UITMCNS SEREMBAN

J E M U R

JOURNAL OF EXPLORATORY MATHEMATICAL UNDERGRADUATE RESEARCH

RECENT **MATHEMATICAL PROJECTS** AT UITM, *SEREMBAN CAMPUS*

Vol. 002 | PREPARED BY: KPPIM, UiTM N. SEMBILAN. | DATE: 05 OCT 2024

Photography courtesy of Porapak Apichodilok

TIME SERIES FORECASTING OF ROAD ACCIDENT IN MALAYSIA BY USING BOX-JENKINS AND UNIVARIATE MODEL

Nur Farah Bahiah Mohd Zarmi[1], Nur Balqish Zainal[2] and Norliana Mohd Lip[3,*]
[1,2,3]Mathematical Sciences Studies, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Negeri Sembilan, Seremban Branch, 70300 Seremban, Negeri Sembilan.
[*]norliana7287@uitm.edu.my

Abstract

This study aims to predict future road accidents in Malaysia using time series forecasting techniques. The objectives are to determine the best model between Box-Jenkins and Univariate methods, and to forecast road accident numbers based on the model with the lowest error measures of MAPE and RMSE. The study analyzes Malaysia's historical and current traffic accident trends to provide insights for forecasters, government agencies, and transportation firms. Researchers can also build upon the findings to further investigate traffic accident prediction in Malaysia. The results show that the Double Exponential Smoothing model provides the most accurate forecasts for 2024 road accident numbers in Malaysia. This information can help authorities implement targeted interventions to improve road safety and reduce the significant human and economic toll of traffic accidents.

Keywords: Box-Jenkins, ARIMA, Univariate Model, Single Exponential Smoothing (SES), Double Exponential Smoothing (DES)

1. Introduction

A road accident is an unexpected event causing loss or injury without the injured person's fault, for which legal relief may be sought. According to the World Health Organization (2023), approximately 1.3 million people die in road accidents annually. Gimino G. and T.A.(2023) mentioned that Malaysia reported 545,588 road accidents in 2022, resulting in 6,080 deaths, as shared by Transport Minister Anthony Loke. Human error, such as driving under the influence, speeding, and distracted driving, is frequently cited as the main cause of traffic accidents. Unfavorable weather and poor road conditions also significantly impact accident rates.

Malaysian roads experience various types of accidents, each presenting distinct risks. Multi-vehicle collisions from tailgating and inadequate following distances cause chaotic pileups. Head-on collisions, the most critical, result in severe injuries or fatalities, especially at high speeds. Low-speed accidents, though seemingly less severe, still risk casualties, particularly for pedestrians and cyclists. Merging accidents highlight the importance of proper speed and blind spot checks in congested areas. Recognizing and addressing these accident dynamics is crucial for enhancing road safety in Malaysia (Ali, 2023).

Predicting traffic accidents through time series analysis involves identifying patterns and trends in past data. Understanding seasonal, monthly, or daily fluctuations in accident frequency is crucial for developing prediction models. Time series analysis helps forecast peak accident periods, allowing authorities to allocate resources more efficiently during high-risk times. Additionally, it evaluates the long-term effectiveness of safety initiatives, enabling the optimization of road safety strategies. Predictive models provide insights into the temporal dynamics of traffic accidents, leading to focused and timely precautions that improve traffic safety.

This study compares Box-Jenkins and Univariate models to determine the most accurate method for forecasting road accidents. The objectives are to identify the most effective model and use it to forecast future road accidents in Malaysia.

2. Methodology

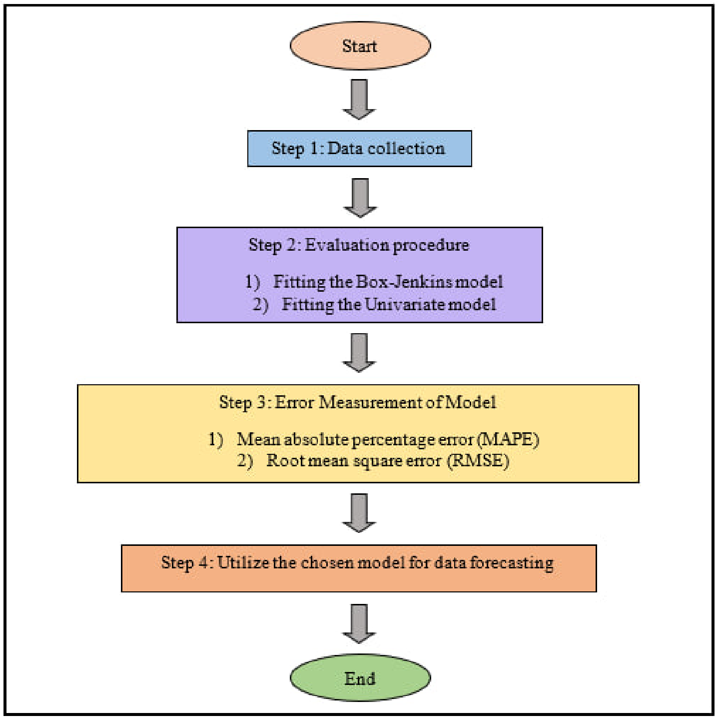

Figure 1: Research Framework Of The Study

2.1. *Data collection*

The data used in this study comes from secondary sources, specifically the Traffic Investigation and Enforcement Department (JSPT) of the Royal Malaysia Police (RMP). The analysis includes monthly data from January 2019 to December 2023.

2.2. *Box-Jenkins Model*

The Box-Jenkins model, also known as the ARIMA (AutoRegressive Integrated Moving Average) model, is a powerful and widely used approach for time series forecasting. Developed by George Box and Gwilym Jenkins, it combines autoregressive (AR) processes, where future values are regressed on their own lagged values; differencing (I), which involves subtracting previous values to achieve stationarity; and moving averages (MA), which model the error term as a linear combination of past errors. The ARIMA model is identified and estimated through a systematic process of model selection, parameter estimation, and diagnostic checking, making it highly adaptable to various types of time series data. Its strength lies in its ability to handle non-stationary data by transforming it into a stationary series, making it suitable for a wide range of forecasting applications. The ARIMA model is expressed by Lazim (2011):

$$Y_t = Y_{t-1} + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} + \epsilon_t \tag{1}$$

where;
Y_t : the time series at time t,
$\phi_1, \ldots, \phi_p$: the autoregressive parameters,
$\theta_1, \ldots, \theta_q$: the moving average parameters,
ϵ_t : the error term at time t,
$Y_{t-1}, \ldots, Y_{t-p}$: the lagged values of the time series, and
$\epsilon_{t-1}, \ldots, \epsilon_{t-q}$: the lagged values of the error term.

The Autoregressive Integrated Moving Average (ARIMA) model, represented as ARIMA (p, d, q), is developed for non-stationary data to achieve stationarity by differencing the variable d times. In this model, p denotes the autoregressive process, and q represents the moving average process. For example, ARIMA(1,1,1) is expressed as:

$$w_t = \mu + \phi_1 w_{t-1} - \theta_1 \epsilon_{t-1} + \epsilon_t \tag{2}$$

where;
$w_t = y_t - y_{t-1}$: the first differenced series,
μ : the constant term,
ϕ_1 : the impact of the previous differenced value,
θ_1 : the influence of the previous error, and
ϵ_t : the current error term.

2.3. ***Univariate Model***

2.3.1. *Single Exponential Smoothing*

The time series forecasting technique known as Single Exponential Smoothing (SES) is applied to univariate data that without a trend or seasonal pattern.The method works by applying exponentially decreasing weights to past observations, which means more recent observations are given more weight than older ones. The smoothing constant, α is the only parameter required. The method can be written as equation (3) mentioned by Rosyid et al. (2019):

$$F_{t+m} = \alpha y_t + (1-\alpha) F_t \tag{3}$$

where;
F_{t+m} : Forecast value for m period ahead
α : Smoothing constant
y_t : Actual value

2.3.2. *Double Exponential Smoothing*

Double exponential smoothing is an extension of single exponential smoothing that is used for time series data with a trend. In order to overcome the drawbacks of single exponential smoothing, a technique to take trends in the data is included. It provides a way to forecast future values by adjusting for both the level and the trend in the data. There are two smoothing constant, α and β that is the parameter required. In order to get the prediction value, step on finding stationary value and trend value, the method can be written as equation (4) and equation (5) mentioned by Rosyid et al.(2019):

Stationary value

$$S_t = \alpha y_t + (1-\alpha)(S_{t-1} + T_{t-1}) \tag{4}$$

Trend value

$$T_t = \beta(S_t - S_{t-1}) + (1-\beta) T_{t-1} \tag{5}$$

Forecasting value

$$F_{t+m} = S_t + T_t m \tag{6}$$

where;
S_t : Stationary value
T_t : Trend value
y_t : Actual value
α : Smoothing constant
β : Smoothing constant
F_{t+m} : Forecast value for m period ahead

2.4. *Error Measurement of Model*

The Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE), which are provided in equations (6) and (7), respectively, were used to evaluate the accuracy of the models in this study.

2.4.1. *Root Mean Square Error (RMSE)*

$$RMSE = \sqrt{\frac{1}{n}\sum_{t}^{n}(x_t - \hat{x_t})^2} \tag{7}$$

2.4.2. *Mean Absolute Percentage Error (MAPE)*

$$MAPE = \frac{1}{n}\sum_{t}^{n}|(\frac{x_t - \hat{x}_t}{x_t}) \times 100| \tag{8}$$

where;
x_t : Actual observed value
$\hat{x}_t$: Predicted value
n : The number of predicted values

3. Result and Discussion

The first findings of the study demonstrate the summary of error test that was utilised to determine which ARIMA models were most effective in predicting the frequency of traffic accidents for Box-Jenkins approach. Table 1 illustrates that, when compared to other models, ARIMA (1, 1, 1) was chosen to be the most effective model to use because its RMSE and MAPE values were the lowest.

Table 1: Selected ARIMA Model

Model	RMSE	MAPE
ARIMA(1,1,1)	16239.8200	45.0952
ARIMA(1,1,2)	16528.1000	45.8986
ARIMA(1,1,3)	16812.3500	46.6681

The data in Table 2 shows the RMSE and MAPE for SES and DES models. The DES model outperforms the SES model with lower RMSE (475.9685) and MAPE (1.1338), indicating its superiority for Univariate Models. The optimal α and β values for DES are 0.9596 and 0.0536, respectively. The RMSE value reflects the average error magnitude, while the MAPE value indicates the average percentage error. Despite potential absolute error variations, the DES model maintains accuracy, balancing relative accuracy and absolute error. This demonstrates the effectiveness of the DES model in forecasting time series data.

Table 2: Comparison of SES and DES

Model	RMSE	MAPE
Single Exponential Smoothing (SES)	4065.5544	9.6934
Double Exponential Smoothing (DES)	475.9685	1.1338

Table 3 shows DES has a smaller MAPE and RMSE than ARIMA(1,1,1), with values of 475.9685 and 1.1338, respectively. High accuracy of the model may be attained when the RMSE and MAPE values are at their lowest. As a result, the DES model is determined to be the most accurate model for estimating the number of Malaysia's road accident for 2024.

Table 3: Comparison of ARIMA(1,1,1) and DES

Model	RMSE	MAPE
ARIMA(1,1,1)	16239.8200	45.0952
Double Exponential Smoothing (DES)	475.9685	1.1338

As shown in Table 3 it is obvious that the DES model is the most accurate model for estimating the number of Malaysia's road accident in 2024. The predicted value's outcome is shown in the Figure 2 and Table 4 below.

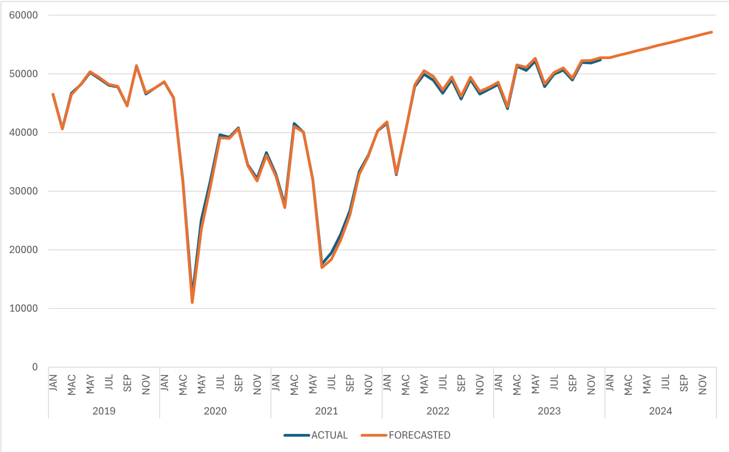

Figure 2: Graph Of Forecast Value Of Road Accident In Malaysia

Table 4: Forecasted Number of Accidents in 2024

Year	Month	Forecasted Number of Accidents in 2024, Y_t
2024	Jan	52,782.03367
	Feb	53,176.70636
	Mac	53,571.37905
	Apr	53,966.05174
	May	54,360.72443
	Jun	54,755.39712
	Jul	55,150.06981
	Aug	55,544.74250
	Sept	55,939.41519
	Oct	56,334.08788
	Nov	56,728.76058
	Dec	57,123.43327

4. Conclusion

The objective of this study was to model 60 data points from January 2019 to December 2023 of road accident data from Malaysia using Box-Jenkins ARIMA model and Univariate models, specifically SES and DES. The second objective invloved comparing error measurements, which is MAPE and RMSE, in order to choose the best model. Box-Jenkins was represented by ARIMA(1,1,1), and the best Univariate model was DES. According to the comparison, DES was the best model since it had the lowest RMSE and MAPE. As a result, from January 2024 to December 2024, DES was employed to forecast road accidents, with the eventual goal of reaching 60,000 cases each month. The study discovered that DES is the most accurate forecasting model, offering crucial information to policymakers and traffic experts to help them put plans into place that would lessen road accidents, which are a major cause of mortality in Malaysia. Besides, there are two recommendation machine learning algorithm which Naïve Bayes and Neural Network to come out with accurate result. The ability of neural networks to handle complex, non-linear interactions between variables like weather, traffic, and time of day makes them ideal for predicting traffic accidents (Gatarić et al., 2023).

References

Ali, M. (2023). 15 Most Common Types Of Road Accidents. https://www.hseblog.com/types-of-road-accidents/. [Accessed 22-03-2024].

Gatarić, D., Ruskic, N., Aleksić, B., Đurić, T., Pezo, L., Lončar, B., and Pezo, M. (2023). Predicting road traffic accidents-artificial neural network approach. *Algorithms*, 16:257.

Gimino G., Rahim R., V. R. and T.A., Y. (2023). 915,874 road accidents recorded throughout 2021 and 2022, says Transport Ministry — thestar.com.my. [Accessed 22-03-2024].

Lazim, M. A. (2011). *Introductory business forecasting : a practical approach*. UiTM Press, Shah Alam, Selangor, 3rd ed edition. Includes 1 computer optical disc (4 3/4 in.).

Rosyid, H. A., Widiyaningtyas, T., and Hadinata, N. F. (2019). Implementation of the exponential smoothing method for forecasting food prices at provincial levels on java island. In *2019 Fourth International Conference on Informatics and Computing (ICIC)*, pages 1–5.

World Health Organization, W. (2023). Road traffic injuries. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries. [Accessed 24-03-2024].