MEKATRONIKA
e-ISSN: 2637-0883
VOL. 3, ISSUE 1, 79 – 88
DOI: https://doi.org/10.15282/mekatronika.v3i1.7155

ORIGINAL ARTICLE

The Analysis of Road Traffic Fatality Pattern for Selangor, Malaysia Case Study

N. Q. Radzuan[1], M. H. A. Hassan[1,*], K. A. Abu Kassim[2], A. A. Ab. Rashid[2], I. S. Mohd Razelan [3] and N. A. Othman[4]

[1]Faculty of Mechanical and Automotive Engineering Technology, Universiti Malaysia Pahang, 26600 Pahang, Malaysia.
[2]Malaysian Institute of Road Safety Research, 43000 Selangor, Malaysia.
[3]Faculty of Civil Engineering Technology, Universiti Malaysia Pahang, 26300 Pahang, Malaysia.
[4]Faculty of Electrical and Electronics Engineering Technology, Universiti Malaysia Pahang, 26600 Pahang, Malaysia.

ABSTRACT – Road traffic fatality is a burden towards low- and middle-income countries including Malaysia. Seeing that Selangor has the highest number of road traffic fatalities in Malaysia for the year 2019, therefore the state is selected as a case study. The aim of the article is 1) to understand the road traffic crash pattern and road traffic fatality pattern in Selangor 2) to determine the ability of 16 road traffic features in classifying road traffic fatality occurrence. The preliminary data screening shows that road traffic crash patterns and road traffic fatality patterns in Selangor have many similarities. However, both of them also have few dissimilarities such as crash time of occurrence, day of occurrence, number of vehicles involved in a crash, and type of vehicle first hit for the crash. Supervised machine learning algorithm in Orange data mining software was considered in this analysis. The analysed algorithms among others are neural network, random forest, decision tree, logistic regression, naïve Bayes, and support vector machine. Neural network was seen as the best algorithm to classify road traffic fatality occurrence with 97.0% classification accuracy outperform other algorithms. The result of the article can be used by the relevant traffic stakeholders to execute safety intervention in a more focused manner in Selangor to reduce the number of road traffic fatalities.

ARTICLE HISTORY
Received: 3rd April 2021
Revised: 14th May 2021
Accepted: 21st June 2021

KEYWORDS
Road traffic crash pattern
Road traffic fatality pattern
Artificial neural network
Supervised machine learning

INTRODUCTION

According to Department of Statistics Malaysia, there were 173,746 fatalities recorded in Malaysia for the year 2019 which the factors of fatality are from various causes. 3.8% of the fatality statistics was contributed by road traffic crashes which listed road traffic crashes among the 5 principal causes of fatality in Malaysia for 2019. By further focusing to fatality statistics by age group, road traffic crashes are the principal causes of fatality for the age group of 0 - 14 years old and 15 – 40 years old by 3.3% and 20.6% respectively outnumbering other listed causes such as pneumonia, leukaemia, accidental drowning and submersion cerebrovascular diseases, malignant neoplasm of breast, and ischaemic heart diseases [1]. The statistic is on par with the statement reported by the World Health Organization (WHO) in the Global Status Report on Road Safety 2018 that stated road traffic crashes are the number one killer of children and young adults aged 5 to 29 years old worldwide. The rate of road traffic crashes also happens three times higher in low to middle income countries than in high income countries [2].

The worldwide road safety trend showed most countries facing road traffic issues such as intoxicated driving, distracted driving, and unmet safety features application viz. motorcycle helmet, vehicle safety belt, and child restraint seat as the contributors to road traffic crashes [2], [3]. Pedestrians, cyclists, and motorcyclists which are under vulnerable road user sections were also being concerned worldwide. For example, 65% of fatalities in Cambodia were involving motorcycles [4]. The trend is no different than the current situation in Malaysia where motorcycle crashes justify for large percentage of all road traffic fatalities among other type of vehicles by 65.7% in 2018 as in Figure 1.

*CORRESPONDING AUTHOR | M. H. A. Hassan | mhasnun@ump.edu.my

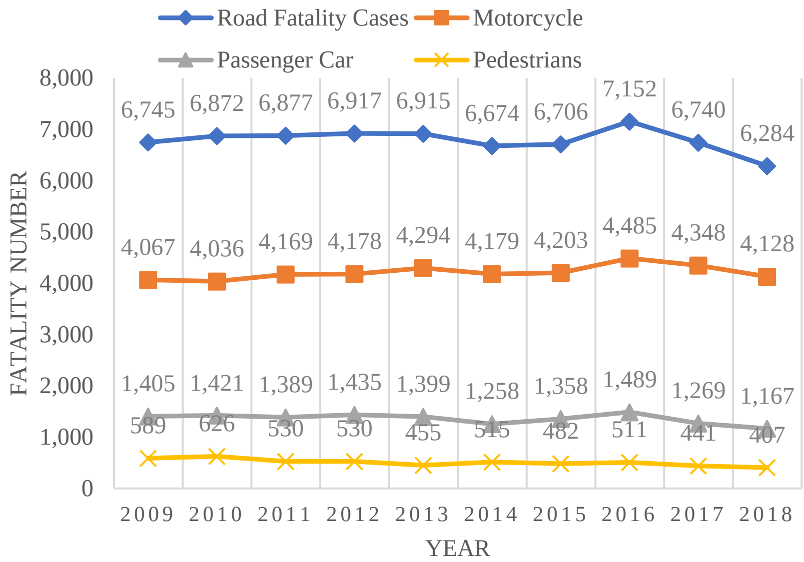


Figure 1: The road traffic fatality cases among transportation modes from 2009 until 2018.

Malaysia as a country that falls into low- and middle-income countries group had responded to United Nations General Assembly declaration to take initiative in reaching Sustainable Development Goal Agenda 2030 item 3.6 by reducing 50% of the number of road traffic fatality and road traffic injury through Decade of Action (UN DOA) for Road Safety 2011–2020 [5]. As the country's road traffic fatality and road traffic crashes in 2010 are 6, 872 and 414, 421 respectively, hence, the statistics should be reduced to 3, 436 and 207, 210 cases in 2020. Therefore, comprehensive implementations were made to five pillars: 1) Pillar 1: Road safety management 2) Pillar 2: Safer roads and mobility 3) Pillar 3: Safer vehicles 4) Pillar 4: Safer road users 5) Pillar 5: Post crash response. The trend of road traffic fatality cases presented in Figure 1 was seen declining starting 2016 which indicates it succeeded after 5 years of UN DOA implementation.

The Highest Statistics

In comparison among states and federal territories in Malaysia, 3.8% out of 27, 700 fatalities in Selangor were caused by road traffic crashes. This also means that Selangor has the highest road traffic fatality cases in 2019 as in Figure 2.

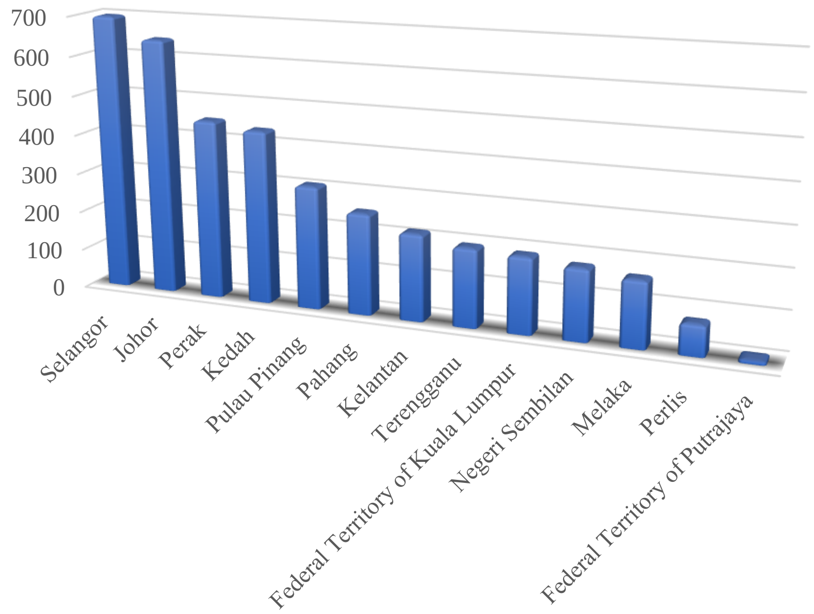


Figure 2: Road traffic fatality was listed among the five principal causes of fatality in all states and federal territories of Peninsular Malaysia in 2019 [1].

It is no-brainer that there are many road traffic fatalities in Selangor because it is densely populated by 6.53 million people. The state consists of Klang Valley, Putrajaya, and Greater Kuala Lumpur which offer massive job opportunities that drive migration from other states to work in Selangor. Kuala Lumpur as a centre of business attraction has been developed rapidly since the 1970s hence increases land prices, living cost, as well as road traffic congestion. Under the circumstances of optimal individual economic reason and preference, the people who work in Kuala Lumpur live distant from Kuala Lumpur, with Selangor and Putrajaya becoming the alternative. Moreover, the transport system within the area is considered good due to expansion of monorail, lightweight transit (LRT), mass rapid transit (MRT), and commuter rail system. However, travelling in and out of Kuala Lumpur by private vehicle is still dominant on the road. Thus, there would be a significant number of road traffic travellers to be involved in road traffic crashes.

Selangor state has become aware of road traffic researchers concerning the road traffic issue there. In a previous observational study, the usage of motorcycle safety helmet and vehicle safety belt was still low among vehicle passengers although it was enacted in road traffic regulation as mandatory [6]. Paiman et. al. also reported the usage of safety helmets for motorcyclist and child pillion riders is still scarce even though the national law compels motorcyclists to buckle up motorcycle safety helmets [7]. Harnen et. al. found that the motorcycle crashes at non-signalised intersection at 53 intersections around Selangor districts was contributed by the technicality of road traffic such as the intersections which located within commercial areas, the increase of non-motorcycle and motorcycle flows at major roads, the increase of approaching speed at the intersections, as well as the small of lane width, number of lanes and shoulder width [8].

Hartika et. al. considered road traffic black spot area in Shah Alam, Selangor as a case study to make a road traffic crash prediction model using machine learning algorithm [9]. In addition to that, Kamaruddin et. al. analysed and made comparison of road traffic crash trends from 2013 to 2017 between Selangor and Perlis which had the highest and the lowest road traffic fatality cases in the time frame. The authors concluded the main factors led to road traffic crashes were traffic behaviour, drivers, road, and structure [10].

OBJECTIVE AND METHODOLOY

The aim of this article is to apprehend the pattern of road traffic crash towards fatality in Selangor in 2019. The road traffic crash data were provided by the Traffic Division, Royal Malaysian Police, Bukit Aman. 1490 data in Selangor was screened first before being processed for further data analysis. The screening method was performed in order to eliminate repetition of same data and incomplete data, leaving a total of 1479 road traffic crash data comprising crashes of 1) fatal, 2) serious injury, 3) slight injury, and 4) non-injury / vehicle damage categories only. Table 1 shows the tabulation of the driver's road traffic fatality cases in Royal Malaysia Police Data is lesser than the data recorded by Department of Statistic Malaysia by 9 cases.

Table 1: Selangor road traffic fatality cases.

Year	Department of Statistics Malaysia Data	Royal Malaysia Police Data		
		Driver	Passenger	Pedestrian
2019	699	690	104	72
			866	

Even though this is a small indifference error by 1.29%, it is worth mentioning that any fatality disclosed in the police report is a fatality that occurred within 30 days after the road traffic crash occurred. Henceforth, the differences could be a man-made error while recording the data. On the other hand, the road traffic fatalities could be either one or multiple victims either the driver, passenger, or pedestrian meanwhile there could be multiple drivers resulting in road traffic fatality.

Firstly, the data was grouped into road traffic crash patterns. Table 2 indicates features as well as details of each feature for better comprehension.

Table 2: Details of road traffic crash pattern features.

No	Road Traffic Features	Details
1.	Time occurrence	Small hours / AM hours / PM hours / evening hours
2.	Day occurrence	Sunday / Monday / Tuesday / Wednesday / Thursday / Friday / Saturday

3.	Number of vehicles involved	(number)
4.	Type of crash	Fatal / non-fatal
5.	Type of road surface	Gravel / bituminous / concrete / unpaved road / others
6.	Type of traffic system	One-way / two-way / three-way / four-way or more
7.	Road design	Straight / curve / roundabout / four-leg intersection / T or Y intersection / junction / elevated intersection
8.	Quality of the road surface	Smooth / pothole / uneven or wavy roughness / sinking
9.	Road condition	Flat / Steep
10.	Road marking	Double lines / single line / u-turn / one-way / median strip / no road marking
11.	Road surface condition	Dry / wet / flood / spilled oil / sand / under construction /
12.	Type of the first hit in the traffic crash	Head-on / rear-end / sideswipe / side impact / hit on object or animal / hit pedestrian / rollover / skidded
13.	Speed limit	50 kmph / 70 kmph /80 kmph /90 kmph / 110 kmph / others
14.	Weather	Fine / strong wind or crosswind / dark with streetlight / dark without streetlight
15.	Light visibility	Daytime / dawn or dusk / night / dark visibility without streetlight
16.	Type of road	Expressway / federal road / state road / municipal road / others
17.	Type of area	City / town / small town / rural area

All features are in categorical type except the number of vehicles involved in item 3. The road traffic fatality data was then changed from numeric into categorical data with the code ‘0’ as non-fatal crash or ‘1’ as fatal crash. The data then was summarized through descriptive statistics to understand the pattern of road traffic crashes.

All 1479 data is categorized into a group of 16 independent variables. These independent variables were considered into classification analysis to determine their ability in predicting one dependent variable by using Orange data mining software. The dependent variable is the type of crash either non-fatal crash or fatal crash. As the study has a large data set, therefore the sampling type was fixed to 70% proportion of data for training analysis and balance of 30% data were stored for testing analysis. This is done by Orange data mining itself.

Orange Data Mining Software

Orange data mining software version 3.20.1 (Orange 3) is a medium application for machine learning and a powerful data mining tool which is perfect to conduct the data analysis. Orange 3 is thorough and versatile with data management, data preprocessing, visual programming, model training and evaluation. As the application is an accession of classification and regression models therefore machine learning algorithms such as decision trees, random forest, neural network, support vector machine, naïve bayes, linear regression, and logistic regression utilized in Orange 3. The Orange 3 workflow for the study is presented in Figure 3.

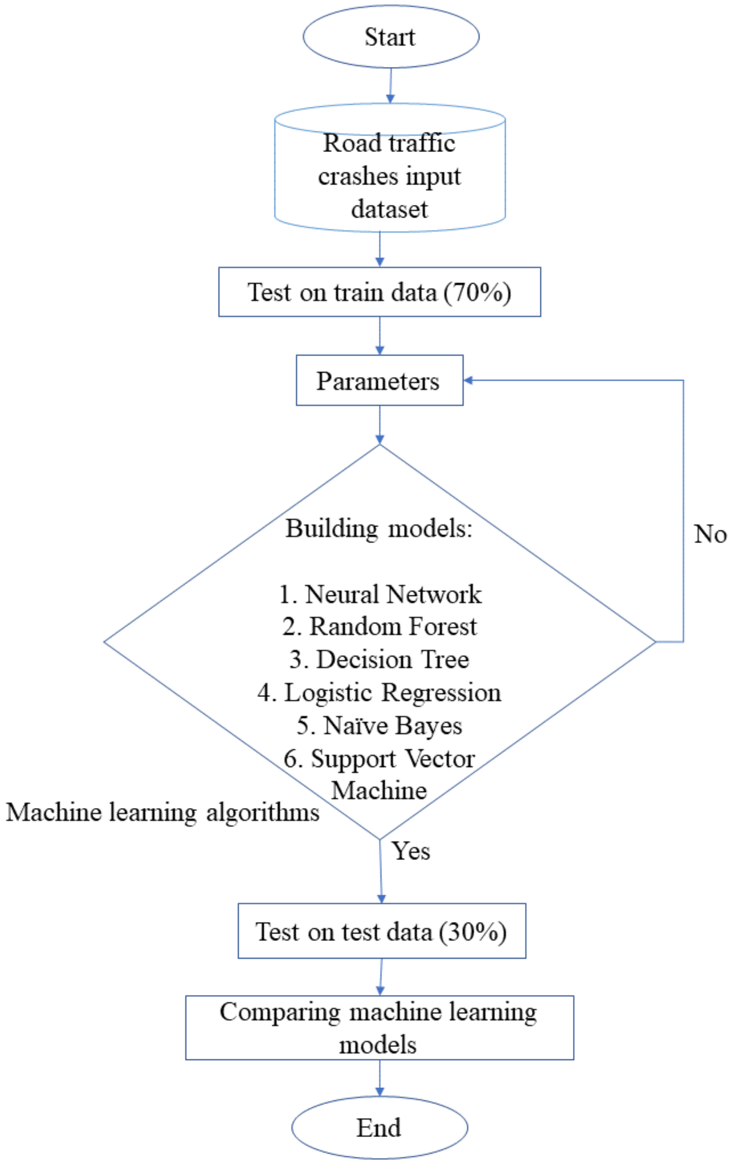


Figure 3: Experimental setup.

Classification Model

The model training in Orange 3 could be classification model or regression model depending on the expected output. The provided data of classification model development was processed with the purpose it may be recognised into its suitable data categorisation cluster. Machine learning methods are normally supervised and unsupervised techniques. It is called a supervised learning algorithm when the independent and dependent variables are recognized. The pattern of independent variables was visualized to interpret the dependent variable. Examples of supervised learning algorithms are decision tree, random forest, neural networks, and support vector machines [11]–[13]. Unlike supervised learning algorithms, the unsupervised learning algorithm on the other hand has no specific method to be dealing with. Any alikeness in the raw data will be classed into its data cluster group. The unsupervised learning algorithm is ideal to ascertain meaningful data classes or patterns in the mined data. Unsupervised learning algorithms can be multidimensional scaling, Manifold Learning, Louvain Cluster-ing, k-means clustering, hierarchical or non-hierarchical clustering, and t-SNE [14], [15].

Neural network consists of neural network topology as shown in Figure 4. The topology has independent variable input layers, hidden layer which placed neurons, and dependent variable output layer [16]–[18].

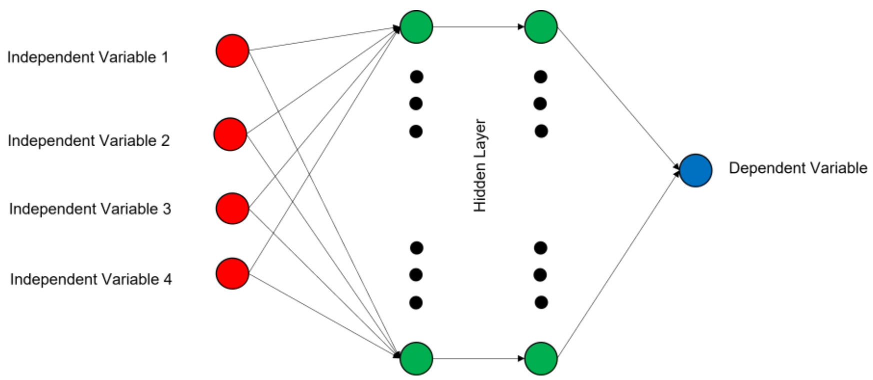


Figure 4: Example of multi-input single output (MISO) Neural Network model with a network topology.

Decision trees can process independent variable input data in many types for either categorical or numerical data in a tree structure [19]–[22]. It also can work with any data type such as nominal, numerical, and alphabetical. The benefit of the algorithm is its ability to handle the independent variable data that comprises missing values and errors. The independent variable data also can be stored easily before being classified further through the algorithm. Its decision node has at least two branches meanwhile leaf node serves as the classification or the decision. The branches will keep on being divided until it cannot be split anymore. The best decision node in the model development process which has good correlation with its best predictor is called root node. Example of a decision tree through literature is shown in Figure 5.

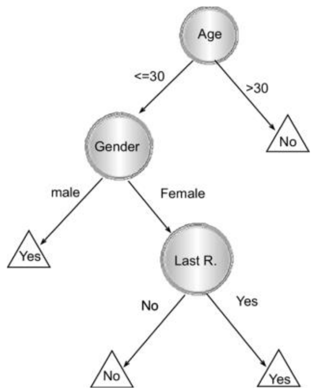


Figure 5: Example of decision tree data presentation [19].

If the analysis has many decision trees altogether in its group, then it is called a random forest. The implausible independent variables will be improved and further becomes accurate. The benefit of random forest is its capacity to estimate accurate predictions even in large data sets. Random and small-depth subset trees will be created before being combined in order to avoid overfitting [23]–[26].

Logistic Regression is commonly used when the dependent variable is in categorical data especially in binary code (0 or 1), for example group of emails either spam or not spam, or for this case study the various crash cases could be non-fatal or fatal crashes. However, logistic regression is also able to predict multinomial logistic regression (e.g., non-injury, slight injury, severe injury, or fatal). The analysis of the logistic regression algorithm is based on logit function probability [27].

The Naïve Bayes algorithm computes the probability of a class in independent variables. The algorithm later may predict the class of the dependent variable with the highest following probability. The algorithm is robust to data randomness and less responsive towards irrelevant data features [28].

Support vector machine is a supervised machine learning algorithm that separates data into at least two categories. Then, the algorithm creates an N-dimensional hyperplane from a margin point in the data plane. The best support vector machine result is the one with the largest distance of the margin divided by optimal hyperplane [29]. Figure 6 shows marginalization of support vector machines from literature where groups of purple dots and red squares are divided by an optimal hyperplane with distance of margin, w.

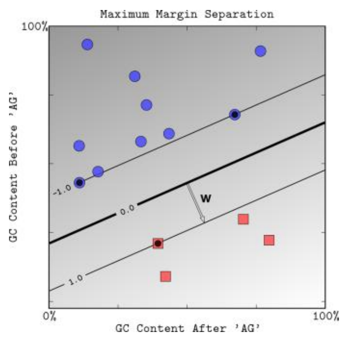


Figure 6: Example of margin separation in support vector machine [30].

RESULT AND DISCUSSION

A descriptive statistic of road traffic crash and fatality cases was performed for Selangor case data in 2019. It is vital to understand the pattern of fatality crashes that happened at one place as traffic authorities may perform a conducive and targeted intervention. Over half of the provided data is a road traffic fatal crash that occurred in Selangor. Table 3 is the comparison of mode for road traffic crash and fatality in Selangor for the year 2019. There are few differences between crash pattern and fatality pattern in terms of time occurrence, day occurrence, number of vehicles involved, and type of the first hit crash.

Table 3: Comparison of the most repetition characteristics in road traffic crash and road traffic fatality in Selangor.

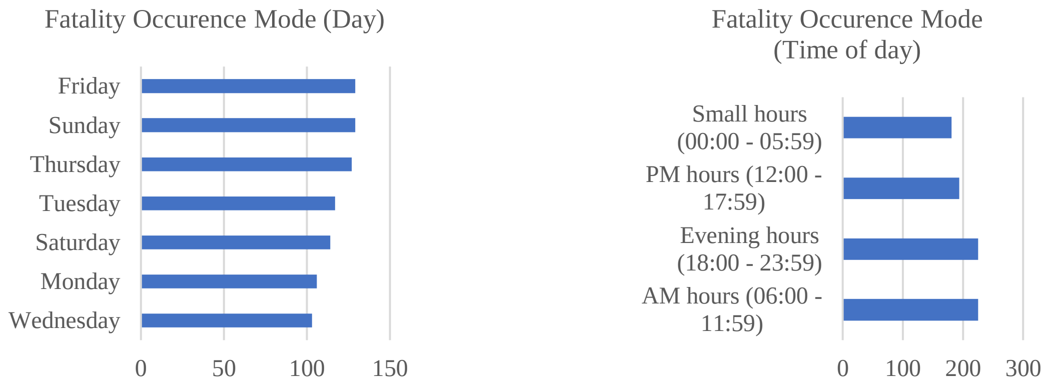
Criteria	Road Traffic Crash Pattern	Road Traffic Fatality Pattern
Number of data / crashes occurrence	1479 data	825 data
Time occurrence	Evening hours	AM hours & Evening hours
Day	Thursday	Sunday & Friday
Number of vehicles involved	2	1
Type of crash	Fatal crash	
Type of road surface	Bituminous road	
Type of traffic system	Two-way-road	
Road design	Straight road	
Quality of the road surface	Smooth surface	
Road condition	Flat road	
Type of road marking	Double lines	
Road surface condition	Dry	
Speed Limit	Others	
Type of the first crash	Head-on crash	Skidded
Weather	Fine weather	
Light visibility	Daytime	
Type of road	Municipal road	
Type of place	Rural area	

Road Traffic Fatality Pattern

The nature of road fatality crashes for Selangor was one fatality case occurred in a fatality crash however the maximum fatal victims for 2019 cases were five fatalities in the crash. Fatality crashes mostly occur on Sunday and Friday with 129 cases on each day. The fatality crashes day occurrence of Friday is in line with a statement by Abas et. al. that heavy congestion normally occurred during weekdays [31].

(a) It is known that road traffic congestion bears a higher risk of crash occurrence likely to happen. Wen et. al. further reported heavy congestion that happens on weekdays are due to drivers commuting during AM and PM peak. The heaviest congestion of AM peak usually takes place on Monday while the heaviest congestion of PM peak usually happens on Friday [32]. It is also important to note that AM peak and PM peak generally start between 7.00 AM to 10.30 AM and 4.30 PM to 7.00 PM respectively. Therefore, it could be resulting in a fairly high road traffic fatality occurrence at these hours. This justifies Friday, AM and evening hours dominated fatal crashes as shown in

Figure 7. Evening hours and AM hours share the same total of fatalities by 225 cases.



(b) (b)

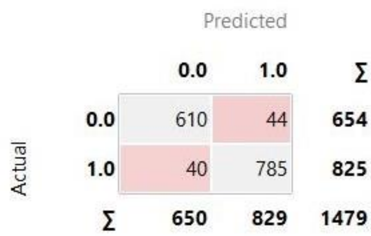
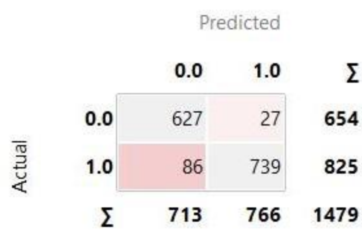
Figure 7: The overview of fatality occurrence in day and time of day.

The special scenario for Selangor was a bit different as the mode of fatality occurs on both Sunday and Friday. The travel made on Sunday might be a non-work-related trip. The geography of Selangor at the centre of Peninsular Malaysia has benefited drivers to easily make long distance travel across states for either vacation or 'balik kampung' (return to hometown) [33]. In addition to that, weekend rest days in four states in Peninsular Malaysia (i.e. Kedah, Johor, Terengganu, and Kelantan) falls on Friday and Saturday instead of Saturday and Sunday. This could explain a significant high fatality occurrence on Thursday until Sunday.

The Prediction of Road Traffic Fatality

A set of 1479 road traffic crash data was being performed an analysis to check the ability of the features to predict the road traffic fatality crash occurrence. A descriptive statistic summarization was checked for determining the normal distribution of the data. The features with normal distribution were selected into a prediction model using Orange 3. Comparison was made between neural network, random forest, decision tree, logistic regression, naïve bayes, and support vector machine as listed in Table 4.

Table 4: The classification accuracy in predicting road traffic fatality cases.

Method	Classification accuracy of train data	Classification accuracy of test data
Random Forest		
	93.9%	94.3%
Decision tree		
	92.4%	92.4%

Random Forest (Predicted / Actual):

	0.0	1.0	Σ
0.0	610	44	654
1.0	40	785	825
Σ	650	829	1479

Decision tree (Predicted / Actual):

	0.0	1.0	Σ
0.0	627	27	654
1.0	86	739	825
Σ	713	766	1479

Model	Train accuracy	Test accuracy
Neural Network	78.2%	78.8%
Logistic Regression	68.1%	68.1%
Naïve Bayes	66.1%	66.1%
Support Vector Machine	50.6%	50.6%

Neural Network

Actual \ Predicted	0.0	1.0	Σ
0.0	472	182	654
1.0	132	693	825
Σ	604	875	1479

Logistic Regression

Actual \ Predicted	0.0	1.0	Σ
0.0	373	281	654
1.0	191	634	825
Σ	564	915	1479

Naïve Bayes

Actual \ Predicted	0.0	1.0	Σ
0.0	390	264	654
1.0	238	587	825
Σ	628	851	1479

Support Vector Machine

Actual \ Predicted	0.0	1.0	Σ
0.0	281	373	654
1.0	357	468	825
Σ	638	841	1479

The best model in predicting road traffic fatality cases in road traffic crash data is the random forest with 93.9% classification accuracy of train data and 94.3% classification accuracy of test data. The confusion matrix shows the actual data that was classified and the prediction that was made based on the features given. The confusion matrix allows researchers to check machine learning algorithm effectiveness. By taking an example of the best prediction model, the confusion matrix depicts the random forest algorithm correctly identifying 610 non-fatal road traffic crash cases and misidentifies 44 non-fatal crashes as fatal cases. 88 fatal crashes were identified correctly with two mistakes only.

CONCLUSION

The study was performed with an intention to understand the road traffic fatality cases that occurred in Selangor for the year 2019. The independent variables of road traffic crashes pattern and road traffic fatality pattern were revealed from police data. The study also shows that the variety of independent variables is able to classify either non-fatal road traffic crashes or fatal road traffic crashes. This information is useful for the traffic stakeholders to focus which attributes needed to be concerned for safety intervention execution. The future work should be expanding the analysis to other states as well.

ACKNOWLEDGEMENT

The authors would like to acknowledge ASEAN NCAP, FIA Foundation, Global NCAP, OEMs and the Society of Automotive Engineers Malaysia (SAE Malaysia) for funding this study under the ASEAN NCAP Holistic Collaborative Research (ANCHOR III) grant. Also, the authors are thankful to the Universiti Malaysia Pahang for providing the facilities to conduct the study.

REFERENCES

[1] Department of Statistics Malaysia, "Penerbitan Perangkaan Sebab Kematian Malaysia, 2020," Putrajaya, 2020.
[2] WHO, "GLOBAL STATUS REPORT ON ROAD SAFETY 2018," Geneva, 2018.

[3] United Nations, “Road Safety - Considerations in Support of the 2030 Agenda for Sustainable Development Considerations in Support of the 2030 Agenda for Sustainable Development,” 2018.

[4] United Nations, “Transport for Sustainable Development The case of Inland Transport,” New York, 2015.

[5] IRTAD, “Road Safety Annual Report 2019,” 2019.

[6] J. Oxley, S. O’Hern, and A. Jamaludin, “An observational study of restraint and helmet wearing behaviour in Malaysia,” *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 56, pp. 176–184, 2018.

[7] N. Paiman R *et al.*, “Safety helmet use rate and child pillion riders’ characteristics in Selangor,” Kajang, 2013.

[8] S. Harnen, S. V. Wong, R. U. Radin Sohadi, and W. I. Wan Hashim, “Motorcycle Crash Prediction Model for Non-Signalized Intersections,” *IATSS Res.*, vol. 27, no. 2, pp. 58–65, 2003.

[9] H. A. Hartika, M. Z. Ramli, M. Zaihafiz, Z. Abidin, and M. Hafiz, “Study of Road Accident Prediction Model at Accident Blackspot Area : A Case Study at Selangor,” vol. 3, no. 5, pp. 466–470, 2017.

[10] M. K. A. Kamarudin *et al.*, “Road Traffic Accident in Malaysia: Trends, Selected Underlying, Determinants and Status Intervention,” *Int. J. Eng. Technol.*, vol. 7, no. 4.34, p. 112, 2018.

[11] A. S. Mohanty, K. C. Patra, and P. Parida, “Toddler ASD Classification Using Machine Learning Techniques,” *Int. J. Online Biomed. Eng.*, vol. 17, no. 07, pp. 156–171, 2021.

[12] S. Puttinaovarat, S. Pruitikanee, J. Kongcharoen, and P. Horkaew, “Machine Learning Based Emergency Patient Classification System,” *Int. J. online Biomed. Eng.*, vol. 17, no. 5, pp. 133–146, 2021.

[13] M. Hamim, I. El Moudden, M. D. Pant, H. Moutachaouik, and M. Hain, “A Hybrid Gene Selection Strategy Based on Fisher and Ant Colony Optimization Algorithm for Breast Cancer Classification,” *Int. J. online Biomed. Eng.*, vol. 17, no. 2, pp. 148–163, 2021.

[14] Z. Liu, “Unsupervised learning with graph theoretical algorithms and its applications to transcriptomic data analysis,” Imperial College London, 2018.

[15] C. A. Muñoz López, S. Bhonsale, K. Peeters, and J. F. M. Van Impe, “Manifold Learning and Clustering for Automated Phase Identification and Alignment in Data Driven Modeling of Batch Processes,” *Front. Chem. Eng.*, vol. 2, no. November, pp. 1–18, 2020.

[16] S. M. Warren and P. Walter, “A Logical Calculus of the Ideas Immanent in Nervous Activity,” *Bull. Math. Biophys.*, vol. 5, pp. 115–133, 1943.

[17] D. O. Hebb, *The Organization of Behavior*. New York: John Wiley & Sons Inc, 1949.

[18] F. Rosenblatt, “The Perceptron: A Perceiving and Recognizing Automaton (Project PARA). Report No. 85-460-1,” Cornell Aeronautical Laboratory, New York, 1957.

[19] C. A. Survey, L. Rokach, and O. Maimon, “Top-Down Induction of Decision Trees,” *IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev.*, vol. 35, no. 4, pp. 476–487, 2005.

[20] M. Somvanshi, P. Chavan, S. Tambade, and S. V. Shinde, “A review of machine learning techniques using decision tree and support vector machine,” in *Proceedings - 2nd International Conference on Computing, Communication, Control and Automation, ICCUBEA 2016*, 2017, pp. 1–7.

[21] R. Messenger and L. Mandell, “A modal search technique for predictive nominal scale multivariate analysis,” *J. Am. Stat. Assoc.*, vol. 67, no. 340, pp. 768–772, 1972.

[22] L. . Breiman, J. . H. . Friedman, R. . A. . Olshen, and C. . J. . Stone, “Classification and Regression Trees,” *Biometrics*, vol. 40, no. 3, 1984.

[23] L. E. O. Breiman, “Bagging Predictors,” *Mach. Learn.*, vol. 26, no. 2, pp. 123–140, 1996.

[24] Y. Amit and D. Geman, “Shape Quantization and Recognition with Randomized Trees,” *Neural Comput.*, vol. 9, no. 7, pp. 1545–1588, 1997.

[25] L. Breiman, “Randomizing outputs to increase prediction accuracy,” *Mach. Learn.*, vol. 40, no. 3, pp. 229–242, 2000.

[26] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[27] F. Thabtah, N. Abdelhamid, and D. Peebles, “A machine learning autism classification based on logistic regression analysis,” *Heal. Inf. Sci. Syst.*, vol. 7, no. 1, pp. 1–11, 2019.

[28] A. I. Dale, *A History of Inverse Probability: From Thomas Bayes to Karl Pearson*, 2nd editio. New York: Springer-Verlag New York, 1999.

[29] A. Y. Chervonenkis, “Early History of Support Vector Machines,” *Emperical Inference*, pp. 13–20, 2013.

[30] A. Ben-Hur, S. O. Cheng, S. Sonnenburg, B. Schölkopf, and G. Rätsch, “Support vector machines and kernels for computational biology,” *PLoS Comput. Biol.*, vol. 4, no. 10, 2008.

[31] M. A. Abas, S. Rajoo, and S. F. Zainal Abidin, “Development of Malaysian urban drive cycle using vehicle and engine parameters,” *Transp. Res. Part D Transp. Environ.*, vol. 63, no. June, pp. 388–403, 2018.

[32] H. Wen, J. Sun, and X. Zhang, “Study on Traffic Congestion Patterns of Large City in China Taking Beijing as an Example,” *Procedia - Soc. Behav. Sci.*, vol. 138, no. 0, pp. 482–491, 2014.

[33] Z. M. Jawi *et al.*, “Automotive Consumerism : A Study of Car User ’ s Practices & Behaviour in Klang Valley , Malaysia,” 2017.