Journal of Sustainability Science and Management
Volume 18 Number 5, May 2023: 92-109

eISSN: 2672-7226

MODELLING ROAD ACCIDENTS IN SELANGOR: A COMPARATIVE ANALYSIS BY USING SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (SARIMA) AND ARTIFICIAL NEURAL NETWORK (ANN)

WAN NUR AFRINA WAN MUHAMMAD AZAN, NURBAIZURA BORHAN* AND SITI MERIAM ZAHARI

Statistical and Decision Science Studies, College of Computing, Informatics and Media, Universiti Teknologi MARA, 40450 Selangor, Malaysia.

**Corresponding author: nurbaizura@tmsk.uitm.edu.my*
Submitted final draft: 12 February 2023 *Accepted: 23 March 2023* http://doi.org/10.46754/jssm.2023.05.007

Abstract: Road accidents have become one of the major problems globally. It was reported that the total number of road accidents in Malaysia has increased by 0.92% within the last ten years. Modelling road accident occurrences is critical for policymakers to understand the trend and pattern of road accidents to provide an appropriate countermeasure. This study attempts to forecast the road accident occurrences on federal and state roads in Selangor, Malaysia. This study utilised monthly road accident data from January 2011 to December 2021. The traditional univariate Seasonal Autoregressive Integrated Moving Average (SARIMA) and Artificial Neural Network (ANN) models were employed and the performance of each model was assessed. The findings found that the ANN model outperformed the SARIMA model in training and validation. Furthermore, the ANN model has the lowest RMSE, MAE, MAPE and MASE values for both sets. This study demonstrates the potential of machine learning in forecasting and predicting road accident occurrences, giving more flexibility and assumption-free methodology.

Keywords: Time series, machine learning, road accidents, forecasting, dynamic impacts.

Introduction

Road accidents have become one of the major problems globally today. Due to rapid globalisation with the increasing population and vehicles, road accidents have increased globally over the years. According to the World Health Organization reports in 2021, the total number of deaths due to road accidents has increased slightly, reaching 1.3 million in 2019 particularly among vulnerable road users such as pedestrians, cyclists, and motorcyclists (World Health Organization, 2021). According to the Ministry of Transport Malaysia (MOT), the total number of road accidents has increased by 0.92%, amounting to 414,421 and 418,237 mishaps in 2010 and 2020, respectively. The number of road accident occurrences increased by 36.64% from 2010 to 2019. However, as expected, road accident occurrences decreased in 2020 and 2021. This could be due to the recent pandemic, which caused the number of road accident occurrences to decline.

According to the report of MOT in 2020, the total number of road accidents reported in Selangor in 2019 indicates the highest number of road accidents compared to other states in Malaysia. Moreover, Selangor has also recorded an increasing number of road accidents by 45.56% from 2010 to 2019. However, in 2020, there was a huge decline in the number of road accident occurrences by 90.39% and the impact of the COVID-19 outbreak could explain this.

To reduce the number of road accident occurrences in Selangor, the government has created enormous road safety programmes in the past. Despite all the measures, road accident occurrences and fatalities are still reported to be higher compared to other states. It is clear that modelling road accident occurrences is crucial and has become an important topic for researchers worldwide. The importance of modelling road accidents can provide better insights to the policymakers on the patterns and the authorities can take corrective action (Junus & Ismail, 2014; Zamzuri *et al*., 2019).

Therefore, this research aims to forecast and predict the future occurrence of road accidents on federal and state roads in Selangor by using Seasonal Autoregressive Integrated Moving Averages (SARIMA) and Artificial Neural Networks (ANN) with past time series data collected as a comparison. The time series models will be compared to determine the best-fitted model to predict road accidents in Selangor on federal and state roads.

This paper presented the literature from prior works in Section 2, followed by a discussion of SARIMA and ANN time series modelling in Section 3. The empirical results of both the SARIMA and ANN method modelling are presented in Section 4, and Section 5 concludes.

Literature Review

Prediction and modelling analysis has been a concern to all experts and practitioners worldwide in the context of various issues worldwide, among others (Suhaimi *et al*., 2019, Nashwan *et al*., 2019; Maulana *et al*., 2019; Tan *et al*., 2022). Over the decades, various methods have been utilised by past researchers to forecast road accidents occurrences worldwide (Borucka, 2018; Gazder *et al*., 2018; Ihueze & Onwurah, 2018; Zamzuri *et al*., 2019; Gutierrez-Osorio & Pedraza, 2020; Radzuan *et al*., 2020; Merabet & Zeghdoudi, 2020; Gilani *et al*., 2021). One frequently employed technique past researchers used is time series analysis, defined as examining a series of data points gathered over time. Time series forecasters gather and analyse historical data to build a model that captures the underlying data generation process. The model is then extrapolated to predict future values (Zhang, 2012).

The Autoregressive Integrated Moving Average (ARIMA) of Box and Jenkins and Seasonal Autoregressive Integrated Moving Average (SARIMA) models were commonly used in these past studies. Generally, ARIMA is a model used to forecast future points in a time series by fitting it to the data, while SARIMA is a method for analysing time series with seasonality. Over the years, various studies have been conducted to forecast and predict road accidents occurrences by using ARIMA (Ghedira *et al*., 2018; Ihueze & Onwurah, 2018; Nasir *et al*., 2018; Wai *et al*., 2019; Hassouna & Al-Sahili, 2020) and SARIMA (Zolala *et al*., 2016; Borucka, 2018; Merabet & Zeghdoudi, 2020). Linear or classical time series analysis of SARIMA is the traditional analysis used to model road accident occurrences for regions or countries and project future road accident patterns. The SARIMA model assumes that the mean for a certain season does not remain constant over time. Seasonality is a regular pattern of changes in a time series that repeats over S periods, where S is the number of periods between repetitions.

Zolala *et al*. (2016) have examined the trend for mortality of road accidents in Kermanshah province by using ARIMA and SARIMA models. The study found that SARIMA$(0, 0, 0)(1,1,1)_{[12]}$ has the lowest value of mean square error (MSE) and concluded that it is the best-fitted model to predict road accident fatalities. Borucka (2018) has employed the SARIMA model to analyse the risk of accidents on Polish roads, Poland. SARIMA$(6,1,2)(2,1,1)_{[12]}$ model with seasonal delay $D = 7$ was found to be the best model with only a 30% of forecast error. In the Skikda region, northeast Algeria, Merabet and Zeghdoudi (2020) conducted a study investigating the number of road accident occurrences. The study has reported that SARIMA$(12,1,1)(0,1,1)_{[12]}$ is chosen as the best-fitted model to forecast road accidents occurrences in the Skikda region, which the model has the smallest criterion values as compared to other SARIMA models. Additionally, SARIMA with the Kalman filter model performs better than the traditional SARIMA model with lower MAE, MAPE and RMSE values.

On the other hand, since many real-time series data appear to follow a non-linear motion, the SARIMA model has been found inadequate to represent their dynamics. It is critical to recognise the shortcomings of linear models. They cannot capture nonlinear data relationships (Zivot & Wang, 2006; Zhang, 2012). In the subject of forecasting road accident occurrences, machine learning approaches have been widely used owing to their capacity to analyse multi-dimensional data, flexible implementation and coding, and excellent predictive skills (Zheng *et al*., 2019). Several machine learning techniques have been employed by past researchers, such as Artificial Neural Network (ANN), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF) and Naïve Bayes (NB) (Çodur & Tortum, 2015; Suganya & Vijayarani, 2017; Lokala *et al*., 2017; AlMamlook *et al*., 2019; Lee *et al*., 2019; Radzuan *et al*., 2020; Silva *et al*., 2020; Kim *et al*., 2022; Bokaba *et al*., 2022), among others. Over the years, various studies have been conducted to forecast and predict road accident occurrences using the ANN model (Alkheder *et al*., 2017; García de Soto *et al*., 2018; Radzuan *et al*., 2020).

Alkheder *et al*. (2017) conducted a study utilising ANN to predict the severity of injuries in road accidents. The study employed an ordered probit model compared to ANN and found that the ANN outperformed the ordered probit model by yielding greater accuracy. In Switzerland, a study was conducted by García de Soto *et al*. (2018) to develop a road accidents risk prediction model. The study reported that ANN is the best-fitted model to forecast the frequency of road accident occurrences compared to the Bayesian Network (BN) model, which has a lower tendency to overestimate the number of road accident occurrences. Moreover, in Malaysia, Radzuan *et al*. (2020) have employed ANN to predict the serious injury cases involved in road accidents and the study found that the developed ANN with seven key features is capable of forecasting the frequency of serious injuries due to road accidents with the lowest value of MAPE and a high degree of accuracy.

Moreover, several past studies have also employed both traditional SARIMA and machine learning techniques to determine the comparison and accuracy of both analyses in forecasting road accident occurrences. Studies that have employed SARIMA modelling with machine learning techniques in forecasting road accident occurrence include, among others, (Qian *et al*., 2020; Slimani *et al*., 2020; Feng *et al*., 2022). In China, Qian *et al.* (2020) conducted a study to estimate road traffic fatalities and to forecast short-term road accident deaths using the traditional SARIMA and the Elman Recurrent Neural Network (ERNN) model. The study found that the proposed and developed ERNN model outperformed the traditional SARIMA model in predicting road accident deaths in 2017 and has the lowest MAPE value compared to the SARIMA model.

Moreover, in Morocco, Slimani *et al*. (2020) conducted a study that utilised the Multilayer Perceptron of ANN (MLP-ANN) to forecast the daily traffic stream on Morocco's highway. The study reported that the proposed MLP-ANN model outperformed the traditional SARIMA and the Support Machine Regression (SMOreg), which yielded the smallest error forecast of RMSE, MAE and MSE as compared to other models. Finally, in the latest year, a study which employed the Long Short-Term Memory (LSTM) model to predict road traffic injury was conducted by Feng *et al*. (2022) in Northeast China. The study also utilised SARIMA and Facebook Prophet (Prophet) models and the performances of each model were compared and evaluated. The study found that the LSTM model had the highest prediction accuracy by predicting the number of road traffic injuries more accurately and robustly than other models.

In general, there is still a lack of studies on predicting and forecasting the occurrences of Malaysian road accidents, especially on federal and state roads. In order to fill this gap, this study is conducted to forecast road accident occurrences on federal and state roads in Selangor by employing two-time series models, the traditional SARIMA and ANN models, for comparative analysis to obtain the model with the most accurate prediction effect. Then, the performance of both models was evaluated and a future forecast one step ahead was conducted to forecast the value of road accidents on federal and state roads in Selangor for 2022.

Methodology

SARIMA Time Series Modelling

This section briefly describes the time series model employed in this study. An ARIMA model has three order parameters as follows: ARIMA (*p, d, q*) whereby *p* refers to the order of the ARIMA model's nonseasonal autoregressive process, *d* is the order of nonseasonal integration or differencing while q refers to the order of the nonseasonal moving average method used in the model. Autoregression (AR) which denotes with parameter p defined as the dependent relationship between an observation and a set of lagged observations used in the model or the incorporation of previous values in the regression equation for the series *Y*. Moving Average (MA) can be defined as the dependence between an observation and a residual error from a moving average model applied to lagged data is used in this model. The number of terms included in the model is determined by the order *q*. The ARMA models are a general and useful class of time series models that combine autoregressive [AR(*p*)] and moving average [MA(*q*)] models to generate a general and useful class of time series models. The ARMA (*p, q*) model:

$$Y_t - \sum_{i=1}^{p} \phi_i Y_{t-i} = \varepsilon_t - \sum_{j=1}^{q} \theta_i \varepsilon_{t-i} \quad (1)$$

$$\phi_p(B) Y_t = \theta_q(B) \varepsilon_t$$

where $\{\varepsilon_t\}$ is random errors which are identically distributed with a mean of zero and a constant variance, σ^2, $\phi_p(B)$ and $\theta_q(B)$ are known as polynomials of backward shift operators such that $B^m Y_t = Y_{t-m}$.

Next, integrated or differencing uses raw observation differencing to make the time series stationary. In the integrated component, the *d* denotes the degree of differencing. A series is differentiated by subtracting the current and previous values *d* times. When the stationarity assumption is violated, differencing is frequently utilised to stabilise the series. If *d* is the order of non-seasonal differencing such that $(1 - B)^d Y_t$, it is stationary and follows an ARMA process, which is said to be an autoregressive integrated moving average process ARIMA.

A seasonal structure can also be used to specify ARIMA models. The model is defined by two sets of order parameters *(p, d, q)* and *(P, D, Q)* representing the seasonal component of m periods. Although selecting *(P, D, Q)* is similar to picking non-seasonal order parameters, seasonal ARIMA necessitates a more extensive model structure design. In this model, *P* is the order of the seasonal autoregressive process, *D* is the order of seasonal integration or differencing and *Q* is the order of the seasonal moving average process. SARIMA (*p, d, q*) x (*P, D, Q*)s model can be written as:

$$\Phi_P (B^s)\, \phi_p(B)\, (1 - B^s)^D (1 - B)^d\, Y_t \quad (2)$$

$$= \Theta_Q (B^s)\, \theta_q(B)\, \varepsilon_t$$

where *s* is the seasonal period, *D* is the order of seasonal differencing, ϕ and θ are the conventional autoregressive and moving average parameters, while ϕ and θ are the seasonal autoregressive and moving average parameters.

The degree of differencing for non-seasonal, *d* and seasonal, *D* non-stationary processes is either 1 or 2 in many scenarios. Two statistical characteristics, Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF), serve a similar role in the design of SARIMA models as they do in the construction of non-seasonal ARIMA models.

The model identification for *p*, *q*, *P*, *Q* can be made using ACT and PACF plots. The series must be stationary to fit the SARIMA model. When the mean, variance, and autocovariance of a series are time-invariant, it is said to be stationary. Additionally, three types of unit root tests which are the Augmented Dickey-Fuller (ADF), Phillips–Perron (PP) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) are employed to identify the model's stationarity. As for the ADF and PP tests, the null hypothesis is rejected if the t-value is larger than the critical value of ADF test or the p-value is less than α=0.05, which indicates that there is no unit root in the series and data is stationary. However, for the KPSS test, the decision rule of the test is the opposite of both ADF and KPSS tests, where the null hypothesis is accepted if the test statistic value is smaller than the critical value of KPSS test or the p-value is greater than α=0.05, which indicates that the process is stationary. If the series is not stationary, it should be changed to become stationary, commonly done via differentiation. Next, Maximum Likelihood Estimation (MLE) can be used to estimate the parameters and their standard errors.

Then, to choose the best model with statistical noise without any regular pattern, model diagnostic checking is employed by plotting the residuals over time to identify the absence of systematic trends, representing the autocorrelation function of residuals. Moreover, the diagnostic checking of the developed models consists of the parameter (coefficients) and residual test, which utilised the Ljung-Box Portmanteau test (LB) and the plots of ACF and PACF. After the model has been identified and the model parameters have been estimated, a model selection criterion is required to select the best-fitting model. Two statistical indicators will be used to evaluate the goodness of fit of the model, which are Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC). The AIC and BIC statistics are calculated as follows:

$$AIC = -2 \ln L + 2k \quad (3)$$

$$BIC = -2 \ln L + kln(n) \quad (4)$$

where k is the number of parameters in the estimated model, L is the maximised value of the likelihood function of the estimated model and n is the number of observations in the data series. The final or best-fitted model of SARIMA is chosen based on a model that executed the smallest value of AIC and BIC, suggesting that the smaller the value, the better the model is. In addition, the goodness of fit of the model could also be evaluated by using the absence of autocorrelation and partial autocorrelation in the residuals.

Artificial Neural Networks (ANNs) Method

Artificial Neural Networks (ANNs) are forecasting techniques based on simple brain mathematical models. By using ANNs, complex nonlinear connections between the response variable and its predictors can be achieved. The neural network becomes non-linear when we add an intermediary layer with hidden neurons. The widely used design characterised the network comprising three layers of basic processing units and each layer of neurons is connected cyclically. Figure 1 shows an example of a commonly used ANNs design known as a multilayer feed-forward network, where each layer of nodes receives inputs from the previous layers. The outputs of one layer's nodes are used as inputs in the following layer. A weighted linear combination is used to combine the inputs to each node. Before being output, the result is changed by a nonlinear function.

The model can be expressed as follows:

$$y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j f\,(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij}\, y_{t-i}) + \varepsilon_t \quad (5)$$

where p is the number of input nodes, q is the number of hidden notes, f is the sigmoid transfer function such as the logistic: $f(x) = \frac{1}{1+\exp(-x)}$. $\{\alpha_j, j = 0,1,2,\ldots,q\}$ is a vector of weights from the hidden to output nodes and $\{\beta_{ij}, i = 0,1,2,\ldots,p; j = 1,2,$ are weights from the input to hidden nodes. α_0 and β_0 are the biased terms which have values always equal to 1.

According to Halim *et al*. (2016), modelling ANNs can be divided into three main phases: Modelling, training and testing. The modelling phase consists of rules, input parameters, and data gathering. In this phase, the independent and dependent variables will be defined as input(s) and output, respectively. Next, data preparation and learning law adaptation are carried out in the training phase. Here, the most appropriate type of ANNs and their parameters will be chosen to solve the study's problem. In addition, choosing the suitable ANNs architecture by adding the number of layers and neurons that will integrate the model into the design will be done in this phase. The Neural Network Time

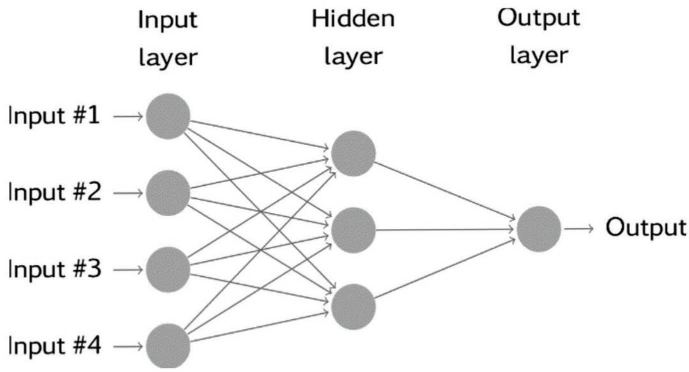

Figure 1: Artificial neural networks (ANNs) design

Series (NNAR) is used in this study to obtain the appropriate ANN model for the dataset. NNAR is one of the methods of ANN in which lagged time series values were used as inputs to a neural network. The *nnetar* function is used and the notation is NNAR(p, P, k)$_m$ to indicate p lagged inputs, P seasonal lagged inputs, k nodes of hidden layer and m seasonal period. In general, a NNAR(p, P, 0)$_m$ is equivalent to an ARIMA(p, 0, 0)(P, 0, 0)$_m$ model but without the parameter limitations that guarantee stationarity. The NNAR(p, P, k)$_m$ model can be expressed as follows:

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}, y_{t-m}, y_{t-2m}, \dots, y_{t-Pm}) + \varepsilon_t \quad (6)$$

where f indicates the neural network model with k hidden nodes in a single layer and is the residuals series.

Finally, in the testing phase, accuracy and performance are assessed by computing the difference between expected and actual outputs. The performance measures of ANNs model will utilise two accuracy measures: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) values. In addition, the Ljung-Box Portmanteau test (LB) will be conducted to check on the residuals of the developed model.

Observations separated by multiples of seasonal period m are frequently associated with seasonal autocorrelations in a seasonal time series. Monthly time series, for example, are likely to be associated with observations made 12 months apart. That is, a particular observation is linked to previous and future observations. If explicitly modelling seasonality is desired, it is necessary to include these seasonal observations as viable input variables in neural network models.

Results and Discussion

Description of Data

In this study, data used is the historical road accidents on federal and state roads in Selangor obtained from Royal Police Malaysia (PDRM). The data comprised monthly road accidents from January 2011 to December 2021 or 132 observations. The data will be partitioned into two categories; training and validation. The training dataset consists of January 2011 to December 2018 (96 months) while January 2019 and December 2021 (36 months) are treated as validation datasets. The analysis of both SARIMA and ANN models was carried out using R Software.

Preliminary Data Analysis

Trend analysis will be conducted after pre-processing the data to check whether there is a trend, seasonal component, apparent sharp changes in behaviour, and outlying observation in the dataset (Brockwell & Davis, 2002). Trend analysis is the process of examining data for patterns in its graphical representation. The trend component in business or economic time series data, according to Lazim (2011), describes the overall rising or downward motions that define all economic and business activities often observed in dynamic economic and business contexts. Figure 2 displays the plot of the monthly time series of road accident occurrences data. This time series has considerable seasonal variation in the number of road accidents per month. Additionally, an additive model is appropriate to describe this as the seasonal fluctuations appear approximately consistent in size across time and do not appear to be dependent on the time series level.

Result of SARIMA Model

According to the Box and Jenkins methodology, time series modelling entails transforming data to achieve stationarity, identifying appropriate models, estimating model parameters, diagnostic checking of the assumption model, and forecasting future data values.

The SARIMA model's result was obtained first to check for the series stationarity condition. The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) were plotted to identify the specification of the SARIMA model. Figure 3 shows the ACF and PACF plots. Based on the ACF plot, the changes vectors of autocorrelation (ACF) in original time series data decrease as time increases, and some even meet at the negative values. It indicates that there is a presence of correlation in the series. Meanwhile, based on the PACF plot, it cuts off at lag 1, and a few exceed the band.

The ADF, PP and KPSS tests were conducted to check the stationarity of the series data. Table 1 shows the results of ADF, PP and KPSS tests. ADF and KPSS indicate non-stationarity of the series data because the p-value of ADF test is greater than 0.05 and the p-value of KPSS test is less than 0.05, while PP test indicates stationarity in the dataset with a p-value of KPSS test is less than 0.05. Since there are few values that exceed

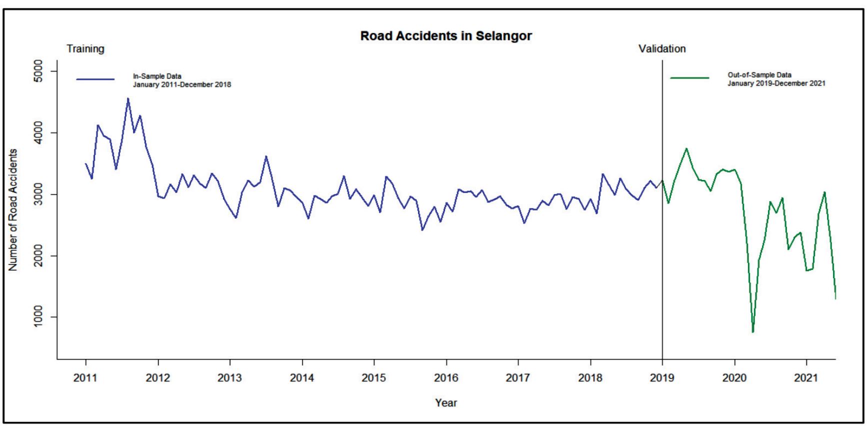

Figure 2: Time series plot of road accidents occurrences in Selangor

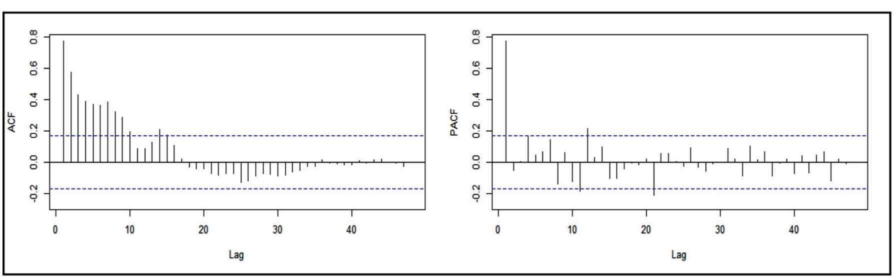

Figure 3: ACF and PACF plots of road accidents occurrences in Selangor

the significance limit in both ACF and PACF plots and based on the results of ADF and KPSS tests, there is a need for this series data to be made stationary by performing the seasonal and non-seasonal differences on the original series data.

Figure 4 shows the time series, ACF and PACF plots with different values of differencing. After performing the seasonal and non-seasonal differencing to the series data, the ACF and PACF plots showed few significant correlations at the beginning of the lags and some correlations exceeded the band and also showed an oscillating pattern around 0 with no strong visible trend. As for the differenced series data with order *D*=1 and *d*=1, the data hovers around the same mean of 0 except after 2020, when the series fluctuated. On the other hand, the differenced series data of order *D*=1 and *d*=2 shows that the data hovers around the same mean of 0 with fluctuation in the year 2021. Moreover, as for the differenced series data with order *D*=2 and *d*=1, there were not many changes in the series plots and ACF, while PACF showed a better result with only a few 4 significant correlations as compared to the first two differenced series data.

The unit root tests (ADF, PP and KPSS) for each seasonal-non-seasonal differenced series data were conducted to check for the stationarity of the series with differencing of order 1 and order 2 as shown in Table 2. Based on the p-value of ADF, PP and KPSS tests for each seasonal-non seasonal differenced, stationarity is assumed as the rejecting null hypothesis since the p-value is less than 0.05 for ADF and PP tests while accepting the null hypothesis as the p-value is more than 0.05 for KPSS test. All unit root tests indicate that all three seasonal-non-seasonal differenced series are stationary

Table 1: ADF, PP and KPSS unit root tests

Unit Root Tests	p-value
ADF Test	0.2339
PP Test	0.01
KPSS Test	0.01

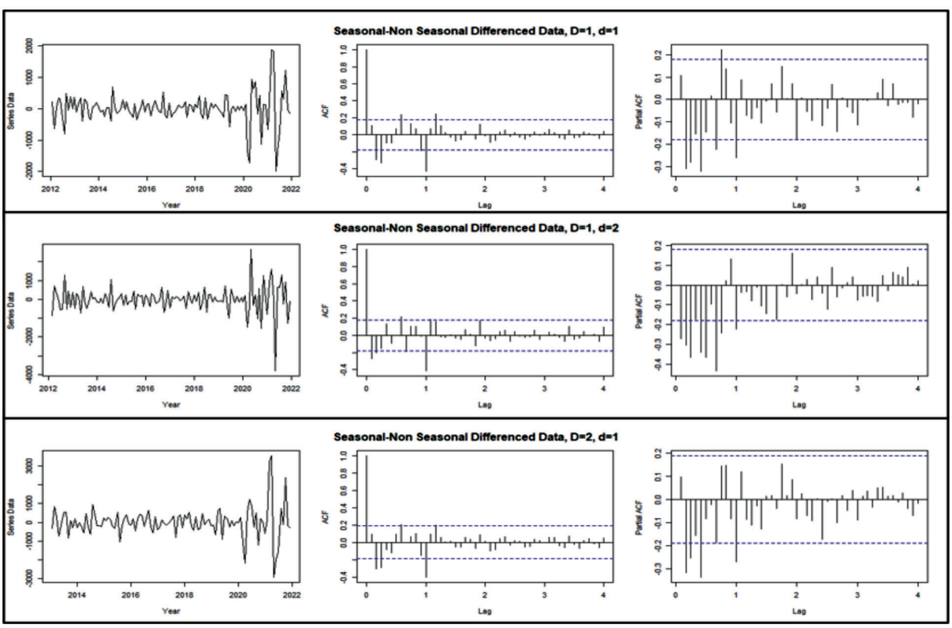

Figure 4: Plots of ACF, PACF and different series data

when differencing of order D=1 with d=1,2 and D=2 with d=1 was applied; hence the assumption of stationarity is met. This suggests that differencing order 1 and 2 terms for both seasonal and non-seasonal terms is sufficient and should be included in the model.

After the model has been identified, a few different SARIMA models have been fitted to the series data and the results of the best model results are chosen based on the model with the lowest AIC value. Each model parameter estimation is presented in Table 3. Based on Table 3, SARIMA(1,1,1)(1,2,1)$_{[12]}$ is the model with the lowest value of AIC as compared to other SARIMA models. Moreover, SARIMA(1,1,1)(1,2,1)$_{[12]}$ is the only model that exhibits

Table 2: Unit root tests on each seasonal-non-seasonal differenced series data

Seasonal-Non Seasonal Order	***D*=1, *d*=1**	***D*=1, *d*=2**	***D*=2, *d*=1**
ADF Test	0.01	0.01	0.01
PP Test	0.01	0.01	0.01
KPSS Test	0.1	0.1	0.1

Table 3: Empirical results of SARIMA models

Model	**Coefficient**		**Value**	**S.E**	**p-value**	**LB**	**AIC**
SARIMA(0,1,2)(1,2,1)$_{[12]}$	θ_1	MA(1)	-0.3674	0.1196	0.0021	24.4760 (0.0574)	1018.3692
	θ_2	MA(2)	-0.2247	0.139	0.1059		
	ϕ_1	SAR(1)	-0.3037	0.1379	0.0277		
	θ_1	SMA(1)	-0.9600	0.2668	0.0003		
SARIMA(1,1,1)(1,2,1)$_{[12]}$	ϕ_1	AR(1)	0.5321	0.1607	0.0009	22.3210 (0.0997)	1016.7743
	θ_1	MA(1)	-0.9022	0.1195	0.0000		
	ϕ_1	SAR(1)	-0.3320	0.1359	0.0146		
	θ_1	SMA(1)	-0.9646	0.2531	0.0001		
SARIMA(1,1,2)(1,2,1)$_{[12]}$	ϕ_1	AR(1)	0.6210	0.2963	0.0361	22.6620 (0.0660)	1018.6719
	θ_1	MA(1)	-1.0169	0.3692	0.0059		
	θ_2	MA(2)	0.0813	0.2494	0.7445		
	ϕ_1	SAR(1)	-0.3330	0.1361	0.0144		
	θ_1	SMA(1)	-0.9634	0.2543	0.0002		
SARIMA(2,1,1)(1,2,1)$_{[12]}$	ϕ_1	AR(1)	0.5320	0.1557	0.0006	22.6360 (0.0665)	1018.6800
	ϕ_2	AR(2)	0.0429	0.1399	0.7592		
	θ_1	MA(1)	-0.9244	0.1436	0.0000		
	ϕ_1	SAR(1)	-0.3346	0.1361	0.0140		
	θ_1	SMA(1)	-0.9642	0.2542	0.0001		
SARIMA(2,1,2)(1,2,1)$_{[12]}$	ϕ_1	AR(1)	-0.3825	0.1724	0.0265	20.0920 (0.0929)	1019.0633
	ϕ_2	AR(2)	0.4175	0.1745	0.0167		
	θ_1	MA(1)	0.1113	0.1601	0.4867		
	θ_2	MA(2)	-0.8799	0.1525	0.0000		
	ϕ_1	SAR(1)	-0.3098	0.1382	0.0250		
	θ_1	SMA(1)	-0.9598	0.2668	0.0003		

statistically significant of all coefficients in the model in which the p-value of the coefficients is less than α=0.05. Then, the diagnostic checking for each model's coefficients was tested by using the LB test. Based on LB test, it shows that the residuals of the chosen SARIMA model are assumed to be white noise and independently distributed from each other since the p-value is greater than 0.05. Hence, the best appropriate and chosen model to forecast the dataset is the SARIMA$(1,1,1)(1,2,1)_{[12]}$ with the value of AIC = 1016.7743. The corresponding ACF plot of residuals for SARIMA$(1,1,1)(1,2,1)_{[12]}$ is presented in Figure 5.

Results of ANN Model

P was set to 1 due to the seasonality in the dataset. Besides the automatic selection by *nnetar* function, multiple *p* values from 1 to 10 were evaluated and the comparison of each ANN model based on two accuracy measures and LB test is presented in Table 4. Based on Table 4, model NNAR$(1,1,2)_{[12]}$ performed well in the validation set as it has the lowest values of all four accuracy measures compared to other ANN models (RMSE=759.4985, MAE=575.1768, MAPE=34.7503 and MASE=2.1823. In contrast, model NNAR$(10,1,6)_{[12]}$ performed well in the training set with the lowest values of RMSE (14.1611), MAE (9.9932), MAPE (0.3380) and MASE (0.0379) compared to other models.

Furthermore, the diagnostic checking on the residuals was conducted. Both models of NNAR$(1,1,2)_{[12]}$ and NNAR$(10,1,6)_{[12]}$ show that there is no autocorrelation in the residuals in the models since the p-value is greater than 0.05 and the residuals are white noise. As for the ACF plot of the residuals, NNAR$(1,1,2)_{[12]}$ showed that there is one significant correlation at lag 6 (Figure 6) while NNAR$(10,1,6)_{[12]}$ model all correlation coefficients are within the confidence band (Figure 7). Considering two indices of accuracy measures both in the training and validation sets, both NNAR$(1,1,2)_{[12]}$ and NNAR$(10,1,6)_{[12]}$ models are selected as the optimal models and used for forecasting the road accidents occurrences at federal and state roads in Selangor.

Comparison of SARIMA and ANN Model

After the best appropriate models for both SARIMA and ANNs had been selected, the performances of these two models in the training and validation sets were evaluated in order to select the best and appropriate model to predict the road accidents occurrences at federal and state roads in Selangor and presented in Table 5. Based on the table, it shows that both ANN models outperform the traditional SARIMA. NNAR$(1,1,2)_{[12]}$ outperformed other models in the validation dataset, while NNAR$(10,1,6)_{[12]}$ outperforms others in the training dataset which in each dataset, both ANN models have

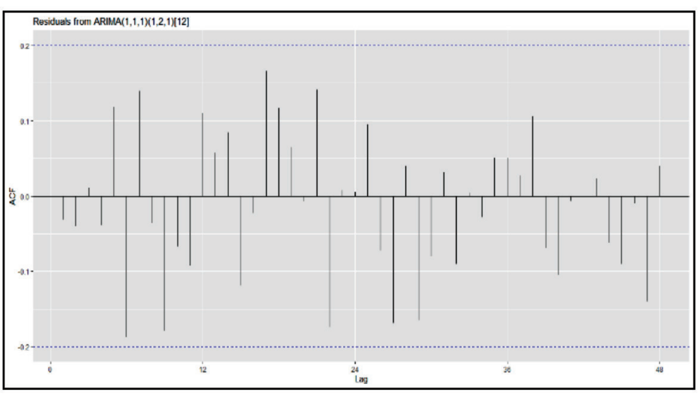

Figure 5: ACF plot of residuals of SARIMA$(1,1,1)(1,2,1)_{[12]}$

Table 4: Results of NNAR models

Model	Training Dataset				Validation Dataset				LB
	RMSE	MAE	MAPE	MASE	RMSE	MAE	MAPE	MASE	
NNAR(1,1,2)$_{[12]}$	161.2505	122.2941	4.1331	0.4640	759.4985	575.1768	34.7503	2.1823	0.1722 (0.6782)
NNAR(2,1,2)$_{[12]}$	149.1066	111.3938	3.7748	0.4227	774.6504	582.5278	35.4469	2.2102	0.3621 (0.5473)
NNAR(3,1,2)$_{[12]}$	148.7257	111.7374	3.7814	0.4240	848.8346	637.4994	38.8468	2.4188	1.0014 (0.3170)
NNAR(4,1,3)$_{[12]}$	108.8163	79.2007	2.6918	0.3005	925.7161	686.4193	42.1174	2.6044	1.4722 (0.2250)
NNAR(5,1,4)$_{[12]}$	89.5111	61.9811	2.1101	0.2352	826.2132	627.6511	37.7921	2.3814	1.1063 (0.2929)
NNAR(6,1,4)$_{[12]}$	71.3787	50.9413	1.7310	0.1933	805.9279	603.1425	36.5953	2.2884	1.8126 (0.1782)
NNAR(7,1,4)$_{[12]}$	56.0552	41.0172	1.3883	0.1556	825.1486	630.6027	37.7768	2.3926	0.0560 (0.8129)
NNAR(8,1,5)$_{[12]}$	45.4619	28.8988	0.9824	0.1096	879.2991	656.3539	40.0298	2.4903	0.3044 (0.5811)
NNAR(9,1,6$_{)[12]}$	19.0666	12.9787	0.4391	0.0492	904.5878	681.7063	41.1724	2.5865	0.6185 (0.4316)
NNAR(10,1,6)$_{[12]}$	14.1611	9.9932	0.3380	0.0379	780.9573	608.9640	35.4811	2.3105	0.0044 (0.9469)

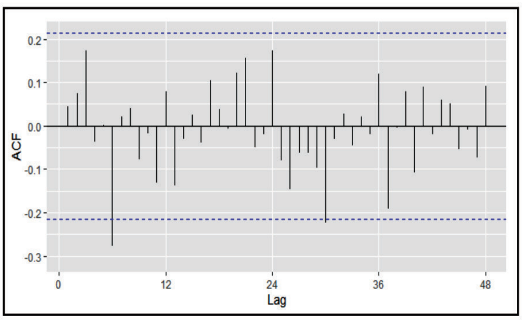

Figure 6: ACF plot of residuals of NNAR(1,1,2)$_{[12]}$.

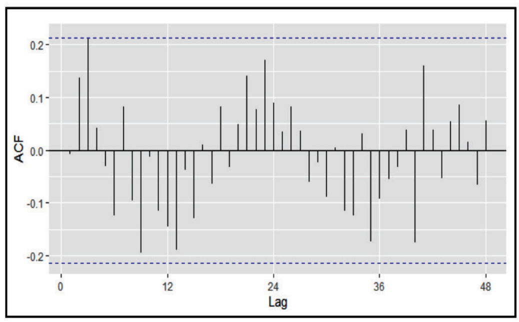

Figure 7: ACF plot of residuals of NNAR(10,1,6)$_{[12]}$.

the lowest values of RMSE, MAE, MAPE and MASE. The SARIMA model performs worst compared to both ANN models, which have the largest values of all four accuracy measures in both training and validation datasets.

Moreover, the in-the-sample and out-of-sample forecasts for the chosen SARIMA and both ANN models were carried out and shown in Figure 8 and Figure 9. As for the in-the-sample forecasts, NNAR(10, 1, 6)$_{[12]}$ fitted well while both SARIMA and NNAR(1,1,2)$_{[12]}$ fitted did not far from the actual series values. In contrast, as for the out-of-sample forecasts, the SARIMA model forecasted well in 2019. However, the forecasts (fitted values) were far higher than the actual series values from 2020 till the end of 2021. The actual series in 2020 went down due to the pandemic and it slowly rose back up and fluctuated till the end of 2021. For both ANN models, the models fitted far from the actual series data, with no increasing or decreasing trend. NNAR(1, 1, 2)$_{[12]}$ model

Table 5: Performances of SARIMA and NNAR models in training and validation set

Model	Training Dataset				Validation Dataset			
	RMSE	MAE	MAPE	MASE	RMSE	MAE	MAPE	MASE
SARIMA(1,1,1)(1,2,1)$_{[12]}$	192.701	127.386	4.309	0.586	1604.844	1238.032	71.656	4.697
NNAR(1,1,2)$_{[12]}$	161.250	122.294	4.133	0.464	759.499	575.177	34.750	2.182
NNAR(10,1,6)$_{[12]}$	14.161	9.993	0.338	0.038	780.957	608.964	35.481	2.311

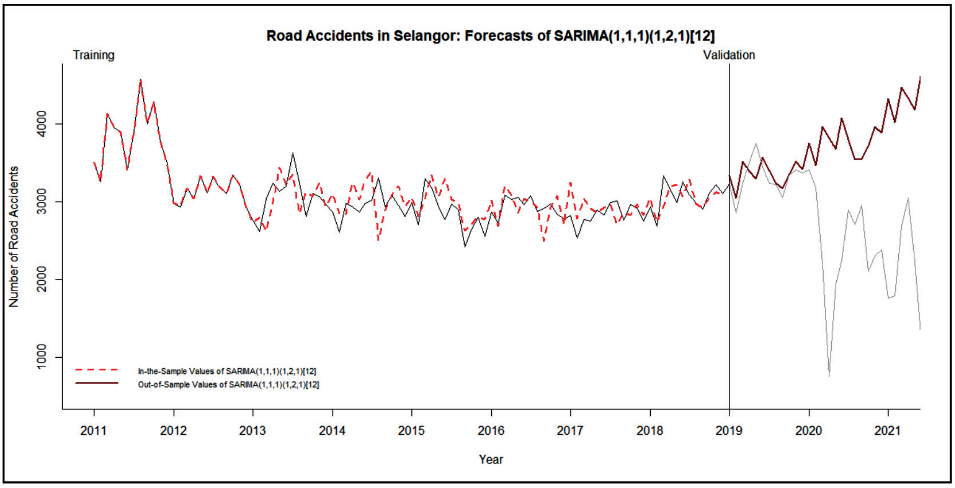

Figure 8: In-sample and out-of-sample forecasts of SARIMA(1,1,1)(1,2,1)$_{[12]}$

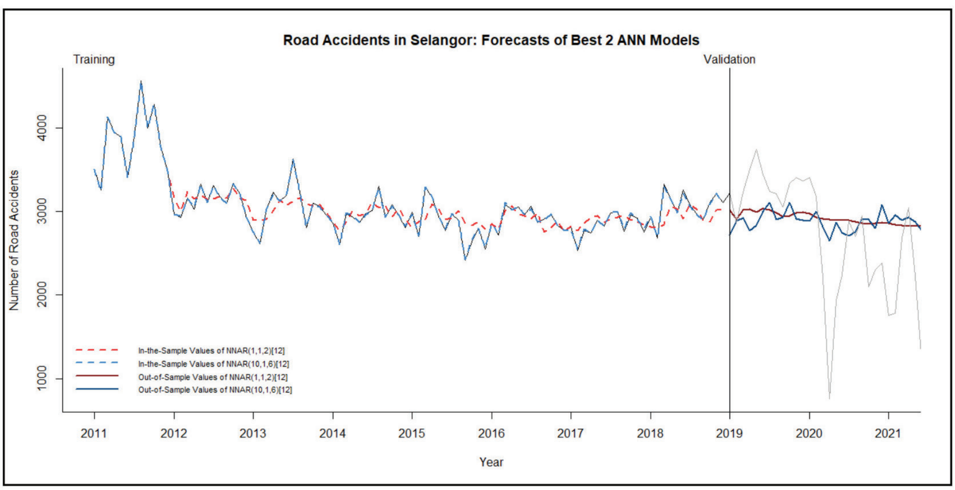

Figure 9: In-sample and out-of-sample forecasts of NNAR(1,1,2)$_{[12]}$ and NNAR(10,1,6)$_{[12]}$

has no fluctuations and the fitted values do not have distinct pattern while for NNAR(10, 1, 6)$_{[12]}$, there is patterns and similarly hovers around the same values.

Future forecasts of road accidents occurrences in Selangor

Finally, after modelled the data with several models and each model was evaluated, a future forecast one step ahead is conducted to forecast the value of road

Table 6: Future forecasts values in 2022

Month	Best Models		
	SARIMA(1,1,1)(1,2,1)$_{[12]}$	**NNAR(1,1,2)$_{[12]}$**	**NNAR(10,1,6)$_{[12]}$**
January 2022	4980.1878	2808.8393	2698.4038
February 2022	4666.8348	2802.2182	3165.1655
March 2022	5093.4355	2797.6320	3155.3891
April 2022	4966.1611	2795.8582	2813.3467
May 2022	4787.7684	2793.4748	3108.0196
June 2022	5372.3849	2794.3066	3017.5159
July 2022	4940.4611	2793.9066	3046.1795
August 2022	4522.4618	2792.0497	2848.1982
September 2022	4617.7829	2788.9792	2994.6565
October 2022	4771.8285	2787.9594	3087.3817
November 2022	5162.3514	2789.7467	2795.8382
December 2022	5132.5290	2790.1388	2736.3413

accidents on federal and state roads in Selangor for the year 2022. Table 6 shows the future forecast value of each model, while visualisation of the future forecasted values of each model is shown in Figure 10, Figure 11 and Figure 12, respectively.

Plot of future forecasts of SARIMA(1,1,1)(1,2,1)$_{[12]}$ model as shown in Figure 10. It shows an increasing trend and pattern in the future forecast values of the model in the year 2022. The number of future forecasts is higher than the existing past values. The future forecasts of the model start with approximately 4980 cases in January 2022, then peak in June 2022, and gradually decrease in August 2022, which has the lowest future forecast values. The number of future forecasts continues to rise until the end of December 2022. On average, there are approximately 4917 cases of road accident occurrences on federal and state roads in Selangor each month in 2022.

Next, for the first ANN model, NNAR(1,1,2)$_{[12]}$, the plot of future forecasts shows that there is no trend or pattern in the number of future forecasts of road accidents occurrences at federal and state roads in Selangor throughout the whole period of the year 2022 as shown in Figure 11. The future forecasts seem stagnant and hover around the same values. There is no increasing or decreasing pattern in the future forecast values of the model. The number of future forecasts starts with approximately 2808 cases in January 2022, then gradually decreases until December 2022. On average, there are approximately 2794 cases of road accident occurrences on federal and state roads in Selangor each month in 2022.

Then, for the second ANN model of NNAR(10,1,6)$_{[12]}$, the plot in Figure 12 shows slightly better future forecast values in 2022 than the first ANN model. The future forecast values hover around the same range and are similar to the range from 2012 to 2018. There is still no increasing or decreasing trend in the future forecast values of road accident occurrences on federal and state roads in Selangor between January 2022 and December 2022. The future forecasts of the model start with approximately 2698 cases in January 2022, then peak in February 2022, and fluctuate throughout March 20200 till December 2022. In general, in each month, on average, the model forecasted approximately 2955 cases.

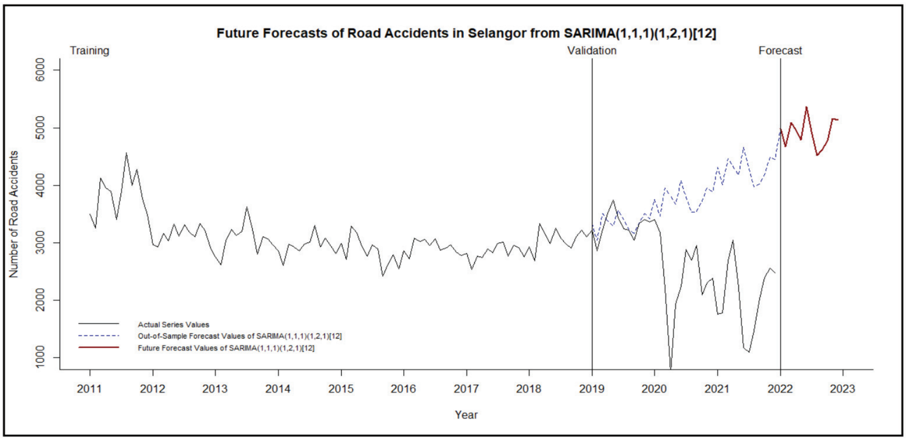

Figure 10: Future forecasts of SARIMA (1,1,1)(1,2,1)$_{[12]}$

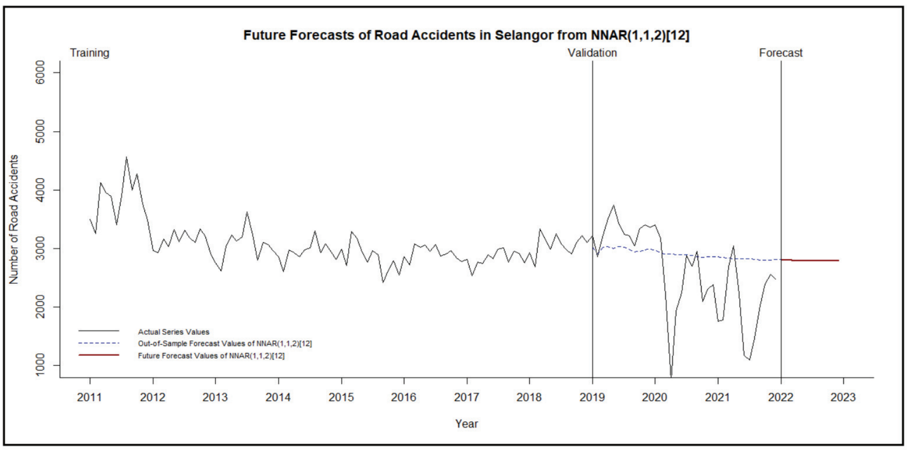

Figure 11: Future forecasts of NNAR(1,1,2)$_{[12]}$

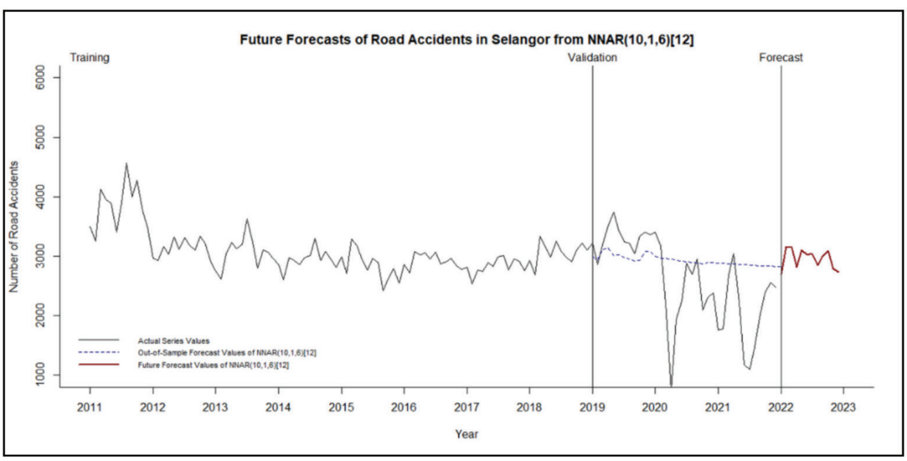

Figure 12: Future forecasts of NNAR(10,1,6)$_{[12]}$

Conclusion

Many sectors face the challenge of improving accuracy, especially time series forecasting. Despite the abundance of time series models accessible, research towards increasing forecasting model efficacy has never stopped. Besides the traditional univariate SARIMA time series model, Artificial Neural Networks (ANNs) are a powerful, general-purpose tool for pattern recognition, classification, grouping, and, most importantly, accurate prediction.

In this study, we model and forecast the occurrence of road accidents on federal and state roads in Selangor based on monthly data from January 2011 to December 2021. This study compares two time series models of SARIMA and ANN to estimate the performance. The study discovered that the ANN model outperformed in fitting or forecasting compared with the traditional univariate SARIMA model. There is no increasing or decreasing trend in the future forecasts made by both ANN models. The highest number of road accident occurrences on federal and state roads in Selangor would be in January 2022, while the lowest number on federal and state roads in Selangor would be in February 2022 or October 2022. Hence, the number of road accident occurrences on federal and state roads in Selangor is expected to be within the same range as in the last observed value in 2019 without increasing or decreasing the number of future forecasts.

This study demonstrates the potential of machine learning in forecasting and predicting road accident occurrences. These findings suggest that the government, policymakers, and other relevant authorities should collaborate to prevent future road accident occurrences in Selangor. For future work, expanding the dataset with more historical values is recommended. Furthermore, it may be possible to forecast the occurrences of road accidents in the future by including some random variables in the design parameters. Future research could focus on how to improve ANN models' prediction performance and in this instance, more advanced techniques, such as a deep neural network, could be explored.

Acknowledgements

A special appreciation to the College of Computing, Informatics and Media, Universiti Teknologi MARA (UiTM) for supporting the publication of this paper.

References

Alkheder, S., Taamneh, M., & Taamneh, S. (2017). Severity prediction of traffic accident using an artificial neural network. *Journal of Forecasting, 36*(1), 100-108. https://doi.org/10.1002/for.2425

AlMamlook, R. E., Kwayu, K. M., Alkasisbeh, M. R., & Frefer, A. A. (2019, April). Comparison of machine learning algorithms for predicting traffic accident severity. In *2019 IEEE Jordan International Joint Conference On Electrical Engineering And Information Technology (JEEIT)* (pp. 272-276). https://doi.org/10.1109/JEEIT.2019.8717393

Bokaba, T., Doorsamy, W., & Paul, B. S. (2022). Comparative study of machine learning classifiers for modelling road traffic accidents. *Applied Sciences*, *12*(2), 828. https://doi.org/10.3390/app12020828

Borucka, A. (2018). Risk analysis of accidents in Poland based on ARIMA model. *Proceedings of 22nd International Scientific Conference, Transport Means*, 162–166.

Box, G. E. P., Jenkins, G. M., & Wisconsin. (1970). *Time series analysis: Forecasting and control* (2nd ed.). San Francisco: Holden-Day.

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control* (Revised ed.). San Francisco: Holden-Day.

Brockwell, P. J., & Davis, R. A. (2002). *Introduction to time series and forecasting* (2nd ed.). New York, NY: Springer New York.

Çodur, M. Y., & Tortum, A. (2015). An Artificial Neural Network model for highway accident prediction: A case study of Erzurum, Turkey. *Promet – Traffic & Transportation*, *27*(3), 217-225. https://doi.org/10.7307/ptt.v27i3.1551

Feng, T., Zheng, Z., Xu, J., Liu, M., Li, M., Jia, H., & Yu, X. (2022). The comparative analysis of SARIMA, Facebook Prophet, and LSTM for road traffic injury prediction in Northeast China. *Frontiers in Public Health, 10*. https://doi.org/10.3389/fpubh.2022.946563

García de Soto, B., Bumbacher, A., Deublein, M., & Adey, B. T. (2018). Predicting road traffic accidents using Artificial Neural Network Models. *Infrastructure Asset Management*, *5*(4), 132-144. https://doi.org/10.1680/jinam.17.00028

Gazder, U., Ahmed, A., & Shahid, U. (2018). Application of Artificial Neural Networks for exploratory analysis of small dataset. *2018 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, 1-6. https://doi.org/10.1109/3ICT.2018.8855768

Ghedira, A., Kammoun, K., & Saad, C. Ben. (2018). Temporal analysis of road accidents by ARIMA model: Case of Tunisia. *International Journal of Innovation and Applied Studies*, *24*(4), 1544–1553.

Gilani, V. N. M., Hosseinian, S. M., Ghasedi, M., & Nikookar, M. (2021). Data-Driven urban traffic accident analysis and prediction using Logit and machine learning-based pattern recognition models. *Mathematical Problems in Engineering*, *2021*. https://doi.org/10.1155/2021/9974219

Gutierrez-Osorio, C., & Pedraza, C. (2020). Modern data sources and techniques for analysis and forecast of road accidents: A Review. *Journal of Traffic and Transportation Engineering (English Edition)*, *7*(4), 432–446. https://doi.org/10.1016/j.jtte.2020.05.002

Halim, Z., Kalsoom, R., Bashir, S., & Abbas, G. (2016). Artificial intelligence techniques for driving safety and vehicle crash prediction. *Artificial Intelligence Review*, *46*(3), 351-387. https://doi.org/10.1007/s10462-016-9467-9

Hassouna, F. M. A., & Al-Sahili, K. (2020). Practical minimum sample size for road crash time-series prediction models. *Advances in Civil Engineering*, *2020*, 1-12. https://doi.org/10.1155/2020/6672612

Ihueze, C. C., & Onwurah, U. O. (2018). Road traffic accidents prediction modelling: An analysis of Anambra state, Nigeria. *Accident Analysis and Prevention*, *112*, 21-29. https://doi.org/10.1016/j.aap.2017.12.016

Junus, N. W. M., & Ismail, M. T. (2014). Predicting road accidents: Structural time series approach. *AIP Conference Proceedings*, *1605*, 816–821. https://doi.org/10.1063/1.4887695

Junus, N. W. M., Ismail, M. T., & Arsad, Z. (2014). Behavior of road accidents: Structural time series approach. *AIP Conference Proceedings*, *1635*, 780–787. https://doi.org/10.1063/1.4903671

Junus, N. W. M., Ismail, M. T., Arsad, Z., & Rahman, R. A. (2017). Malaysia road accidents influences based on structural time series analysis. *Applied Mathematics and Information Sciences*, *11*(4), 1029–1039. https://doi.org/10.18576/amis/110409

Kim, H., Kim, J. T., Shin, S., Lee, H., & Lim, J. (2022). Prediction of run-off road crash severity in South Korea's highway through Tree Augmented Naïve Bayes learning. *Applied Sciences*, *12*(3), 1120. https://doi.org/10.3390/app12031120

Lazim, M. A. (2011). *Introductory business forecasting a practical approach* (3rd ed.). Shah Alam, Selangor: UiTM Press.

Lee, J., Yoon, T., Kwon, S., & Lee, J. (2019). Model evaluation for forecasting traffic accident severity in rainy seasons using machine learning algorithms: Seoul city study. *Applied Sciences*, *10*(1), 129. https://doi.org/10.3390/app10010129

Lokala, U., Nowduri, S., & Sharma, P. K. (2017). Road accidents big data mining and visualization using Support Vector Machines. *World Academy of Science, Engineering and Technology-International Journal of Computer and Systems Engineering*, *10*(8). Retrieved from https://corescholar.libraries.wright.edu/knoesis/1137

Maulana, S. I., Syaufina, L., Prasetyo, L. B., & Aidi, M. N. (2019). Spatial logistic regression models for predicting peatland fire in Bengkalis regency, Indonesia. *Journal of Sustainability Science and Management, 14*(3), 55-66.

Merabet, F., & Zeghdoudi, H. (2020). On modelling seasonal ARIMA series: Comparison, application and forecast (Number of injured in road accidents in Northeast Algeria). *WSEAS Transactions on Systems and Control*, *15*, 235–246. https://doi.org/10.37394/23203.2020.15.25

Ministry of Transport Malaysia. (2011). *Statistik Pengangkutan Malaysia 2010*. https://www.mot.gov.my/my/Statistik%20Tahunan%20Pengangkutan/Statistik%20Pengangkutan%20Malaysia%202010.pdf

Ministry of Transport Malaysia. (2019). *Statistik Pengangkutan Malaysia 2018.* https://www.mot.gov.my/my/Statistik%20Tahunan%20Pengangkutan/Statistik%20Pengangkutan%20Malaysia%202018.pdf

Ministry of Transport Malaysia. (2020). *Statistik Pengangkutan Malaysia 2019.* https://www.mot.gov.my/my/Statistik%20Tahunan%20Pengangkutan/Statistik%20Pengangkutan%20Malaysia%202019.pdf

Nashwan, M. S., Ismail, T., & Ahmed, K. (2019). Non-stationary analysis of extreme rainfall in Peninsular Malaysia. *Journal of Sustainability Science and Management, 14*(3), 17-34.

Nasir, A. S. M., Husainizam, H. H., Halim, N. S., Malek, I. A., & Malek, H. A. (2018). Modelling road accidents in Malaysia. *ESTEEM Academic Journal*, *14*, 66-76.

Qian, Y., Zhang, X., Fei, G., Sun, Q., Li, X., Stallones, L., & Xiang, H. (2020). Forecasting deaths of road traffic injuries in China using an Artificial Neural Network. *Traffic Injury Prevention, 21*(6), 407-412. https://doi.org/10.1080/15389588.2020.1770238

Radzuan, N. Q., Hassan, M. H. A., Majeed, A. P. P. A., Kassim, K. A. A., Musa, R. M., Razman, M. A. M., & Othman, N. A. (2020). Forecasting road deaths in Malaysia using Support Vector Machine. In *ECCE2019* (pp. 261–267). Lecture Notes in Electrical Engineering (Vol. 632). Singapore: Springer. https://doi.org/10.1007/978-981-15-2317-5_22

Radzuan, N. Q., Hassan, M. H. A., Majeed, A. P. P. A., Musa, R. M., Razman, M. A. M., & Kassim, K. A. A. (2020). Predicting serious injuries due to road traffic accidents in Malaysia by means of Artificial Neural Network. In Jamaludin, Z., & Mokhtar, A. M. N. (Eds.), *Intelligent Manufacturing and Mechatronics* (pp. 75-80). Lecture Notes in Mechanical Engineering (LNME). Singapore: Springer. https://doi.org/10.1007/978-981-13-9539-0_8

Silva, P. B., Andrade, M., & Ferreira, S. (2020). Machine learning applied to road safety modeling: A systematic literature review. *Journal of Traffic and Transportation Engineering* (English ed.), *7*(6), 775-790. https://doi.org/10.1016/j.jtte.2020.07.004

Slimani, N., Slimani, I., Amghar, M., & Sbiti, N. (2020). Road traffic forecasting using a real data set in Morocco. *Procedia Computer Science, 177*, 128–135. https://doi.org/10.1016/j.procs.2020.10.020

Suganya, E., & Vijayarani, S. (2017). Analysis of road accidents in India using data mining classification algorithms. *2017 International Conference on Inventive Computing and Informatics (ICICI),* 1122-1126. https://doi.org/10.1109/ICICI.2017.8365315

Suhaimi, N., Ghazali, N. A., Nasirt, M. Y., Mokhtar, M. I. Z., Ramli, N. A., Yusof, N. F.

F. M., & Ul-Saufie, A. Z. (2019). Daytime ozone concentration prediction using statistical models. *Journal of Sustainability Science and Management, 14*(3), 7-11.

Tan, I. A. W., Tan, M. C., Sim, D. H. H., Lim, L. L. P., & Chin, S. F. (2022). Modelling the effect of wood and maize cob-derived biochar application on soil dynamics and maize growth for sustainable agriculture. *Journal of Sustainability Science and Management, 14*(3), 3-23.

Wai, A. H. C., Seng, S. Y., & Fei, J. L. W. (2019). Fatality involving road accidents in Malaysia: A comparison between three statistical models. *ACM International Conference Proceeding Series*, 101–105. https://doi.org/10.1145/3343485.3343494

World Health Organization (WHO). (2020). *World Health Statistics 2020: Monitoring Health for the SDGs, Sustainable Development Goals.* https://apps.who.int/iris/bitstream/handle/10665/332070/9789240005105-eng.pdf

World Health Organization (WHO). (2021). *World Health Statistics 2021: Monitoring Health for the SDGs, Sustainable Development Goals*. https://apps.who.int/iris/bitstream/handle/10665/342703/9789240027053-eng.pdf

Zamzuri, Z. H., Shabadin, A., & Ishak, S. Z. (2019). Bayesian network of traffic accidents in Malaysia. *Journal of Information and Communication Technology*, *18*(4), 473-484. https://doi.org/10.32890/jict2019.18.4.4

Zhang, G. P. (2012). Neural Networks for time-series forecasting. In Rozenberg G., Bäck T., & Kok J. N. (Eds.), *Handbook of natural computing* (pp. 461–477). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-92910-9_14

Zheng, M., Li, T., Zhu, R. U. I., Tang, M., Cui, Z., Chen, J., Ma, Z., & Wang, Z. (2019). Traffic accident's severity prediction: A deep-learning approach-based CNN network. *IEEE Access*, *7*, 39897–39910. https://doi.org/10.1109/ACCESS.2019.2903319

Zivot, E., & Wang, J. (2006). Nonlinear time series models. In Zivot, E & Wang, J. (Eds.), *Modeling financial time series with S-PLUS* (2nd ed., pp. 653-712). New York, NY: Springer. https://doi.org/10.1007/978-0-387-32348-0_18

Zolala, F., Haghdoost, A. A., Ahmadijouybari, T., Salari, A., Bahrampour, A., Baneshi, M. R., & Razzaghi, A. (2016). Forecasting the trend of traffic accident mortality in west Iran. *Health Scope*, *5*(3), e31336. https://doi.org/10.17795/jhealthscope-31336