Multi-Layer Perceptron Approach for Predicting Road Traffic Accidents Based on the Driver Age

SN Ariyathilake[1], RMKT Rathnayake[2]

[1] *Department of Computing & Information Systems, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka*

[2] *Department of Physical Science & Technology, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka*

[#1] SN Ariyathilake; pbsnariyathilake@std.appsc.sab.ac.lk

[#2] RMKT Rathnayake; kapila.tr@gmail.com

Abstract— Road accidents are grown to be a huge disaster over the globe despite geographical boundaries. It is challenging to eradicate the road accident occurring rate but necessary countermeasures can be taken to reduce the accident rate. The driver of the vehicle can be considered as the main responsible person for the accident. In Sri Lanka most of the times drivers try to escape from the accident location. To take necessary legal actions, identifying the driver of the accident is the main consideration. Therefore, the objective of this study is to identify the age group of the driver by using several road accident factors, which can be successfully applied for driver identification. The researcher has used 3 attributes; Direction of vehicle moving, a human crash factor related to the accident and the location of the accident (nearest km post) to identify the driver age group. This research develops a Multilayer Perceptron (MLP) approach to identifying the driver age group with considerable accuracy. Decision tree approach has used to compare the results of the MLP model and the MLP outperformed with 85.52% accuracy. Sensitivity analysis by using essential hyper-parameters was performed in order to identify the best MLP model. For the prediction, analysis results revealed that the 140 epochs with RMSprop optimizer can increase the accuracy of the model and the low learning rates optimized the accuracy of the model. The developed model will be effectively used for identifying driver age limit and the model will be able to use by relevant authorities to make accurate decision making towards the accident investigations.

Keywords— **Multilayer Perceptron (MLP), RMSprop, Decision tree**

I. INTRODUCTION

With the Rapid development of the industrial and transportation sectors road accidents have increased dramatically. A considerable amount of accidents are happening on each day in Sri Lanka, despite the rural or urban conditions. The rate for the accidents for one day has increased according to statistics reports of the ministry of transportation (*Traffic Police - Road Traffic Accidents*, 2018).There are over 5 % of increases in road accidents in 2017 than in 2016. Moreover 1000 accidents per week happening in Sri Lanka and out of them there is a high possibility, that 50 of them are dead and also 150 has admitted to the hospital daily. As average 4000 deaths per year happening in Sri Lanka (*Traffic Police - Road Traffic Accidents*, 2018).Accidents can happen in different situations, in different ways at different times. Therefore prediction of road accidents can consider a better way to reduce the number of accidents occurring in Sri Lanka. Not only the normal roads of Sri Lanka but also in the highway road system of Sri Lanka exposed to the road accidents. So the analysis of these road accidents is a compulsory Social need.

Driver's age and the road accident occurring has very strong relationship according to the statistical analysis of the Sri Lanka police. Young drivers are becoming major victims of road accidents very often when considering Sri Lanka. Nearly 7929 of young people had admitted to the hospital in 2017 and the due to motorcycle 3933 victims had hospitalized among them. In addition to that the probability of driver run away from the accident incident has become high(Sunday times Sri Lanka, 2017).

In some cases police officers only decide the driver fault only by looking at the accident and the eyewitnesses. Analysis of the records is not done by the authorities. Hense sometimes judgments regarding the driver age get wrong. In order to predict driver age of the accident it is necessary to have a model which can determine the driver age limit. By using the age provided by the model can identify the run-away driver into some extend. So from that accurate judgments can be made and only the responsible party for the accident will be punished.

Applying Machine learning in road accident prediction, can identify the hidden and uncovered pattern within the road accident data set and can accurately predict the different circumstances of an accident occurring. Machine learning is said to be studying on pattern recognition and learning through the experience which is related to data mining as well as the deep learning side(Vogt, 2019). Machine

learning techniques are used by many types of researches because machine learning can learn from data sets and make accurate predictions on different domains(Vogt, 2019). Many types of research have indicated that the machine learning techniques are a success in uncovering abnormal behaviours within the data sets rather than the normal behaviours (Lee, Chung and Hwang, 2016)(M.~Chong, A.~Abraham and M.~Paprzycki, 2005).

This study focuses on predicting an age group of the driver who engages with the accident by using several factors of the road accident. Most of the time, drivers are trying to get away from the accident location due to several reasons. So identifying the driver age group before taking any countermeasures will be effective in decision making in road safety. After identifying the age group of the driver, the responsible driver will be able to take to the custody as well. In order to predict the age group model will be suggested by using several techniques.

II. METHODOLOGY FOR THE STUDY

Driver age is one of the major contributing factor to road accidents. By identifying the age group of the driver who is responsible for the accident is very important when it comes to taking necessary legal actions. Several machine learning methods have applied to take accurate predictions and to compare the accuracy of each algorithm.

A. Data Set and Study Area

The road accident data which need for the model development was gathered from Traffic police headquarters Colombo Sri Lanka. Nearly 193217 data were gathered from 2012 to 2016 which contain all the accidents happened in Sri Lanka at that time period. In order to perform study accurately, Colombo – Batticaloa road was selected for the study because it is the most vulnerable road according to the data set. The road consists of 426km with approximately 300 accidents per year. That road can be considered the most accident-prone road in Sri Lanka. Drivers who engage with the accidents are categorized to several age group. Starting from age 16 to 70 drivers were categorized into 10 groups. So the model will predict whether the responsible driver is in which age group.

B. Data Pre-processing

Data pre-processing can be considered a very valuable step in model development. In order to fit the data into the model, data must be without missing values, outliers and extreme values. Python used for the data preprocessing tasks. Sci-py, numpy, pandas, sklearn libraries are used for the data preprocessing. Interquartile range was used to identify the outliers within the data set and outliers were replaced with a value which resides between the interquartile ranges. The interquartile range is a measure of statistical dispersion which resides between upper and lower quartiles.

$$IQR = Q3 - Q1 \quad (1)$$

Any data which is not between these quartiles consider as the outlier and the replacement of the outlier was done within a value within this range(Renze, 2019).

C. Feature Selection

Features are the most important for the model. In order to have a better model, features play a major role. To identify the features for the model, Tree classifier algorithm with feature importance properties was selected. It provides the score for each feature with regarding its relationship between the target variable.

To calculate Feature importance node impurity and the probability for that link is considered. Node probability will be able to calculate by samples that reach to the node and the total samples. The higher the value gets the importance of the feature increase(*The Mathematics of Decision Trees, Random Forest and Feature Importance in Scikit-learn and Spark*, 2018).

Nodes importance calculates using the Gini index.

$$nij = wjCj - wleft(j)Cleft(j) - wright(j)Cright(j) \quad (2)$$

ni sub(j)= the importance of node j

w sub(j) = number of samples reaching to node j

C sub(j)= impurity of node j

left(j) = node j left child

right(j) = node j right child

Next importance of each feature calculates by,

$$fij = \sum j.nij / \sum k.nik \quad (3)$$

fi sub(i)= the importance of feature I

ni sub(j)= the importance of node j

Out of 20 variables 3 of them selected as the most valuable feature for the model. The age limit of the drivers was set as the output classes. 10 age limits were used to get prediction regarding the driver age. Relevant age group according to the data is predicted by the model.

The selected features for the model are,

i. Direction of vehicle moving

ii. Human crash factor related to the accident

iii. The location of the accident (nearest km post)

D. Multi-Layer Perceptron

MLP is the one of neural network type which mostly used by several researchers in order to archive high accuracy with models. MLP consist of interconnected neurons which provide non-linear relationships between input and output features. Weights are the medium which connects nodes inputs of the nodes are modified according to the weights by an activation function(Gardner and Dorling, 1998).

The neural network is described as a graph. Hence,

$$G = (V,E) \quad (4)$$

Eight functions for the edges,

$$w: E \rightarrow R \quad (5)$$

Each neuron is modeled as a simple scalar function,

$$\sigma: R \rightarrow R \quad (6)$$

There are 2 possible functions for σ

i. Sign function σ (α) = sign(α)

ii. Sigmoid function σ (α) = 1/(1+exp(-α))

α – activation function of the neuron

Neuron input is gained by taking a weighted sum of the outputs of all the neurons which are connected to the input. Where weight is w(Lee, Chung and Hwang, 2016).

Xj - input signals

Wkj- synaptic weights

Vk- summing output (value)

$$Vk = \sum Wkjxj \; P \; j=1 \quad (7)$$

E. Decision Tree Classifier

A CART decision tree is a classification and regression tree. In order to classify instances Gini index or the twinge criterion can be used. Each and every node of the decision tree has two edges. Impurity criterion of an edge is splitting the data by most suitable numeric value or categorical value. CART Algorithm will identify the most important features and others(*The Mathematics of Decision Trees, Random Forest and Feature Importance in Scikit-learn and Spark*, no date).The mathematical formulation of the decision tree is shown below. In given training vectors,

$$xi \in Rn \quad i= 1, 2,\ldots \quad (8)$$

Label vector

$$y \in Rl \quad (9)$$

Decision tree partitioned the samples with the same labels together.

Data at m node = Q

For each time Split

$$\Theta = (j, tm) \quad (10)$$

(which feature j and tm as the threshold)

Partition data in to left and right side, as

Qleft (Θ) and Qright (Θ) subsets,

Then

$$Qleft(\Theta) = (x, y)/\ xj < = tm \quad (11)$$

$$Qright(\Theta) = Q/\ Qleft(\Theta) \quad (12)$$

III. RESULTS OF THE STUDY

Comparative results which are taken from using several machine learning approaches is illustrated in the result section. Several python libraries were used to create approaches towards the driver age prediction. Basic libraries like pandas, numpy, sklearn, scipy were most commonly used with model development. MLP model was developed based on Tensorflow Google API.Results which was gained through the algorithms are shown below.

A. Multi-Layer Perceptron Model

In order to find the best model within the multilayer perceptron, several approaches were conducted. By using several epochs, different learning rates and different optimizers, best and accurate model was selected. Sensitivity analysis results of the MLP model is shown further. After performing the following sensitivity analysis final MLP model selected.

The developed model comprises of two hidden layers and an output layer with drop out function. The input dimension to the first hidden layer is equal to the Four. Outputs of the first hidden layer are results from the activation function, RELu(Rectified Linear Unit). RELu is applied to the weighted sum of both input and the output of the layers. The second hidden layer has used TANh as the activation function. Fully connected layers were trained in order to classify the two classes in the output layer. Two dropout functions were used to each hidden layer to maximize accuracy. In order to reduce the mean of the error, softmax_cross_entropy_with_logits loss function has used with RMSPropOptimizer. 5 folds in k fold validation were used to evaluate the model with validation.

1) Hyper Parameter Tunning with Different Epochs: Neural

network model trained by changing the number of epochs in the algorithm. Starting from the 10 epochs to 200 epochs were tested with the neural network model. The neural network which is consist of two hidden layers with 20 and 10 neurons for each were tested. Data set was split into an 80:20 ratio when performing the evaluation with epochs. Figure 1 shows the accuracy variation with respect to the 10 to 200 epoch (Number of iterations). According to figure 1, model accuracy on the testing and validation has increased after every iteration (epoch). But in the 100 epoch both validation and testing accuracy has decreased. Drop down function can be considered as the reason for the fluctuation. At the 140 epoch both testing and validation accuracy has increased to some point (0.8118 and 0.8552). It is the point which has gets maximum accuracy for the validation. Testing accuracy is also considerable high. According to the graph, at 140 epoch shows the best accuracy rather than the other epochs.

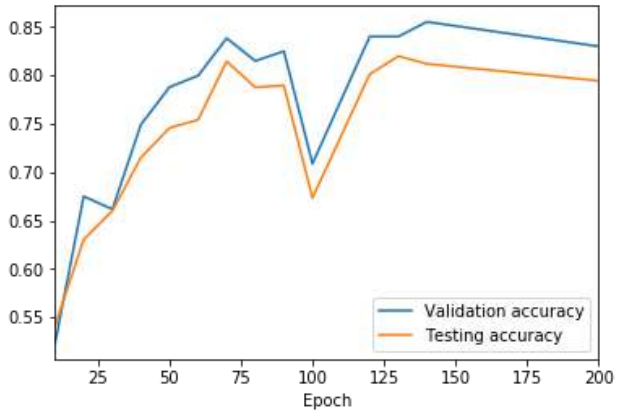

Figure 1 Accuracy performance of the MLP approach

Validation loss and the testing loss variation with the number of epochs is shown in Figure 2. Reducing loss means maximizing accuracy. Both validation loss and the testing loss reduce over time and at the 100 epoch, validation loss has increased but in the 140 epoch it has reduced to a certain limit. Then after several epochs both validation and the testing loss has dropped to the considerable amount (0.385 and 0.478). 140 epoch shows less loss rather in the other steps. Hense 140 epochs can be considered as the idle epoch.

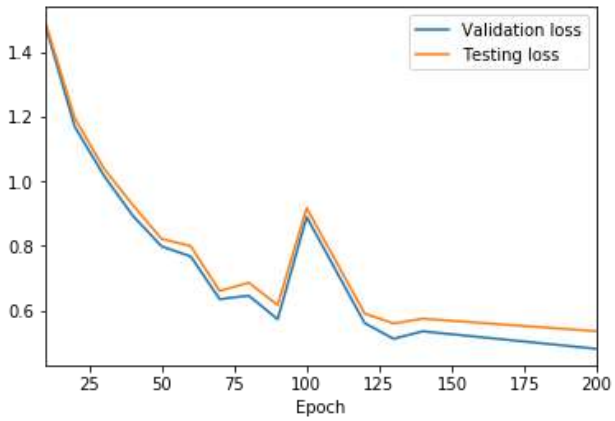

Figure 2. The loss for different epochs in MLP

Relationship between the loss and accuracy of the model is illustrated in Figure 3. The loss function is generally used for optimization purpose while the accuracy used to measure the performance of the algorithm. At the 140 epoch accuracy rises to the 0.81, while the loss drops to the 0.57. So it can be considered as the time both the loss and accuracy performs effectively according to characteristics of the curve.

When minimizing the training data set, the performance of the model is also reducing. Both validation and the testing accuracy are considerable high when the data set was split into an 80:20 ratio.

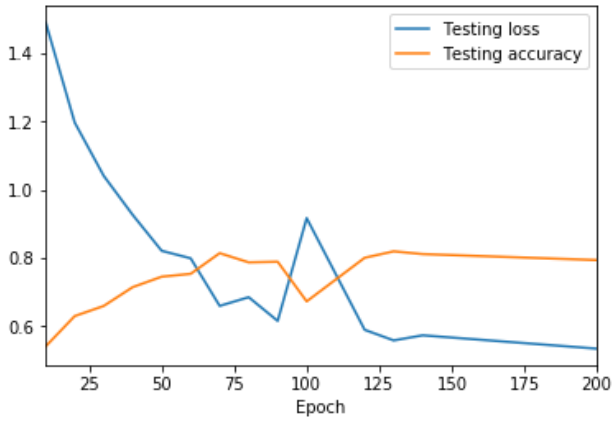

Figure 3. Accuracy and the loss performance of the MLP for 200 epochs

K fold validation is the validation technique used by the model. 5 folds were validated with the model separately. In every epoch performed, the accuracy of the 5 folds became differ. So the folds accuracy with the epoch is shown in figure 4. Fluctuation of the accuracy is seen within the graph. But most of the time, training accuracy improved at the final fold. (fold 4). All the other folds accuracy was improved rather than at fold 0. At the epoch 200, all the folds have improved.

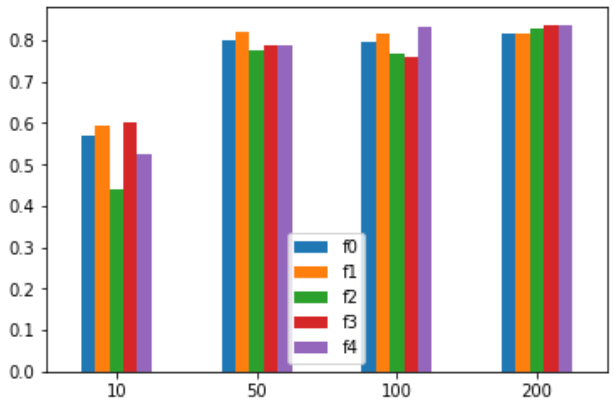

Figure 4. K fold results of the MLP

2) Sensitivity Analysis Using Different Optimization Functions: Neural networks need to be optimized to reduce the loss which optimizing the accuracy. For that purpose several optimization algorithms are in the practical usage as well as in the research domain.(Sameen and Pradhan, 2017) Adagard, SGD, RMSprop and Adam are most common among them. Weight and biases are adjusted by training by the mentioned Optimization functions. Each algorithm has its own advantage and disadvantage, selecting the optimization technique is one of the key function in the neural network. There is no exact optimization to use at particular problems. In order to find the best optimization algorithms to use, a study has evaluated 4 different optimization algorithms as shown in Table 1.

Table 1. Performance of the different optimization functions

Optimize function	parameters	Validation accuracy	Testing accuracy
SGD	Lr=0.01, decay=0.0, momentum=0.0	0.771679	0.75438
RMSprop	Lr=0.001, decay=0.0, momentum=0.0	0.791438	0.79122
Adagard	Lr=0.01, decay=0.0, momentum=0.0	0.762897	0.74824
Adam	Lr=0.001, decay=0.0, momentum=0.0	0.798024	0.78596

Algorithms have shown with their own characteristics and the training and testing accuracy for each optimization function is in the table. According to the evaluation results, the best accuracy reached for the algorithm is RMSprop with 0.7912280 accuracies. Which have a learning rate of 0.001. The lowest performance reached with Adagard and SGD. but other functions have performed considerable well rather than Adagard and SGD.

3) Sensitivity Analysis of the Different Learning Rates: The effect of the learning rate with the accident location prediction model was evaluated by using several learning rates to train the model. 0.5, 0.1, 0.05, 0.01 and 0.001 learning rates were used to evaluate the model with validation accuracy. Figure 5 illustrates the learning rate with validation accuracy. Learning rate 0.001 was the highest validation accuracy which archived by the model. Results revealed that by reducing larger learning rates does not maximize the validation accuracy. The selected model validation accuracy is 85.52% and testing accuracy was 81.18% which has 140 epochs, 0.003 learning rate and the optimizer as the RMSprop.

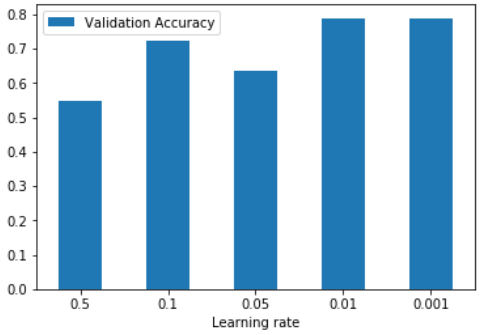

Figure 5. Accuracy for the different learning rates

B. *Comparative Analysis*

In the comparative analysis proposed MLP model was compared with the decision tree. Decision trees are known as one of the best and flexible classification model in machine learning. Developed decision tree achieved accuracy with 80.96%. The results achieved through the cross-matrix is shown in Table 2.

Table 2. Confusion matrix summary of the Decision Tree

	Precision	Recall	F1 - score	support
1	0.50	0.02	0.05	41
2	0.61	0.97	0.75	119
3	0.90	0.72	0.80	129
4	0.98	0.95	0.96	140
5	0.99	0.99	0.99	155
6	0.91	0.94	0.93	105
7	0.95	0.90	0.92	89
8	1.0	1.0	1.0	85
9	1.0	0.96	0.98	52
10	0.95	1.0	0.97	35

Then the MLP and decision tree results were compared with the correctly classified instances and incorrectly once. Comparison results are shown in Table 3.

Table 3. Comparison between MLP and decision tree

Approach	Correctly classified instances	Incorrectly classified instances
MLP	85.52%	14.48%
Decision tree	80.96%	19.04%

Results revealed that the MLP outperformed the decision tree with the highest validation accuracy.

IV. DISCUSSION & CONCLUSION

The proposed MLP model is a promising accident prediction model that can be applied in the practice. Most of the times when an accident occurred driver who is responsible for the driver tend to run away from the incident. In order to identify the age limit of the driver by using several accidents factor is a great accomplishment. Police authorities and other relevant parties may use the model to identify the exact age group of the driver. Therefore necessary legal actions can be taken.

In order to accurately predict the age limit using accident factors Multilayer perceptron model using Tensorflow has used with two hidden layers which achieved 85.52% accuracy. Direction of vehicle moving, a human crash factor related to the accident and the location of the accident (nearest km post) were used as the attributes for the model. After a number of sensitivity analyzing the best model for the prediction was selected. The critical hyper parameters which directly affect the model performance were evaluated. From the evaluation steps it was found that the most suitable epoch for the model is 140 steps and the most optimized function was the RMSprop. When critically analyzing the learning rate, learning rate with a value of 0.001 was obtained with analysis. And the used dropout function was helped to increase the accuracy of the model. K fold validation was used to validate the model while increasing the accuracy of each fold. Most of the time validation accuracy of the model was higher than the testing accuracy. When compared with the Decision tree, MLP model was outperformed with 85.52% accuracy.

The proposed model was accurately predicted the driver age which can be used for the investigation steps in the accident. Further studies must be a focus on identifying driver characteristics by using road accident factors. In addition to that theoretical studies are encouraged to perform with regarding maximizing the MLP accuracy.

ACKNOWLEDGEMENT

Authors would like to appreciate the support of the staff members of Traffic Branch of the Traffic Police Head Quarters Colombo and road safety division for providing the necessary data for the study.

REFERENCES

Gardner, M. W. and Dorling, S. R. (1998) 'Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences', *Atmospheric Environment*, 32(14–15), pp. 2627–2636. doi: 10.1016/S1352-2310(97)00447-0.

Lee, K. Y., Chung, N. and Hwang, S. (2016) 'Application of an artificial neural network (ANN) model for predicting mosquito abundances in urban areas', *Ecological Informatics*. The Authors, 36, pp. 172–180. doi: 10.1016/j.ecoinf.2015.08.011.

M.~Chong, A.~Abraham and M.~Paprzycki (2005) 'Traffic Accident Data Mining Using Machine Learning Paradigms', *Informatica*, Volume 29(June 2014), pp. 89–98. doi: 10.5815/ijitcs.2014.02.03.

Renze, J. (2019) 'Outlier', *MathWorld*. Available at: http://mathworld.wolfram.com/Outlier.html (Accessed: 15 June 2019).

Sunday times Sri Lanka (2017) *Road deaths in Sri Lanka are as natural as they are tragic | The Sunday Times Sri Lanka*. Available at: http://www.sundaytimes.lk/170910/news/road-deaths-in-sri-lanka-are-as-natural-as-they-are-tragic-258651.html (Accessed: 28 March 2019).

The Mathematics of Decision Trees, Random Forest and Feature Importance in Scikit-learn and Spark (2018). Available at: https://medium.com/@srnghn/the-mathematics-of-decision-trees-random-forest-and-feature-importance-in-scikit-learn-and-spark-f2861df67e3 (Accessed: 12 June 2019).

Traffic Police - Road Traffic Accidents (2018). Available at: https://www.police.lk/index.php/traffic-police/56 (Accessed: 14 October 2018).

Vogt, M. (2019) 'An Overview of Deep Learning and Its Applications', (January), pp. 178–202. doi: 10.1007/978-3-658-23751-6_17.