Data mining applied for accident prediction model in Indonesia toll road

Andri Irfan, Ronal Al Rasyid, and Susanty Handayani

Citation: AIP Conference Proceedings **1977**, 060001 (2018); doi: 10.1063/1.5043013 View online: https://doi.org/10.1063/1.5043013 View Table of Contents: http://aip.scitation.org/toc/apc/1977/1 Published by the American Institute of Physics

Data Mining Applied for Accident Prediction Model in Indonesia Toll Road

Andri Irfan^{1, a)} Ronal Al Rasyid^{2, b)} and Susanty Handayani^{3, c)}

¹Civil Engineering, Universitas Internasional Batam, Indonesia ²Jasa Marga Jalanlayang Cikampek Toll Road Company, Indonesia ³Greater Jakarta Transportation Authority, Ministry of Transportation, Indonesia

> Corresponding authors: ^{a)} andrirfan@yahoo.com ^{b)} ron.alraz@gmail.com ^{c)} susanhandayani707@gmail.com

Abstract. The current research on toll road accident (TRA) is mainly conducted using conventional descriptive statistics, which, however, fail to properly identify cause-effect relationships and are unable to construct models that could predict accidents. Alternative to decrease traffic accident is by developing accident prediction model. The model relates accident frequencies with traffic flow and various roadway environment characteristics contributing to accident occurrences. This paper presents the TRA prediction model for Jakarta Outer Ring Road Toll Road (JORR), to identify the most important causes of accidents and to develop predictive models. Data mining (DM) techniques (artificial neural networks (ANNs) and support vector machines (SVM)) were used to model accident and incident data compiled from the historical data. Based on the R-Tools, results were compared with those from some classical statistical techniques (logistic regression (LR), revealing the superiority of ANNs and SVM in predicting and identifying the factors underlying accidents in toll road.

INTRODUCTION

Traffic accidents have been one of the most causes of death and injury in Indonesia. Last year, about 25,859 were killed and hundred thousand people were injured in traffic accidents in Indonesia. In addition, traffic accidents often result in enormous costs to society, including an excessive delay for roadway users and public property damage. In fact, it happens globally in which more than 1.17 million people die in road crashes around the world [1], and over 10 million are crippled or injured each year. Based on the data from the Traffic Corp of Indonesian National Police, the national trend of traffic accident has increased from year to year. The number has increased since 2014. Throughout 2014, there were 95,906 cases, which increased into 98,970 cases in 2015 and into 105,374 cases in 2016. Such increases are contrast with the number of cases in 2012, which was 117,949 cases and decreased into 100,106 cases in 2013 [2]. There are several factors allegedly related with the increasingly number of accidents from year to year, including the human factor, vehicle factor and environments factor. From these three factors, human factor is estimated to be the highest contributor for road accidents. However, the annual number of traffic accidents remains nearly the same every year. Clearly, there is an increasing need for efficient methodologies for identifying the risk factors for accidents.

Toll roads have a relatively high risk of accidents in compared to other road types. In relation with the factors as mentioned previously, approximately 75% of accidents on toll roads were caused by human factors [2]. On toll roads, individual vehicle operators have a wide range of physical and mental abilities, different perceptions of risk, different reactions to external stimuli, and their operating abilities may be further complicated by varying degrees of self-inflicted impaired driving. Very little is known about the primary contributing factors of highway traffic accidents in developing countries, simply because most of them either do not keep any good records or have no records at all [3].

Human-Dedicated Sustainable Product and Process Design: Materials, Resources, and Energy AIP Conf. Proc. 1977, 060001-1–060001-9; https://doi.org/10.1063/1.5043013 Published by AIP Publishing, 978-0-7354-1687-1/\$30.00 The fundamental effort required to improve the toll road safety level in Indonesia is by developing TRA prediction model. The function of this model is to provide estimation or prediction of traffic accident numbers which are free of impact bias from the regression-to-the-mean phenomenon. At the toll operator company level, understanding the relationship between the standards and the real accident number can be important to determine the effect of regulating one standard to the safety level of the road. Furthermore, it can serve as a guidance to a toll road improvement program, especially for developing countries, like Indonesia that are struggling to build roads in compliance with their toll road standard [4]. By identifying which elements of toll road standards contribute to the real number of accidents, the authority can effectively target highway expenditure to the improvement of deficient roads. At the same time, while the accident database system is being improved in operator level, the development of TRA prediction model can help the other operator with weak accident database. Thus, the final goal of the development of TRA prediction model is to improve the level of toll road traffic safety.

This paper presents TRA prediction model developed for JORR. The application of the developed model is expected to identify and rank the causes of the accident. Traditional statistical models used accident record to predict current road safety conditions, and their disadvantages highlight that the complete situation of the accident cannot be completely known [5]. Based on condition the accident rate is discrete and cannot be negative and unique, the TRA prediction model has been built by using a non-linear regression model with ANNs and SVM approach. The developed model utilizes large data owned by toll road operators through DM processing with r-miner. Due to the record of large amounts of data, it is necessary to produce a good and continuous pattern [6]. Approach to new techniques and the use of the latest technology is necessary, therefore the set of data that has been collected through the measurement of accidents can be utilized in a structured and scalable scheme to support the interpretation and prediction of accurate data [7]. The best model performance of DM results is used to develop TRA prediction model on toll road in Indonesia.

TOLL ROAD ACCIDENT PREDICTION

The occurrence of TRA can be caused by several causes, i.e. the user, the vehicle and the road geometry. Environmental and weather conditions also become the causes of accidents. Of the factors causing traffic accidents, infrastructure factor is the factor that can be directly engineered by toll road service providers in an effort to improve road safety through a quality improvement. This is different from other efforts related to human factors, where the role of the regulator is more directed to the rules, policy, and education, which will eventually return to the character of the perpetrator. Similarly, on the efforts to improve safety related to facilities, the regulator cannot provide direct action on the process of preparing the vehicle to be used. It is linked to the authority of the company as the producer of vehicles and the owners of vehicles as prospective users.

The TRA has many impacts. In addition to causing physical harm to the victims of accidents, it will also cause derivative effects, such as economic losses, social, and environmental damage. The consequences of an accident can be calculated based on the potential impacts, such as the estimation of medical costs, production losses, human losses, property damage, settlement costs and accident-induced congestion costs. Regulator has several possibilities to regulate the accident externalities, including efficient control of the insurance markets, upgrade the road network, and control of road users' attitudes and behavior [8]. In order to reduce the accident rate, it is essential to know the possibility of an accident in the future. Some traffic accident prediction techniques have been continually developed. One of the most widely used methods is multivariate modeling. This model is the most common technique used in predicting building accidents. Accident prediction model currently constitutes the main tool for estimating the safety performance of road locations. Traditionally, the models are developed using the poisons-gamma hierarchy, which leads to the negative binomial regression model. The used regression model can be a linear regression model or non-linear regression model. The TRA is a chain of events which results in one or more road users causing damage or injury to oneself or others. While some recent studies present that the number of accidents on a given highway section during a certain period of time is probabilistic in nature and is a non-negative integer [9].

Statistical modeling is used to develop TRA prediction models in relation with the accident occurrence on road facilities to their various traffic and geometric characteristics. The model can be used to predict the accident potential of similar road facilities and also to conduct several safety studies, such as before-and- after studies to evaluate the effectiveness of safety improvement measures and those to identify and rank hazardous or accident-prone locations. Statistical models, moreover, are able to empirically link the accident situation and a mix of risk factors to explain these accidents, for example, road traffic and vehicle characteristics [10]. Furthermore, these statistical models are capable of incorporating a large number of variables and their interactions to explain changes in road accident involvement and severity. The development of statistical science encourages the development of other scientific

prediction methods. Statistical methods were commonly utilized in most previous studies, such as linear regression model, logistic regression model, poisons model, negative binomial model, zero-inflated negative binomial model and generalized linear regression model methods. These methods are subject to strong assumptions and limitations in application. In contrast, ANN has been proven efficient and effective in many fields [9].

The various scientific approaches to learn about TRA have been scientifically proven to provide suggestions for the improvement of safety values and the resolution of other safety issues. There are two main engineering approaches for dealing with traffic safety problems, namely the reactive approach and the proactive approach. The reactive approach or retrofit approach entails the necessary improvements in the existing hazardous sites to reduce accident frequency and severity at these sites. Meanwhile, the proactive approach is an accident prevention approach that tries to prevent unsafe road conditions from occurring in the first place [11]. These various engineering approaches are expected to improve the quality of human life.

Artificial Intelligence

Interpretation and prediction data is one of the important things in a toll road safety management. Numerous sets of data or information will be meaningless without any interpretation and the precise and accurately prediction. In line with this issue, the development of a model that can provide a good approach to the process of prediction is required. DM is a widely used approach in the interpretation of data in various disciplines. Through an approach of artificial intelligence (AI), DM has a huge potential to assist the interpretation and prediction process [12]. The utilization of computer and the availability of data which can be utilized for increasing the traffic safety should be continually developed [13], including the utilization of AI to determine the character of the driver and the vehicle in relation with the safety level [14]. In accordance with the library research, the AI approach and DM techniques have not been developed to devise predictive modeling of TRA prediction model.

The artificial intelligence (AI) is the development of the computer ability in interpreting various algorithms language. Soft computing method is performed by imitating the processes that found in nature, such as brain and natural selection [15]. Soft computing techniques allow the processing of data with uncertainly, imprecise and ambiguous characteristics. In the early mid-1960s, a new branch of computer science began to attract the attention of many scientists. This new branch, known as AI, can be defined as the study of how to make computer support the quality of people better performance in working world. To achieve these objectives, the computer is developed by imitating human behavior. In 1970, the AI is more focused on the development of expert systems which is designed to support decision-making through computed opinions of experts. Subsequently, in the 1990s, there was a shift of AI development where various issues were studied directly from the data [16]. Until now, the AI continues to grow and includes several methods and solutions for many disciplines. The scientific approach to civil engineering is increasingly familiar with the utilization of AI, such as in the development of transportation science [17].

The development of the information technology industry is very fast as indicated by the rapid growth of scientific data collection. Large size databases are not a problem anymore due to the advantage of computer technology with a range of major applications and supporters. Currently, all data are collected and stored in the database that can be a very valuable knowledge to support making decision and optimization of an action. Classical statistics have limitations to do the data analysis on a large number of data or when there is a complex relationship between the data variables. To overcome such limitations, the development of computer-based data analysis tools with greater capabilities and automaticity is required [18]. This effort is commonly defined as knowledge discovery in databases (KDD). Wang [19] mentioned the development of this tool is increasingly recognized with the term KDD DM.

Data Mining

The Data Mining (DM) as the extraction of useful knowledge from raw data has been received a lot of attention from both the research community and industry. Indeed, many case studies suggest that companies are increasingly investigating the potential of DM technology to deliver competitive advantage. DM in R-Tools uses an object-oriented language, so r-miner can present a collection of important functions that can be applied to various problems. The r-miner is equipped with various functions, including fit, predict, mining, mgraph, metric and mmetric. Regression is one of the main tasks of DM. This task consists of mapping several inputs as independent variables to a given numeric output as the dependent variable [6]. Several regression approaches of DM are used including LR, ANNs, and SVM.

In the present research, DM was implemented in the R software (http://www.r-project.org/) under the r-miner library [6, 15, 17]. R is an open-source computational environment and high-level language that integrates powerful

statistical and graphical features for databases. R adopts a very flexible and objective-oriented design. The r-miner library consists of several regression techniques, each of them has its own advantages. MR is relatively easy to interpret due to its additive nature while the ANNs and SVM require more computational resources but are capable of modeling more complex data.

METHODOLOGY

In this study, the purpose of the methodology was to provide a framework for the TRA prediction model. Key components of the model were introduced and their relevance to the model was discussed. The discussion focused on the accident involvement and severity models and their integration. Model form, error distribution, concerns, and uncertainty were also discussed. The purpose of the TRA prediction model is to predict the number of vehicle accident involvements on road sections and their resulted injury profiles. Developing a TRA prediction model involves several problems, the first was regarding with imbalance data. Imbalance data means within the data, there is a considerable difference between the observed sizes of one data set. The second problem involves data processing to create a training data set. The training data set has a set of multiple features that can affect the prediction result. The research aimed to solve these two problems and performed efficiently prediction processing using the DM algorithm. In this paper, a novel method to determine each step for classification model analysis of TRA prediction from large data is explained in detail.

Model Approach

One of the basic issues in safety management system is the development of TRA prediction models. Several accident prediction models have been proposed over the years, some of which are simple and others more complex. The success of safety management process in safety toll road depends on the TRA prediction which is executed by the system. To enhance the performance of safety toll road, successful prediction of TRA is considered importance. The researcher conducted the TRA prediction models based on SVM and ANN, which are empirical (data-driven) methods with the occurrence based on the non-linier programming. SVM and ANN are popular method approaches in DM. DM aims at the extraction of useful knowledge from raw data and has been received a lot of attention from both the research community and industry.

The DM in R-Project for statistical computing (R-tools) is an open-source computational environment and highlevel language that integrates powerful statistical and graphical features for data. R-Tools adopt a very flexible and object-oriented design [6]. The tool can be easily extended by the creation of packages. TRA prediction model was developed from the r-miner library with the inclusion of several variables. This research developed DM methodology for TRA prediction without any restrictive assumption by considering data input showed in Table 1 as input variables. TRA prediction was devised by following several processes. The *first* was the cleaning and examination of data that might be used in the TRA prediction model. When the data were "cleaned", the incorrect and irrelevant data were removed from the dataset. This process included correcting typos, ensuring the consistency of data formats and removing records with incomplete data.

Second, the examination of data. Initially, histograms or bar charts of the variables should be created to determine their frequency. Subsequently, the correlation between variables should be found. Knowing the distributions and correlations between variables would help the modeler to choose an appropriate form for the data, and to evaluate the model more effectively once it had been created. As the data were examined, the inconsistencies and inaccuracies might be found signifying that further cleaning of the data was required. Correlation refers to the relationship between two variables. A high correlation means that the two variables are closely related – as one variable changes, the other changes proportionally. If they are continuous variables, they form a line when plotted against each other. A very low correlation means that the two variables change randomly and are not associated. Most data fit somewhere between the two extremes. Correlation test is showed by matrix correlation.

Third, the selection of model type. After considering each model type that had been investigated previously (linier, dynamic, and artificial intelligence), the most appropriate type of model was selected. Creating the best TRA prediction model is usually an iterative process where the modeler may change various aspects of the model form to create the best model using the available data. The aspects of the model form that can be changed vary, ranging from model type to the software used to develop the model. In the case of TRA prediction model types, some factors that affect the model form include the base equation, the variables used in the model, and how those variables are grouped.

The remaining data will be used to evaluate the model. From the library research, this research chose the log regression as a comparator and DM as a proposed model.

R Tools

The R is a unit of software that integrates multiple facilities with the manipulation, calculation, and reliable graphics performance. R is a free version of the language S of software (free) that is similar with S-PLUS and has been widely used by researchers and academics in scientific activities. The R environment is an open source and high-level matrix programming language, and broadly used for statistical and data analysis. The R community is very active, and the new packages are continuously being developed, hence, in this perspective, the R is an open tool for worldwide sharing of RDCT, 2009. One of such packages is the r-miner [6], which is available in http://www3.dsi.uminho.pt/pcortez/ rminer.html. The purpose of this package is to facilitate the usage of data mining algorithms in both classification and regression tasks. Considering that R uses an object-oriented language, the r-miner presents a series of important functions than can be applied on these objects, such as fit – create and adjust a given DM model using a dataset; predict – returns the predictions for new data; mining – a powerful function that trains and tests a model under several runs and a given validation method; mgraph, metric and mmetric – which return several mining graphs or metrics.

EXPERIMENT AND DISCUSSION

As the case study, the JORR Toll Road was selected. The road network in JORR has a complete characteristic. The outer ring road in metropolitan area is characterized by the presence of the urban and non-urban area that serves as the main transportation lines and this corridor is passed by all types of vehicles. JORR toll road connects the port city of Jakarta with other cities in the west, south and east region of Jakarta.

Toll Road Accident Data

The models were verified using the data from *Jalan Lingkar Jakarta (JLJ)* Toll Road Company and Indonesia Toll Road Authority. The dataset from the JORR toll road included 1,725 results derived from 45 toll road segments and 10 input parameters, which are referred to as the influential parameters in empirical studies of TRA. Those parameters were given an order code as input as presented in Table 1 below.

TABLE 1. Input Code										
No	Code	Measurement	Description							
Input 1	TRA	Accident/1.000.000 vehicles km	Number of accident							
Input 2	ADT	Number	Daily Traffic Average							
Input 3	TD	Vehicles/km	Traffic Density							
Input 4	V	Km/hours	Traffic Speed							
Input 5	L	Number	Number of lines							
Input 6	STA	Station-segment	Position of accident							
Input 7	IRI	Meter/kilometer (m/km)	International Roughness Index							
Input 8	G	%	Gradient							
Input 9	Ramp	Number	Number of Ramps							
Input 10	Skid	Mm	Skid resistance							

Accident data were firstly recorded by the police and the JLJ officers. The day and location of each accident was also recorded in the data. In addition, the death in toll, the number of casualties, and the information about those related to each accident were also recorded in the data. The weather condition and shapes of lines were also provided in the data in the form of ID categories.

The TRA data were collected from the history of accident reports. Roughness and geometric data were collected every year using the NAASRA roughness meter and then compiled in the database program. The NAASRA roughness meter measures the relative movement between the axle and the vehicle body at a given speed. ADT data were obtained

from the annual traffic data. The data chosen for this research were the data from 2010 until 2016. The statistics of the parameters used in this paper are presented in Table 2.

TABLE 2. Statistics of the database values											
	TRA	ADT	TD	V	L	STA	IRI	G	Ramp	Skid	
Min	0.00	344,302	0.76	30.15	3	0	1.83	-0.80	0	0.33	
Max	10.00	439,859	0.89	98.46	3	45	3.75	0.90	1	0.67	
Average	1.59	398,406	0.80	67.65	3	-	2.86	0.20	0.25	0.54	
Dev.	1.62	34,903	0.03	17.78	0	-	0.32	0.01	0.02	0.12	

TABLE 2. Statistics of the database values

Toll Road Accident Prediction Model

Using DM with SVM and ANNs model will result in predictive TRA that can be used for each toll road segment on the JORR. In this study, we used the r-miner package of the R tool to test the SVM and ANNs model. DM process used 75% of the whole data in the training process and predicted 25% of the result with the fit model. The quality of the prediction obtained by applying DM model as presented in Figure 1, which shows the results of TRA prediction model. This performance was confirmed by the values of R^2 , Mean Absolute Deviation (MAD), and Root Mean Squared Error (RMSE). Among the 20 runs performed, the SVM hyper-parameters that best fit the data are $\epsilon = 0.08 \pm 0.02$ and $\gamma = 0.04\pm0.00$. The ANN hyper-parameters that best fit the data are $H = 4 \pm 1$.



FIGURE 1. The measurement of performances.

For each model, a total of 1.000 runs of a 20-cross validation procedure were applied. Three DM techniques were tested by using the dataset that had been previously described to perform the predictions. Among the tested three DM algorithms, the best results were obtained by ANNs models. The lower the MAD and RMSE values, the better the predictive model, while a perfect model should have an R runs and with the respective 95% confidence intervals according with the t-student distribution. The computed regression error metrics, for ANNs in terms of the MAD of 0.62 ± 0.01 , RMSE of 0.75 ± 0.03 and coefficient of determination (R²) of 0.87 ± 0.03 . The lower the MAD and RMSE values, the better the predictive model, while a perfect model, while a perfect model should have an R² value close to 1.0.

The predictive results (measured unseen data) were shown in terms of observed versus predicted scatterplots. In such scatterplots, the better the predictions, the closer they are to the diagonal line (perfect model). Figure 2 shows the scatterplots of TRA predictive models with ANNs, which reveal a good fit. The scatterplots show the results of

the learning stage modelling with the total amount of 72-month data, and Figure 2 (b) demonstrates the iteration of validation stage.

Figure 2 (a) illustrates some scatterplots showing the results of the learning stage modelling with the total amount of 60-month data, and Figure 2 (b) demonstrates the iteration of validation stage. The computed regression error metrics, in terms of coefficient of determination (R^2) was 0.96 ± 0.02 , while a perfect model should have an R^2 value close to 1.0. The results are presented in terms of the average of the runs and with the respective 95% confidence intervals according with the t-student distribution. Analyzing the results, a good fit is achieved by the ANNs model.



a. Learning Stage b. Validation stage

FIGURE 2. ANNs TRA outputs

In the validation stage, the features of the r-miner library were used to describe and obtain the relative contribution value of each input value. Figure 3 shows the importance attributed by ANNs to each input parameter that was obtained by applying sensitivity analysis.



FIGURE 3. ANNs TRA outputs.

Based on the figure, it is concluded that the ranking of the variables is in the following order: speed, traffic density, traffic volume, gradient, roughness, ramp number, skid resistance, location, and line number. At the learning and validation stage, speed has the highest contribution level to the value of the TRA.

CONCLUSION

This study developed a TRA prediction model to generate an optimal model of safety management in the toll road. Both objectives were achieved simultaneously. The factors which allegedly because accidents have been investigated for providing road safety, and subsequently, the accident prediction models which include relations between these factors have been established. The TRA reports from 2010 to 2016 were used to form the database. By using DM techniques and the ANNs technique, a model with excellent predictive capacity for large data was obtained. It is also possible to predict the TRA for several different locations using the DM techniques. The proposed model also allows for the identification of the input parameters that control the behavior of the speed, traffic density, traffic volume, gradient, roughness, ramp number, skid resistance, location, and line number. Furthermore, the results showed that the DM techniques, namely ANNs and SVM, have a strong relationship ($R^2 \ge 0.75$) and should be considered in the model. This model can be used in the TRA prediction model.

ACKNOWLEDGEMENTS

The authors are grateful to the editor and reviewers for very helpful comments on the earlier version of the paper, the Greater Jakarta Transportation Authority which supported this research, JLJ Toll Road Company, and Indonesia Toll Road Authority. We would also like to thank those in Universitas Internasional Batam, Indonesia, who provided assistance in this study.

REFERENCES

- 1. K. G. B. Akram, N. Farnoosh, and M. Viliam, (2014). "Highway Accident Modeling and Forecasting in Winter." Transportation Research Part A 59, 384–396.
- 2. BPS DKI Jakarta. (2017). "Transportation Statitistic of DKI Jakarta 2016", Katalog BPS 8301007.31.
- 3. A. C. Mbakwe, A. A. Saka, K. Choi and Y. Lee, (2016). "Alternative method of highway traffic safety analysis for developing countries using delphi technique and Bayesian network." Accident Analysis and Prevention 93, 135–146.
- 4. R. Arief, D. Vinayak, P. Anurag, and J. Rizky Adelwin (2017). "Predicting casualty-accident count by highway design standards compliance." International Journal of Transportation Science and Technology.
- 5. Park, Hyoshin. (2015). "Dispatching and relocation of emergency vehicles on freeways: Theories and applications." Ph.D. diss., University of Maryland, College Park.
- P. Cortez, (2010). "Data mining with neural networks and support vector machines using the r/rminer tool." Advances in Data Mining: Applications and Theoretical Aspects, 10th Industrial Conference on Data Mining, 83, Berlin, Germany: J In P. Pemer, editor.
- 7. S. Liao, P. Chu, and P. Hsiao, (2012). "Data mining techniques and applications A decade review from 2000 to 2011." Expert Systems with Applications 39, 11303–11311.
- 8. M. Dementyeva, P. R. Koster, and E. T. Verhoef, (2015). "Regulation of road accident externalities when insurance companies have market power." Journal of Urban Economics 86: 1-8.
- 9. Ç. M. Yasin, and T. Ahmet, (2015). "An Artificial Neural Network Model for Highway Accident Prediction: A Case Study of Erzurum, Turkey." PROMET-Traffic & Transportation 27, no. 3: 217-225.
- 10. Prieto, Faustino, E. Gómez-Déniz and J. M. Sarabia, (2014). "Modelling road accident blackspots data with the discrete generalized Pareto distribution." Accident Analysis & Prevention, 71: 38-49.
- 11. Leveson, Nancy. (2015). "A systems approach to risk management through leading safety indicators." Reliability Engineering & System Safety 136: 17-34.
- 12. W. Ertel, (2009). "Introduction to Artificial Intelligence." Springer, ISBN 9780857292988.
- T. Tambouratzis, D. Souliou, M. Chalikias, and A. Gregoriades, (2014). "Maximising accuracy and efficiency of traffic accident prediction combining information mining with computational intelligence approaches and decision trees." Journal of Artificial Intelligence and Soft Computing Research 4, no. 1: 31-42.
- 14. Meiring, Marthinus, Gys Albertus., and Myburgh, Hermanus Carel. (2015). "A review of intelligent driving style analysis systems and related artificial intelligence algorithms." Sensors 15, no. 12: 30653-30682.
- 15. J. Tinoco, A. G. Correia and P. Cortez, (2012). "Support vector machines applied to uniaxial compressive strength prediction of jet grouting columns." Computers and Geotechnics, 55: 132-140.

- 16. S. Liao, P. Chu and P. Hsiao, (2012). "Data mining techniques and applications A decade review from 2000 to 2011." Expert Systems with Applications 39: 11303–11311.
- A. Rifai, S. P. Hadiwardoyo, A. G. Correia and P. Pereira, (2016). Genetic Algorithm Applied for Optimization of Pavement Maintenance under OverloadTraffic: Case Study Indonesia National Highway. Applied Mechanics and Materials ISSN: 1662-7482, Vol. 845, Trans Tech Publications, Switzerland, pp 369-378.
- 18. F. A. Rahman, et al. (2014). "Knowledge Discovery Database (KDD)-Data Mining Application in Transportation." Proceeding of the Electrical Engineering Computer Science and Informatics, 1(1), 116-119.
- 19. X. Z. Wang, (2012). Data mining and knowledge discovery for process monitoring and control. Springer Science & Business Media.