted value (target variable). Machine learning also overcomes the complex interactions of variables. The number of potential predictors in this study was rather large at 26. Because it was necessary to preserve as much information contained in the predictors as possible and avoid removing any predictors completely, three machine learning techniques based on ensembles of regression trees were used: random forest (Krishnaveni and Hemalatha, 2011), gradient boosting machines (Pradhan and Sameen, 2020), and bagging regression trees (Sutton, 2005). A number of combinations of predictors were created to see how the various combinations of predictors performed.

The steps of the methodology were as follows:

1. The three machine learning models were trained on 2013–2015 data (786 observations), and their predictive validity was compared using the data from 2016 (262 observations). Mean absolute error (MAE) was used as the criterion for the comparison.

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (1)$$

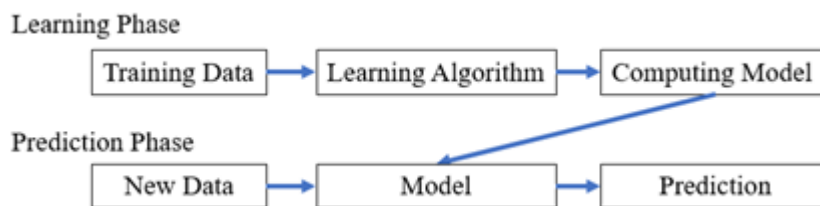
where  $y_i$  is the measured value,  $\hat{y}_i$  is the predicted value, and  $N$  is the size of the training/test set.

2. The best model in terms of the MAE was rebuilt on the entire 2013–2016 dataset.

3. Permutation-based feature importance measures (accumulated local effects) plots were used to assess the marginal effects of each variable on accident counts.
4. The generalizability of the best model was assessed for identifying risky road segments (in terms of the probability of fatal accidents). The sorting quality was assessed using the area under the receiver operating characteristics (ROC) curve measure using out-of-sample data from 2019.

Being ensembles of trees, the three models repeatedly try various combinations of predictors when building each tree, making it possible to look at how well various combinations of predictors perform and thus allowing each predictor to enter the model in at least some of the individual trees. The process allows for the use of all information without any losses of variance and interpretability associated with traditional approaches to dimensionality reduction, such as principal component analysis. The number chosen for trees built was 5,000. The three techniques can be explained as follows:

- 1) The random forest technique applies the bootstrap aggregation (bagging) approach to overcome the bias–variance trade-off. Bias and variance are the factors that constitute learning errors. When the bias is high, the forecasted results are not accurate compared to the actual results, and when the variance is high, the forecast fits very well for specific data sets but not for others, which indicates low levels of forecasting stability.
- 2) Gradient boosting is a machine learning technique for regression and classification problems that produces a prediction model in an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion, as other boosting methods do, and it generalizes them by allowing for the optimization of an arbitrary differentiable loss function.
- 3) Bagging maintains low levels of bias and reduces high levels of variance by randomly sampling observations and features in the training set. The random forest algorithm is the most popular one using the bagging approach; its accuracy (low bias) remains the average value of a decision tree, and the variance is reduced by the use of the central limit theorem.



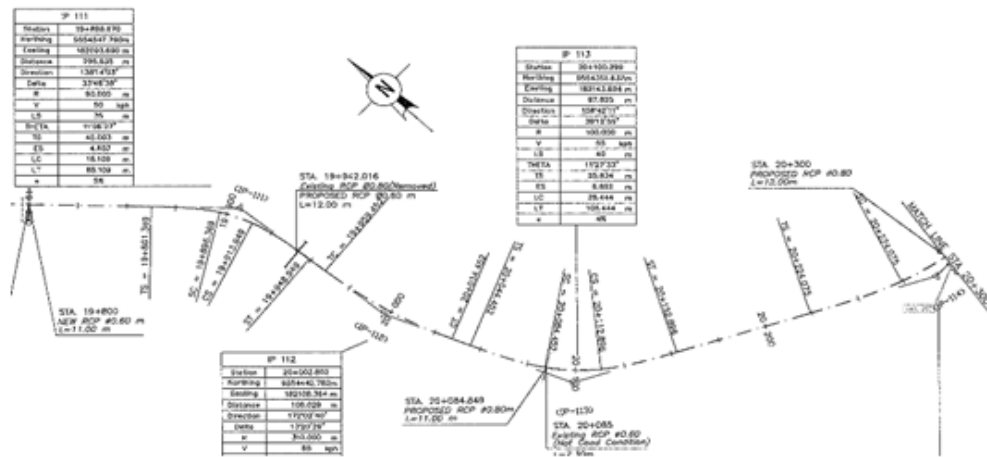
**Figure 1** Machine learning phases

### 2.1. Data Set

The data set used in the analysis comprises South Sulawesi provincial inter-urban roads obtained from the EINRIP Fifth Monitoring Survey ([Australia–Indonesia Partnership, 2017](#)). The accident data were accessed from the national traffic police accident record Integrated Road Safety Management System (IRSMS) for 2013–2016 and 2019. The study was conducted on the following South Sulawesi roads: Jeneponto–Bantaeng, Bantaeng–Bulukumba, Bulukumba–Tondong, and Tondong–Sinjai with the 2013–2016 historical data. The 2019 data were used to check the generalizability of the model.

The traffic characteristics were represented by the average daily traffic (adt), the average speeds (as), and the speed standard deviations (sd) of six vehicle categories that represented the traffic: passenger cars, angkots, pickups, buses, trucks, and motorcycles.

Angkots were classified as a separate group of vehicles because these para-transits often operate without fixed routes and schedules that would normally result in specific speed behaviors. The road features included segment length, horizontal curve, road gradient, and international roughness index (iri). In this study, non-motorized vehicles were not considered. Road segments were defined based on changes in the road geometry and classified into curved and straight and negative or positive gradients as obtained from the detailed engineering design (DED) of roads (Australia–Indonesia Partnership, 2007). A typical DED of an alignment layout is shown in Figure 2. The road surface was represented by the international roughness index (IRI) values. There are 85 road segments along the Jenepono–Bantaeng road, 105 road segments along the Bantaeng–Bulukumba road, 53 road segments along the Bulukumba–Tondong road, and 19 road segments along the Tondong–Sinjai road, which resulted in a total of 1,068 observations.



Source: Australia–Indonesia Partnership (2007)

Figure 2 Alignment layout of a section of Sengkang-Impa-Impa-Tarumpake Road

2.2. Data Statistics

The combined total data trained resulted in 1,048 observations and were used as input in the analysis. The statistics are presented in Table 1. The number of accidents was highly skewed: zero in most of the road segment/year combinations, with 1–5 accidents for most other cases and only a few cases when the number of accidents was higher (Figure 3). The number of accidents for each road segment was a count variable that can be modeled with a particular class of regression models for count data.

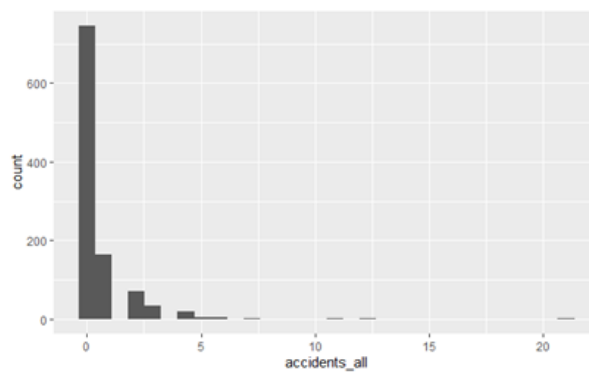


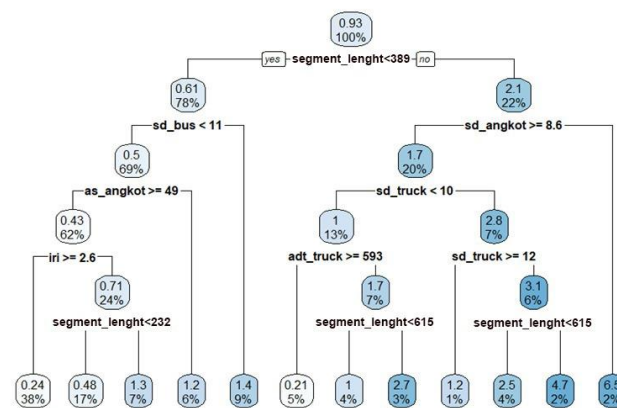
Figure 3 Histogram of the distribution of the annual total number of accidents across road segments

**Table 1** Summary of statistics

Statistic	N	Mean	St. dev.	Min	25%ile	75%ile	Max
<i>segment_length</i> (m)	1,048	298.43	220.57	60.28	160.13	369.14	1,402.24
<i>geometry_curve</i> ( <i>curve</i> = 0, <i>straight</i> = 1)	1,048	0.41	0.50	0	0	1	1
<i>gradient_positive</i> ( <i>positive</i> = 0, <i>negative</i> = 1)	1,048	0.56	0.50	0	0	1	1
<i>IRI</i> (m/km)	1,048	2.79	0.38	2	2.5	3	4
<i>sd_motorcycle</i> (kph)	1,048	10.85	0.77	9.30	10.30	11.20	12.40
<i>sd_passenger_car</i> (kph)	1,048	11.72	0.93	9	11.20	12.30	13
<i>sd_angkot</i> (kph)	1,048	10.46	1.25	8.50	9.90	11.10	12.80
<i>sd_pickup</i> (kph)	1,048	11.934	1.06	10.00	11.00	12.45	14.20
<i>sd_bus</i> (kph)	1,048	5.51	4.74	0	0	9	25
<i>sd_truck</i> (kph)	1,048	10.08	0.87	8.30	9.60	10.50	11.70
<i>as_motorcycle</i> (kph)	1,048	46.13	19.97	9.70	44.03	57.80	63.00
<i>as_passenger_car</i> (kph)	1,048	60.23	3.19	54.20	58.40	61.80	65.70
<i>as_angkot</i> (kph)	1,048	53.57	3.01	48.40	50.90	55.70	60.20
<i>as_pickup</i> (kph)	1,048	55.62	2.73	50	53.3	58	60
<i>as_bus</i> (kph)	1,048	44.51	19.20	0.00	46.30	53.03	62.80
<i>as_truck</i> (kph)	1,048	50.37	2.53	44.30	48.10	52.40	53.80
<i>adt_motorcycle</i> (vehs)	1,048	6,380.32	2,504.21	4,307	4,878	7,649	14,040
<i>adt_passenger_car</i> (vehs)	1,048	2,282.45	631.80	1,485	1,785	2,657	5,309
<i>adt_angkot</i> (vehs)	1,048	1,597.38	822.20	398	1,205.5	1,644	4,799
<i>adt_pickup</i> (vehs)	1,048	998.93	358.94	696	752	942	2,139
<i>adt_bus</i> (vehs)	1,048	80.23	123.32	30	35	75	606
<i>adt_truck</i> (vehs)	1,048	823.30	679.46	465	497	689	2,821
<i>accidents_fatal</i>	1,048	0.16	0.48	0	0	0	4

### 3. Results and Discussion

A classification and regression (CART) tree was created to explain the interaction effects and how the values of an outcome were predicted based on other values. The CART tree was a fully data-driven algorithm. Each node of the tree shows the average number of accidents and the size of the node (% of the total sample size). According to the tree in Figure 4, the average number of accidents across all years, roads, and road segments was 0.93, but there were combinations of factors that were associated with, on average, as few as 0.21 accidents (segment length > 389, sd\_angkot ≥ 8.6, sd\_truck < 10, adt\_truck ≥ 593) or as many as 6.5 accidents (segment length > 389, sd\_angkot < 8.6). Despite offering a number of insights into high- and low-risk combinations of road characteristics that would otherwise be hard to detect in a large volume of data, a single tree is capable of accounting for only some of the significant predictors.



**Figure 4** CART tree of the annual total number of accidents

### 3.1. Correlation Matrix

Pearson correlations indicated the strength of linear correlations of road features with each of the dependent variables based on 1,048 observations from 2013–2016. Coefficients  $|r| > 0.06$  were statistically significantly different from zero at the 5% significance level. The correlations are given in Table 2.

**Table 2** Correlation matrix

Statistic	All accidents	Fatal accidents
segment_length	0.346	0.148
adt_pickup	0.190	0.105
as_motorcycle	0.186	0.079
sd_motorcycle	0.156	0.082
sd_truck	0.108	0.074
adt_motorcycle	0.097	0.045
adt_angkot	0.091	0.054
adt_bus	0.083	0.070
sd_bus	0.048	0.056
adt_passenger_car	0.013	-0.010
sd_passenger_car	0.010	-0.007
gradient_positive	0.009	-0.045
as_passenger_car	0.003	0.011
adt_truck	-0.007	0.004
as_truck	-0.011	-0.022
geometric_curve	-0.029	-0.008
as_pickup	-0.055	-0.035
as_bus	-0.101	-0.016
iri	-0.107	-0.055
sd_pickup	-0.108	-0.071
as_angkot	-0.145	-0.055
sd_angkot	-0.155	-0.095

Unlike the conventional statistical analysis for choosing predictive variables, machine learning utilizes the black-box mechanism, which can be viewed in terms of its inputs and outputs. The effects of correlations of the individual predictor variables and the interaction effects or how features influence the prediction of a machine learning model can be described by partial dependence (PD) or accumulated local effects (ALE) plots. In PD plots, the features have to be uncorrelated, while in ALE plots, the correlations among features can be ignored. ALE plots can correctly interpret model predictions on correlated variables. In this study, we used ALE because it allows for the visualization of the individual predictor variables and the low-order interaction effects.

### 3.2. Train/Test Split Testing of Algorithms

Bagging regression trees outperformed the two other tree-based algorithms and resulted in very low MAEs on the testing sample (2016 data), with all accidents and fatal accidents as dependent variables. On average, the absolute errors were 0.24 for all accidents and 0.14 for fatal accidents, respectively, as shown in Table 3.

**Table 3** Comparison of machine learning methods by MAE on the testing sample (2016)

Method	Dependent variable	
	All accidents	Fatal accidents
Random forest	1.51	0.35
Gradient boosting machines	1.72	0.39
Bagging regression trees	0.24	0.14

### 3.3. Variable Importance Analysis

The most competitive of the three models—bagging regression trees—was estimated

using the complete 2013–2016 data set. The features importance in Table 4 was analyzed for the predictions by shuffling and measuring how much the performance dropped in terms of MAE.

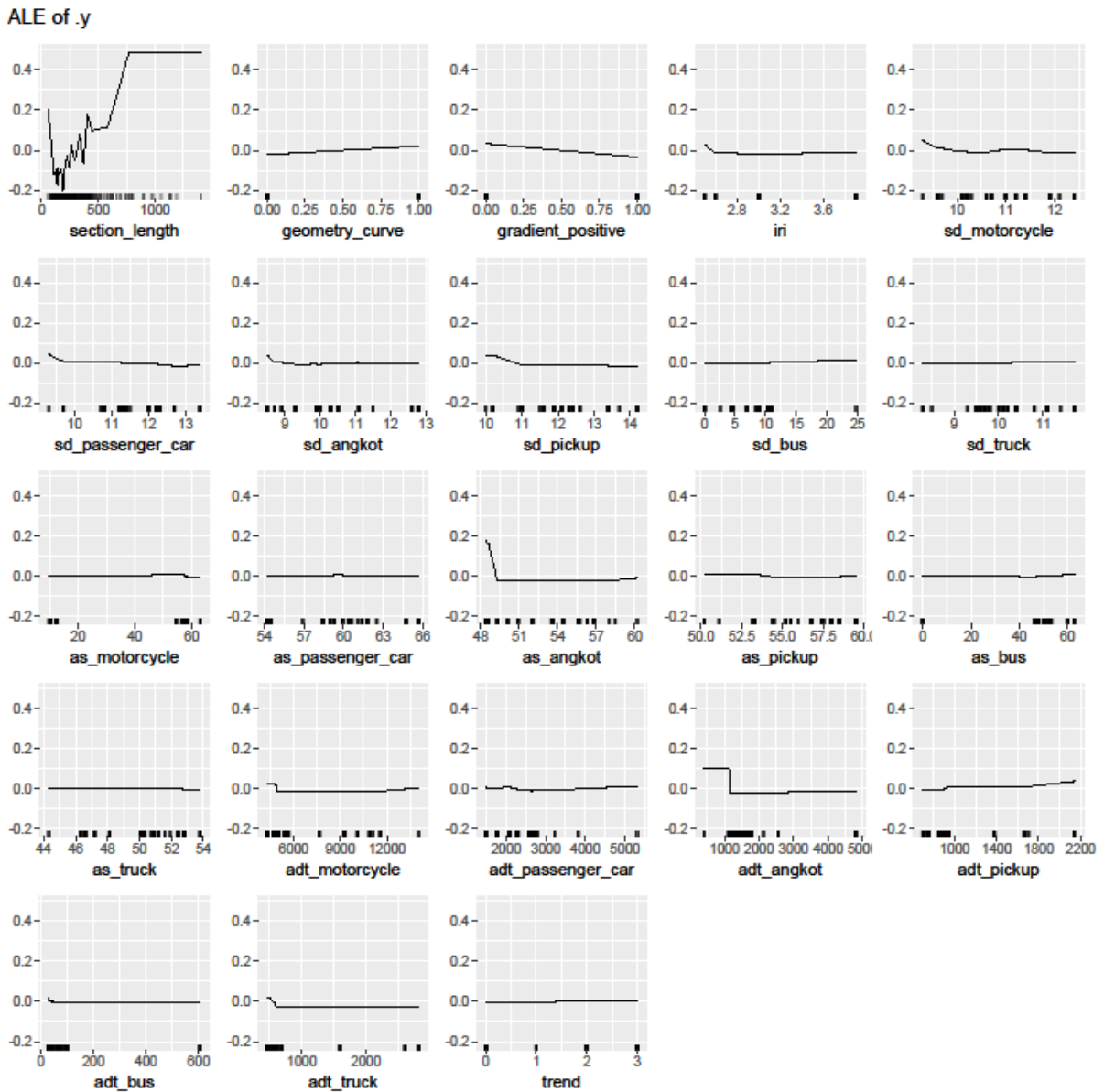
**Table 4** Importance of road characteristics for explaining each of the dependent variables (top 10 features are **bold**)

Variable	All accidents	Fatal accidents
<b>segment_length</b>	<b>3.46</b>	<b>2.69</b>
<b>iri</b>	<b>1.36</b>	<b>1.42</b>
<b>gradient_positive</b>	<b>1.35</b>	<b>1.31</b>
<b>geometry_curve</b>	<b>1.31</b>	<b>1.28</b>
<b>as_angkot</b>	<b>1.30</b>	<b>1.25</b>
<b>adt_pickup</b>	<b>1.30</b>	<b>1.21</b>
<b>adt_truck</b>	<b>1.24</b>	<b>1.16</b>
<b>sd_motorcycle</b>	<b>1.23</b>	<b>1.15</b>
<b>sd_bus</b>	<b>1.21</b>	<b>1.14</b>
<b>sd_angkot</b>	<b>1.18</b>	<b>1.14</b>
as_motorcycle	1.16	1.11
sd_pickup	1.14	1.11
as_bus	1.12	1.10
sd_truck	1.12	1.09
adt_bus	1.10	1.09
sd_passenger_car	1.10	1.06
adt_angkot	1.07	1.05
adt_passenger_car	1.06	1.05
adt_motorcycle	1.05	1.05
as_passenger_car	1.05	1.04
as_pickup	1.03	1.04
as_truck	1.02	1.01

The importance was measured as the factor by which the model's MAE increased when the feature was shuffled. By shuffling the feature values, the association between the outcome and the feature is destroyed. The greater the increase in prediction error, the more important the feature was (Molnar, 2020). Feature importance was remarkably consistent across dependent variables (all accidents and fatal accidents), as shown in Table 4. The road features of segment length, IRI, and gradient and the geometric factors demonstrated the greatest importance in explaining the change in all accidents and fatal accidents. This finding is in line with that of Siregar et al. (2015), which concluded that the road geometry of segments was the underlying cause of accidents involving heavy vehicles.

The relative importance shown in Table 4 describes how features influenced the prediction of a machine learning model on average and how much the features added to the model's predictive power. The ADTs of pickups and trucks were more significant than the vehicles' speeds and SDs. Among the SDs, that of motorcycles showed the greatest importance. Motorcycle contributions to fatal accidents was more significant through the speed SDs than through the average speed. This finding is strongly supported by da Costa et al. (2018), who showed that motorcycle accident risks are related to inappropriate speed choices. The average speed of angkots exhibited the greatest substantial effect among the average speeds. The effect might result from the speed behavior of angkots as a demand–response type of transport that picks up and drops passengers at irregular points, leading to accident occurrences. In line with Zuraida et al. (2017), the importance of speed-related features in fatal accidents may also be associated with drivers' characteristics, leading to varying speed behaviors. Siregar et al. (2020) found that road geometry features showed indirect effects on fatality rates through speed and speed

deviations as the mediating factors. This discrepancy may be due to the accident data that were based not on segments but on road lengths.



**Figure 5** Accumulated local effects of each feature (horizontal axis) on the total number of fatal accidents (vertical axis; 0 corresponds to the mean number of accidents)

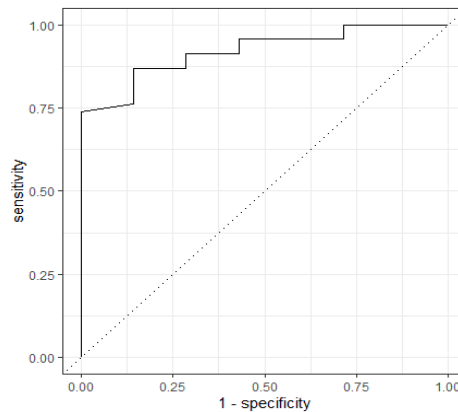
An ALE plot that described the effects, on average, of the feature on the prediction of a machine learning model was created, as shown in Figure 5. Compared to pairwise correlations, these local effects reflected partial associations after adjusting for the rest of the factors. It can be seen that some of the major risk factors where there was substantial variation in the dependent variable as the predictor changed were section length, low average speed, and speed deviation of angkots, along with low-speed deviation of pickups. Straight segments and positive gradients proved to potentially increase the number of fatal accidents more than curves and negative gradients. An increase in IRI lowered the effects only until the IRI reached 2.4. The number of fatal accidents increased (positive local effect) when an angkot's average speed was either too low (< 50 km/h) or too high (> 58 km/h). Angkots' ADT increased the number of fatal accidents until the ADT reached



1,000 vehicles. Additional ADT of passenger cars tended to have no effect until the ADT reached around 3,300 passenger cars per day on the road segment. Then, the effect of every additional car increased as the ADT went up. Motorcycle traffic was a factor that increased the number of fatal accidents up until the motorcycle ADT reached about 5,200 motorcycles per day, and then the effect of each additional motorcycle was close to zero.

### 3.4. Best Model's Generalizability

The bagged regression tree model was used and built on the full data set from 2013–2016 and applied to 2019 data on 53 segments of the Bulukumba–Tondong road. This validation data set was not used in model training and was thus completely independent.



**Figure 6** ROC curve (AUC = 0.92) for the bagged regression tree model applied to 2019 validation data on fatal accidents

The area under the receiver operating characteristic (ROC) curve in Figure 6 turned out to be 0.92, implying that a road segment where fatalities actually occurred was positioned above a random road segment where fatalities did not actually occur, with 92% probability. Such a high area under the curve (AUC) corresponds to the very high diagnostic power of the model.

## 4. Conclusions

The results indicate that there are different factors determining the number of fatal accidents. ADT, speeds, and speed SDs are vehicle-specific variables with different levels of importance. The findings of the present study should also be intuitively interpreted, considering that there might be some underreporting of fatal accidents due to failure to change the status from severe to fatal. Because the high contributions of some vehicle variables in an accident do not necessarily indicate the involvement of the vehicles in the accidents, further analysis is required to give a more integrated and comprehensive view of the relationships between different categories of vehicles and accidents. Having the greatest importance in percentage among the speed features, the average speed of angkots and the ADT of passenger cars and motorcycles necessitate special treatment to reduce the effects on fatal accidents. The results of this study can help traffic safety agencies target specific features for improvement measures.

## Acknowledgements

This research was funded by TADOK Grant Universitas Indonesia 2019 No. NKB-0167/UN2.R3.1/HKP.05.00/2019. The authors would also like to thank the Project

Management Unit of the Australian-funded EINRIP Monitoring & Evaluation Programme, Fifth Monitoring Survey, Final Report 2017 for permission to use the data.

## References

- Alkheder, S., Taamneh, M., Taamneh, S., 2017. Severity Prediction of Traffic Accident using an Artificial Neural Network. *Journal of Forecasting*, Volume 36(1), pp. 100–108
- Australia–Indonesia Partnership, 2007. *EINRIP Alignment Layout*
- Australia–Indonesia Partnership, 2017. *EINRIP Monitoring & Evaluation Programme*, Fifth Monitoring Survey, Final Report
- Chen, S., Saeed, T.U., Alinizzi, M., Lavrenz, S., Labi, S., 2019. Safety Sensitivity to Roadway Characteristics: A Comparison across Highway Classes. *Accident Analysis & Prevention*, Volume 123, pp. 39–50
- Choudhary, P., Imprialou, M., Velaga, N.R., Choudhary, A., 2018. Impacts of Speed Variations on Freeway Crashes by Severity and Vehicle Type. *Accident Analysis & Prevention*, Volume 121, pp. 213–222
- da Costa, D.G.N., Malkhamah, S., Suparma, L.B., 2018. Use of the Safety Factor and Margin of Safety in Motorcyclist Accident Risk Management. *International Journal of Technology*, Volume 9(4), pp. 737–750
- Gitelman, V., Doveh, E., Bekhor, S., 2017. The Relationship between Free-Flow Travel Speeds, Infrastructure Characteristics and Accidents, on Single-Carriageway Road. *Transportation Research Procedia*, Volume 25, pp. 2026–2043
- IRSMS Korlantas POLRI, n.d. Available Online at <https://irsms.korlantas.polri.go.id/login>. Accessed on March 23, 2019
- Krishnaveni, S., Hemalatha, M., 2011. A Perspective Analysis of Traffic Accident Using Data Mining Techniques. *International Journal of Computer Applications*, Volume 23(7), pp. 40–48
- Lee, J., Yoon, T., Kwon, S., Lee, J., 2020. Model Evaluation for Forecasting Traffic Accident Severity in Rainy Seasons using Machine Learning Algorithms: Seoul City Study. *Applied Sciences*, Volume 10(1), pp. 1–15
- Mitra, S., Haque, M., King, M.J., 2017. Effects of Access, Geometric Design, and Heterogeneous Traffic on Safety Performance of Divided Multilane Highways in India. *Journal of Transportation Safety & Security*, Volume 9(Sup1), pp. 216–235
- Molnar, C., 2020, *Interpretable Machine Learning: A Guide for Making Black Box Models Interpretable*. Leanpub
- Papadimitriou, E., Filtness, A., Theofilatos, A., Ziakopoulos, A., Quigley, C., Yannis, G., 2019. Review and Ranking of Crash Risk Factors Related to the Road Infrastructure. *Accident Analysis & Prevention*, Volume 125, pp. 85–97
- Pradhan, B., Sameen, M.I., 2020. Predicting Injury Severity of Road Traffic Accidents using a Hybrid Extreme Gradient Boosting and Deep Neural Network Approach. *In: Laser Scanning Systems in Highway and Safety Assessment*, Springer International Publishing, Cham, Switzerland, pp. 119–127
- Siregar, M.L., Alawiyah, T., Tjahjono, T., 2015. Remedial Safety Treatment of Accident-Prone Locations. *International Journal of Technology*, Volume 6(4), pp. 680–688
- Siregar, M.L., Sumabrata, R.J., Kusuma, A., Samosir, O.B., Rudrokasworo, S.N., 2019. Analyzing Driving Environment Factors in Pedestrian Crashes Injury Levels in Jakarta and the Surrounding Cities. *Journal of Applied Engineering Science*, Volume 17(4), pp. 482–489

- Siregar, M.L., Tjahjono, T., Nahry, 2020. Endogenous Relationship of Accident Occurrence with Speed, Traffic Heterogeneity and Driving Environment on Inter-Urban Roads in Indonesia. *Journal of Applied Engineering Science*, Volume 18 (4), pp. 475–484
- Sutton, C.D., 2005, *Classification and Regression Trees, Bagging, and Boosting*. In: Handbook of Statistics Volume 24, Rao, C.R., Wegman, E.J., Solka, J.L. (eds.), Amsterdam, The Netherlands, Elsevier, pp. 303–329
- Tjahjono, T., 2010. The Effect of Traffic and Road Conditions to the Fatality Rates on Rural Roads in Eastern Indonesia. *Journal of the Eastern Asia Society for Transportation Studies*, Volume 8, pp. 2201–2213
- Wang, X., Zhou, Q., Quddus, M., Fan, T., Fang, S., 2018. Speed, Speed Variation and Crash Relationships for Urban Arterials. *Accident Analysis & Prevention*, Volume 113, pp. 236–243
- Zhang, J., Li, Z., Pu, Z., Xu, C., 2018. Comparing Prediction Performance for Crash Injury Severity Among Various Machine Learning and Statistical Methods. *IEEE Access*, Volume 6, pp. 60079–60087
- Zuraida, R., Iridiastadi, H., Satalaksana, I.Z., 2017, Indonesian Drivers' Characteristics Associated with Road Accidents. *International Journal of Technology*, Volume 8(2), pp. 311–319