Modeling the potential accidents at the guarded railroad crossing line Surabaya–Lamongan

F P Luthfiyani and H Suprayitno*
Civil Engineering Department, Institut Teknologi Sepuluh Nopember Surabaya, 60111, Indonesia

*fitria.luthfiyani@gmail.com

Abstract. Traffic safety at railroad crossings considered poorly understood by the road users. A good safety understanding at railroad crossings is necessary to prevent potential accidents. This condition is possibly achieved if factors causing the accidents are clearly known. This study aims to find out factors with potentially causing accidents in a railroad crossing, especially related to traffic volume, vehicle queue length, past year accidents, and land use around the railroad crossing. Data were collected from the data inventory survey, which was then analyzed using regression analysis. The results found three equations types based on the land use, respectively as: Type 1 (Railroad Crossing-Local Road/Collector) $Y = 0.97 + 2.89x1 + 0.042x2$; Type 2 (Railroad Crossing-Public Facilities) $Y = 0.693 + 4.4x1 + 0.006x2$; Type 3 (Railroad Crossing-Approaching Intersection) $Y = -0.209 + 3.140x1 + 0.018x2$. This study was benefited to design the policy as well as the layout of railroad crossings in order to minimize the potential accidents in the railroad crossing areas.

1. Introduction
High level of traffic accidents becomes one of the main issues faced by all countries in the world [1]. Risk of traffic accidents varies referring to each area. Mostly, those accidents occur in the area with the worst safety to motorcyclists, cyclists, and pedestrians [2]. World Health Organization (WHO) announced that traffic accidents are treated as one of the non-communicable diseases with the highest number of deaths. According to data released by the Indonesian National Transportation Safety Committee (NTSC), from 2010 to 2016 there had been 41 traffic accidents investigated. The highest numbers of accidents were caused by collision (68.29%) with 443 death toll and 791 injured [3]. Accidents caused by collisions do not only occur on highways but also occur at railroad crossings, especially on the level railroad crossings.

There are 1,267 total railway crossings recorded both maintained and unmaintained in East Java Province. 286 crossings are maintained by PT KAI (Indonesian Railroad Company) and 70 crossings are maintained by the Provincial Government. However, the remaining 911 crossings are left unmaintained. Generally, problems arising at the railway crossing in East Java are due to the growth of settlements along the railroad in which increase the number of the crossings. These problems very likely become the causes of railway crossing accidents. Therefore, this article aims to find the potential factors of the environment causing railway crossing accidents. This study presents a model from the first attempt in order to observe the appropriateness of this methodological approach.
2. Method

This study aims to find the connection of between accident and traffic characteristics based on the land use in the railway crossing of Surabaya - Lamongan. Data were taken from traffic record in Surabaya - Lamongan route. The data represented the model of the intended land use. The land use is divided into three types; type 1 is railroad crossing-local road/collector, type 2 is railroad crossing-public facilities, and type 3 is railroad crossing-approaching intersection. The main approach to analyze the accident possibility is by collecting a number of non-accident cases according to assumptions [4].

For years, many researchers have developed GLM (Generalized Linear Models) to analyze the relationship between accident frequency and contributing parameters. The number of accidents has a discrete, random, and not negative effect on the selection of the analysis model. In previous studies, this type of data is suitable to be modeled using the GLM [5,6].

Generalized Linear Model (GLM) can be assumed as [7]:

- There is a response \( y \) observed independently at fixed values of stimulus variables \( x_1, \ldots, x_p \)
- The stimulus variables may only influence the distribution of \( y \) through a single linear function called the linear predictor
  \[
  \eta = \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_p x_p
  \]  
- The distribution of \( y \) has a density of the form
  \[
  f(y; \theta_i; \varphi) = \exp \left[ A_i \{ y_i \theta_i - y(\theta_i) \} / \varphi + \tau(y_i, \varphi / A_i) \right]
  \]

Where \( \varphi \) is a scale parameter (possibly known), \( A_i \) is a known prior weight and parameter \( \theta_i \) depends upon the linear predictor.

2.1. Data source

This study was located in Surabaya - Lamongan route covering 6 segments (Pasarturi, Kandangan, Cermé, Lamongan, Sumlanar, and Babat) with 128 railway crossings. Based on land use, 22 points represent the population. Data details include the number of accidents [8] and data V/C Ratio [9] hereinafter referred to as the Degree of Saturated (DS), Delay in Traffic (DT) [10]. The sampling used in this study is representative because it was directly obtained from the research location.

2.2. Data coding

Average Daily Traffic (ADT) is analyzed based on land use using the Indonesian Road Capacity Manual calculation formula [11] to obtain the value of the Degree of Saturated (DS). From the 22 traffic data at the intersection of the railway line, the average value is obtained according to the land use. The average value of Degree of Saturated (DS) is denoted by \( X_1 \) and the average number of Delay in Traffic (DT) is denoted by \( X_2 \). Land use is divided into type 1 railroad crossing-local road/collector, type 2 railroad crossing-public facilities, and type 3 railroad crossing-approaching intersections.

| Problem                      | Data Input                                        | Analysis                                      | Output                                |
|------------------------------|--------------------------------------------------|-----------------------------------------------|---------------------------------------|
| Degree of Saturated (DS)     | Average Daily Road Traffic                        | Indonesia road capacity manual and            | Value of Degree of Saturated          |
| (DS) calculation             | Geometric Land use                                | Indonesian road capacity software             |                                       |
| Number of accidents          | Data of the latest accidents in railway crossing   | Interview with railway crossing watchman      | Value of accident based on land use    |
| calculation                  | Data of accidents in East Java                    | Descriptive analysis of accidents in East Java|                                       |
|                              | Data accident in operational area of 8 KAI        |                                               |                                       |
|                              | Land use                                         |                                               |                                       |
| Delay in traffic (DT)        | Distance between vehicles                         | Indonesian road capacity manual and           | Value of Delay in Traffic (DT)        |
| calculation                  | Travel time between vehicles                      | equation of Delay Traffic                     |                                       |
|                              | Velocity between vehicles                         |                                               |                                       |
|                              | Land use                                         |                                               |                                       |
3. Result and discussion

3.1. Analysis Degree of Saturated (DS)
Above discussion showed that accidents are not only caused by traffic flow. In this study, traffic flow will be analyzed with other supporting data. This method uses various operational configurations along with various indicators, such as geometric conditions and traffic conditions. The main output of this method is the level of service which is an indication of the Degree of Saturated (DS). The geometric based approach [12] includes road functions, road width, road shoulder, number of lanes and lanes. Approaches to traffic conditions include the type of passing vehicles, the number of passing vehicles, vehicle speed, and the number of pedestrians. The calculation of traffic volume is converted according to the unit type of vehicle. In some cases, calculations are carried out during peak hours. In his research, Mathew (2014) explained that the Degree of Saturated is the number of volumes of vehicles per hour that can be accommodated in one segment assuming that the green phase displayed is 100% (i.e., $g / C = 1.0$).

3.2. Analysis Delay in Traffic (DT)
Traffic delay at work zones is caused by a reduced number of lanes for traffic and lower vehicle speed. When the traffic volume exceeds the work zone capacity, traffic congestion occurs and therefore results in vehicle queues and traffic delays. Assuming a uniform deceleration, the delay for each vehicle before entering a work zone can be calculated using the basic equations of dynamic. Since vehicles decelerate gradually, this assumption is believed to be reasonable and should result in a fairly accurate estimation of the vehicle delay time.

3.3. The latest accident record
The parameter of this study is relevant to the number of accident. The safety which a railway system provides is evaluated by the accidents that occurred during a specific time period, for example, one year [13]. This study used the number of accident in 2017.

3.4. Model evaluation
Generalized Linear Model (GLM) was used to estimate the model coefficients using the SPSS software package and all consistent with earlier research. Model parameter (coefficients) was estimated using the maximum likelihood approach. The procedure which was adopted in the model development was the forward procedure in which the variables were added to the model one by one [10]. The traditional Generalized Linear Models (GLM) (e.g., Poisson, Negative Binomial, and Zero-Inflated Poisson models) are typically used to estimate the relationship between the number of crashes and covariates because of the stochastic nature of crash occurrence [14].

Proses of data analysis used Indonesian road capacity software. From the analysis, the highest value of Degree of Saturated (DS) is on the land use Type 1 on 213 the railway crossing points, Type 2 on 253 the railway crossing points, and Type 3 on 415 the railway crossing points. From the analysis, the highest value of Delay in Traffic (DT) is on the land use Type 1 and located in 213 railway crossing point, Type 2 is on the 344 railway crossing point, and Type 3 is on 406 railway crossing point. The highest accident in land use Type 1 at the point 213, Type 2 at the point 344, and Type 3 at the point 406.

Based on the value of $R^2$ square on Type 1 Land use, 33.9% of independent variables affect the dependent variable. The magnitude of the variation of the independent variable that affects the dependent variable produces a calculated F value of 1.539 with a significance of 0.289 greater than 0.05. Thus simultaneously the independent variable cannot predict the number of accidents. The value of Degree of Saturated (DS) and Delay in Traffic (DT) do not have a statistically significant value, 0.05 with a 95% truth level [15]. Thus, the regression equation that connects the logarithms of accidents in the form of accident likelihood with Degree of Saturated (DS) and Delay in Traffic (DT) is expressed in the following equation:

$$Y = 0.97 + 2.819x_1 + 0.042x_2$$ (3)
Table 2. Related variables according at land use type 1.

| Variable | Regression Coefficient | t_count | Sig. |
|----------|------------------------|---------|------|
| Constanta | 0.97 | | |
| DS (X1) | 2.819 | 0.528 | 0.616 |
| DT (X2) | 0.042 | 1.663 | 0.147 |
| F_count | 1.539 | | 0.289 |
| R square | 0.339 | | |

Based on the value of R square on Type 2 Land use is 86.7% the independent variable affects the dependent variable. The magnitude of the variation of the independent variable that affects the dependent variable produces a calculated F value of 6.6492 with a significance of 0.133 greater than 0.05. Thus simultaneously the independent variable cannot predict the number of accidents. The value of Degree of Saturated (DS) and Delay in Traffic (DT) do not have a statistically significant value, each of which is 0.271> 0.05 and 0.274> 0.05. Thus, the regression equation that connects the logarithm of the accident in the form of accident likelihood with Degree of Saturated (DS) and Delay in Traffic (DT) is stated in the following equation:

\[ Y = 0.693 + 4.4x_1 + 0.006x_2 \]  \hspace{1cm} (4)

Table 3. Related variables according at land use type 2.

| Variable | Regression Coefficient | t_count | Sig. |
|----------|------------------------|---------|------|
| Constanta | 0.693 | | |
| DS (X1) | 4.4 | 1.504 | 0.271 |
| DT (X2) | 0.006 | 1.491 | 0.274 |
| F_count | 6.66492 | | 0.133 |
| R square | 0.867 | | |

Based on the value of R square on Type 3 Land use, 79.6% of the independent variables affect the dependent variable. The magnitude of the variation of the independent variable that affects the dependent variable produces a calculated F value of 9.737 with a significance of 0.019 smaller than 0.05. Thus simultaneously the independent variable can predict the number of accidents. The value of Degree of Saturated (DS) does not have statistically significant value (0.371> 0.05), however, the value of Delay in Traffic (DT) has a significant relationship of 0.027 <0.05. Thus, the regression equation that connects the logarithm of the accident in the form of accident likelihood with Degree of Saturated (DS) and Delay in Traffic (DT) is stated in the following equation:

\[ Y = -0.209 + 3.14x_1 + 0.018x_2 \]  \hspace{1cm} (5)

Table 4. Related variables according at land use type 3.

| Variable | Regression Coefficient | t_count | Sig. |
|----------|------------------------|---------|------|
| Constanta | -0.209 | | |
| DS (X1) | 3.14 | 0.982 | 0.371 |
| DT (X2) | 0.018 | 3.098 | 0.027 |
| F_count | 9.737 | | 0.019 |
| R square | 0.796 | | |
4. Conclusion
Not all equations based on land use have significant value. There are two equations that are not significant as a whole, namely the Type 1 and Type 2 equations. The value of Degree of Saturated (DS) and the Delay in Traffic (DT) cannot simultaneously predict the number of accidents. However, in the Type 3 equation, there is one variable that has a significant value to predict the number of accidents. The only statistically significant explanatory variable was found to be DT ($X_2$), through its logarithmic transformation. Delay in Traffic (DT) based on speed, The consistent negative sign of the beta coefficient of logistic average speed in all models may seem counter-intuitive but is consistent with similar past studies [4]. The other study reveals the significant explanatory variables which are based on accidents based on geometry were: traffic, lane balance, and the traffic control devices [16]. Another significant variable explanatory other than geometric are needed to obtain an accident model in a railroad crossing. Three models are described: identifying the accident-prone location, the correlation between the explanatory factor accident (DS and DT).

Acknowledgment
This study was supported by previous research and Civil Engineering of Institut Teknologi Sepuluh November Surabaya. Best regard is delivered to all parties who motivated and support until this study concluded.

References
[1] Ihueze C C and Onwurah U O 2018 Road traffic accidents prediction modelling: An analysis of Anambra State, Nigeria 112 21–9
[2] Organization W H 2015 Global status report on road safety 2015 (Italy: World Health Organization)
[3] KNKT D 2016 Data Investigasi Kecelakaan Llaj Tahun 2010 – 2016 1–20
[4] Theofilatos A, Yannis G, Kopelias P and Papadimitriou F 2016 Predicting Road Accidents: A Rare-events Modeling Approach Transportation Research Procedia vol 14 (Elsevier B.V.) pp 3399–405
[5] Zhang Y, Xie Y and Li L 2012 Crash frequency analysis of different types of urban roadway segments using generalized additive model J. Safety Res. 43 107–14
[6] Lu P and Tolliver D 2016 Accident prediction model for public highway-rail grade crossings Accid. Anal. Prev. 90 73–81
[7] Kafadar K, Koehler J R, Venables W N and Ripley B D 1999 Modern Applied Statistics with S-Plus Am. Stat. 53 86
[8] Agbelie B R D K 2016 Random-parameters analysis of highway characteristics on crash frequency and injury severity J. Traffic Transp. Eng. (English Ed. 3 236–42
[9] Zhou M and Sisiopiku V P 1997 Relationship Between Volume-to-Capacity Ratios and Accident Rates Transp. Res. Rec. J. Transp. Res. Board 1581 47–52
[10] Ackaah W and Salifu M 2011 Crash prediction model for two-lane rural highways in the Ashanti region of Ghana IATSS Res. 35 34–40
[11] Anonim 1997 Highway Capacity Manual Project (HCM) (Direktorat Jenderal Bina Marga)
[12] Oh J, Washington S P and Nam D 2006 Accident prediction model for railway-highway interfaces Accid. Anal. Prev. 38 346–56
[13] Pyrgidis C, Papacharitou E and Eleftheriadis A 2016 Risk Management at Railroad Grade Crossings: Proposal for a Decision Support System Transp. Res. Procedia 14 1394–402
[14] Wang X, Wu X, Abdel-Aty M and Tremont P J 2013 Investigation of road network features and safety performance Accid. Anal. Prev. 56 22–31
[15] Hartono 2008 SPSS 16.0 Analisis Data Statistika dan Penelitian (Pekanbaru)
[16] Vieira Gomes S 2013 The influence of the infrastructure characteristics in urban road accidents occurrence Accid. Anal. Prev. 60 289–97