Bindura University
of Science Education

COMMITMENT DISCIPLINE

B U S E

EXCELLENCE

FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF STATISTICS AND MATHEMATICS

TIME SERIES ANALYSIS OF MOTOR VEHICLE ACCIDENTS IN ZIMBABWE (1993-2023). A COMPARATIVE STUDY OF AUTOREGRESSIVE INTEGRATED MOVING AVERAGE AND ARTIFICIAL NEURAL NETWORK MODELS.

BY

ZVOMUYA TANYARADZWA (B202545B)

A DISSERTATION SUBMITTED IN PARTIAL FUFILMENT OF THE REQUIREMENTS FOR BSC.HONOURS IN STATISTICS AND FINANCIAL MATHEMATICS

SUPERVISOR: DR T. W. MAPUWEI

JUNE 2024

APPROVAL FORM

The undersigned certify that they have read and recommended to the Bindura University of Science Education for acceptance a dissertation entitled “TIME SERIES ANALYSIS OF MOTOR VEHICLE ACCIDENTS IN ZIMBABWE (1993-2023). A COMPARATIVE STUDY OF AUTOREGRESSIVE INTEGRATED MOVING AVERAGE AND ARTIFICIAL NEURAL NETWORK MODELS Submitted by ZVOMUYA TANYARADZWA, Registration Number B202545B in partial fulfilment of the requirements for the Bachelor of Science Honour’s degree in Statistics and Financial Mathematics.

ZVOMUYA TANYARADZWA…………. …….10/06/2024

B202545B Signature Date

Certified by

DR T. W. MAPUWEI ……..… ·……………… 10/06/2024

SUPERVISOR Signature Date

Certified

DR. M. MAGODORA …… Magodora ……………. 10 /06/2024

Chairman of Department Signature Date

DECLARATION

I Zvomuya Tanyaradzwa hereby declare that this submission is my own work apart from the references of other people’s work which has duly been acknowledged. I hereby declare that this work has neither been presented in whole nor in part for any degree at this university or elsewhere.

Author: Zvomuya Tanyaradzwa

Registration Number: B202545B

Signature: ………………

Date: 10 June 2024

DEDICATION

I dedicate this dissertation to Tobias Binyoli, Elizabeth Zvomuya, Lilian Zvomuya, Aunty G, CC Fae and Aunty N who have made sacrifices towards my personal and professional endeavours. I thank you for believing in my dreams.

ACKNOWLEDGEMENTS

I express my sincere gratitude in writing this acknowledgement as I could not have completed my thesis without the invaluable assistance of my supervisor, family and friends. I extend my deepest appreciation to my supervisor, DR T. W. Mapuwei, Mrs Hlupo and Mr Kusotera for consistently providing me with support throughout the development of my dissertation. Their patience, guidance, motivation, and extensive knowledge proved invaluable during the analysis and writing process. I also thank God for providing me with the care, strength, knowledge, and opportunity to pursue my education to this level. Additionally, I would like to thank the lecturers and staff of the Department of Statistics and Mathematics at Bindura University of Science Education for their academic support and knowledge sharing. To my family, I express my sincere gratitude for their unwavering encouragement, advice, and prayers all through my studies. I pray that the Lord bless them abundantly. Finally, to my colleagues, I consider you all as family, and I ask for God's guidance and protection to be with you always.

ABSTRACT

This study is a comprehensive time series analysis of yearly motor vehicle accidents in Zimbabwe from 1993 to 2023, comparing the performance of Multilayer Perceptron model (MLP) and the Autoregressive Integrated Moving Average (ARIMA). The aim was to identify the most accurate and reliable modelling approach for forecasting number of motor vehicle accidents in Zimbabwe. A quantitative research design for 31 data points of yearly motor vehicle accidents was employed in the study and the R-Software was used to perform data analysis. ARIMA model was developed using the Box-Jenkins model building strategy. The Augmented Dickey Fuller test revealed that the accident data was non-stationary. After first order differencing, the data became stationary. The model with the smallest corrected Akaike Information Criteria (AICc) and Bayesian Information Criteria (BIC) was chosen as the best model which is the ARIMA (1,0,0) model. The best-performing MLP model among the three that were created was 1-(10,5)-1. The performance evaluation metrics were used to compare the models against observed data from 2017 to 2023. Mean Absolute Error and Root Mean Square Error, were used as performance evaluation metrics. This study's conclusions indicate that MLP out performed ARIMA model and it was used for forecasting number of future accidents for the next 5 years. Future projections indicate a downward trend in the number of motor accidents. Even if the trend in future accidents was decreasing there is still a need to exercise more caution so as to reduce the occurrence of accidents related to road traffic crashes.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ARIMA	Autoregressive Integrated Moving Average
MLP	Multilayer Perceptron
WHO	World Health Organization
GDP	Gross Domestic Product
ADF	Augmented Dickey Fuller
ANN	Artificial Neural Networks
FNN	Feedforward Neural Networks
MA	Moving Averages
AR	Autoregressive
TLNN	Time Lagged Neural Networks
ZIMSTAT	Zimbabwe National Statistic
PACF	Partial Autocorrelation Function
RMSE	Root Mean Square Error
MAE	Mean Absolute error
RTA	Road Traffic Accidents

CHAPTER 1: INTRODUCTION

1.0 Introduction

This chapter lays the foundation for the research study. Using Artificial Neural Networks and Autoregressive Integrated Moving Averages with historical time series data collected as a comparison, it seeks to predict the future occurrence of road traffic accidents in Zimbabwe. The best fitting model for the accident data from 1993 to 2023 is found by comparing the time series models. The chapter therefore outlines the background of the study, research problem, objectives, research questions, significance of the study, assumptions of the study, potential limitations and implications for different stakeholders who may be interested in the findings.

1.1 Background of the Study

Motor vehicle accidents have profound implications for public safety and societies worldwide. Every year, approximately 1.35 million people die on the world's roads as a result of road traffic crashes, which is one of the leading causes of death globally with an additional 20-50 million non-fatal injuries resulting in long term disabilities (WHO, 2020). More than half of all road traffic deaths occur among vulnerable road users like pedestrians, cyclists and motorcyclists (WHO, 2018). Negesa & Dessie (2017) coined that road traffic accidents are a major cause of injury and death worldwide, as well as a public health problem. Global statistics shows that the most affected age group is between 15-44. Kuikel et al (2022), accounted that road traffic collisions are the eighth largest cause of death for all ages around the world.

The impact of motor vehicle accidents is disproportionately felt in low- and middle-income countries which account for 93% of global road fatalities despite having only 60 % of the world's motor vehicles (WHO, 2018). The economic burden of motor vehicle accidents is substantial with an estimated annual cost of 2.5trillion worldwide. Globally, the economic impact of traffic collisions is staggering with the costs associated with medical expenses, productivity loss and property damage estimated to be around 3% of the world's gross domestic product (GDP). In low- and middle-income countries these costs can be high as 5% of GDP, further straining already limited resources.

The African continent, home to approximately 16% of world's population, bears disproportionate share the global burden of road traffic fatalities. Africa has the highest rate of road traffic fatalities globally. World Health Organisation estimated 26.6 deaths per 100000 population in Africa as compared to an average of 17.4 deaths per 100000 globally. WHO (2020), reported the severance of road accidents in developing countries where millions are lost and properties are damaged. Vehicle safety standards in many African countries is often lacking. The prevalence of older poorly maintained vehicles on the roads increases the risk of accidents and their severity. Urbanisation has led to increased number motor vehicles and traffic congestion, often outpacing infrastructure developments.

Zimbabwe is a landlocked country and road transport is widely used. Zimbabwe's population has grown significantly over the last few decades, and the country has become more motorised and urbanised, which has increased the number of vehicles on its roads. The total number of accidents in Zimbabwe increased significantly between 2009 and 2010, rising by 61.2% from 20553 in 2009 to 26841 in 2010 (Mutangi, 2015). Similarly, the number of deaths from road accidents increased between 2009 and 2010 by 7.3%, and the trend continued in 2011, with a 12.8% increase from 2010 to 2011 (Mutangi, 2015). Mutangi (2015) also proposed that the increase in the number of motor vehicles in Zimbabwe from 2009 originated from the economy's dollarization and this has made it easier for individuals to acquire vehicles. This surge in motor vehicles has been accompanied by an alarming increase in the occurrence of road accidents, posing a serious threat to public safety and necessitating a comprehensive analysis of the underlying patterns and trends.

Zimbabwe National Road Traffic Accident Database reports that, the number of motor vehicle accidents in Zimbabwe has increased over the previous few decades. This disturbing trend underlines the urgent need for the development of a predictive model to forecast the frequency of traffic accidents and devise appropriate mitigation techniques. The results of the study added to the current body of information on road safety in Zimbabwe, providing evidence-based insights for policymakers, transportation authorities, and law enforcement organizations.

1.2 Statement of the Problem

Motor vehicle accidents have emerged as a significant public health and safety concern contributing to a substantial number of fatalities, injuries and economic losses. It is prudent that the concerned authorities like TSCZ know the future accidents for planning, budgeting and healthcare improvement purposes. The concerned department is not aware of the trend in future accident. This uncertainty makes it difficult for them to plan in terms of research allocations. This research therefore proposes a forecasting model which allows the concerned authorities to anticipate the direction of accidents and hence use that information for planning purposes. The purpose of this study is to compare the performance of ARIMA (Autoregressive Integrated Moving Average) and Artificial Neural Networks (ANN), in evaluating and forecasting motor vehicle accidents in Zimbabwe from 1993 to 2023. The aim is to determine the best model in capturing and identifying motor vehicle accidents.

1.3 Objectives of the Study

Ultimately, the study aims to perform a time series analysis of Zimbabwe motor vehicle accidents using data from 1993 to 2023.

The specific objectives are as follows:

1. To analyse the temporal patterns and trends of motor vehicle accidents in Zimbabwe over the period 1993 to 2023.
2. To fit ANN and ARIMA models to the data.
3. To compare the performance of the ARIMA and ANN models based on various evaluation metrics
4. To forecast future accidents from 2024 to 2028 using the best performing model.

1.4 Research Question

1. What are the temporal patterns and trends of motor vehicle accidents in Zimbabwe from 1993 to 2023?
2. Can ARIMA and ANN models accurately fit the accident data?
3. How well does the neural network approach compare with ARIMA approach?
4. Can future accidents be forecasted over the study period?

1.5 Scope of the Study / Delimitation of the Study

The study examined how well the ARIMA and ANN models performed on modelling Zimbabwe's Road accidents from 1993 to 2023. It specifically examined traffic collisions within the context of Zimbabwe. The study collected historical accident data from Traffic Safety Council of Zimbabwe, covering a substantial time span for analysis. It was only limited to one variable which is, the number motor vehicle accidents.

1.6 Significance of the Study

The study is significant in terms of informing evidence-based decision making, developing predictive models, and contributing to the knowledge base. The study holds immense potential to facilitate targeted interventions and policies by informing accident rates, improve crash prevention, and ultimately save lives in Zimbabwe.

1.7 Assumptions of the Study

1. The mean and variance of the time series data are stationary, which means they do not vary over time.

2. The data is not affected by autocorrelation meaning that the data points are not correlated.

1.8 Limitations of the Study

This study was on time series analysis of motor vehicle accidents in Zimbabwe from 1993 to 2023 and had limitations that were considered. It only focused on the time series data and neglected other factors that can cause motor accidents. The findings cannot be generalized beyond Zimbabwe due to its unique characteristics. The study only compared ARIMA and ANN models and did not consider other forecasting models.

1.10 Definition of Terms

Time series

A time series is a collection of data points gathered over time with the intention of identifying and interpreting trends and patterns (Brockwell & Davis, 2016).

Road Traffic Accident

Any event where at least one motor vehicle is involved and the public has the right of access to a private or public road and at least one person is hurt or killed is considered a road traffic accident (WHO, 2018).

Artificial Neural Networks

Russell and Norvig (2016), defined it as a computational model composed of interconnected nodes, known as artificial neurons or nodes inspired by the biological neural networks in the human brain.

Forecasting

Is a process of making predictions about the future trends based on historical data (Makridakis et al, 2018).

Forecasts

Forecasts refers to predictions or estimates of future events, trends or values based on historical data, analysis, and modelling (Hyndman and Anthanasopoulos, 2018).

1.11 Chapter Summary

This chapter introduced a topic of the study, highlighted its background, objectives, problem statement, limitations and delimitation associated with the study. The subsequent chapter, which reviews theoretical and empirical literature focusing on the ARIMA and ANN approaches, was made possible by the work that preceded it.

CHAPTER 2: LITERATURE REVIEW

2.0 Introduction

In this section, the researcher gives details of theoretical literature, empirical evidence and the research gap related to motor vehicle accidents in Zimbabwe from 1993 to 2023. Additionally, a proposed conceptual model was presented to guide the analysis and interpretation of the data.

2.1 Theoretical literature

Road traffic accidents

Motor vehicle accidents are a complex phenomenon influenced by various factors such as driver behaviour, road conditions, vehicle characteristics, and socioeconomic factors. Time series analysis techniques offer a valuable framework for comprehending trends in motor vehicle accidents, enabling accurate forecasting.

Theories associated with the occurrence of motor vehicle accidents

Risk Perception Theory

Slovic et al. (1988), suggested that individuals' subjective perception of risk influences their behaviour. They went on suggesting that drivers' perception of the likelihood and severity of accidents affects their driving decisions and adherence to safety regulations. It has been observed that personal experience, familiarity with the road environment, and perceived control influence risk perception and subsequently driver behaviour.

Theory of Planned Behaviour:

Ajzen (1991), described how attitudes, subjective norms and perceived behavioural control form basis of the theory. It suggested that individual attitudes toward risky driving behaviours, subjective norms influenced by social factors, and perceived control over driving behaviour play a significant role in predicting driver actions. He also argued that people’s attitudes influence their intention to perform the behaviour as summarized in Figure 2.1 below.

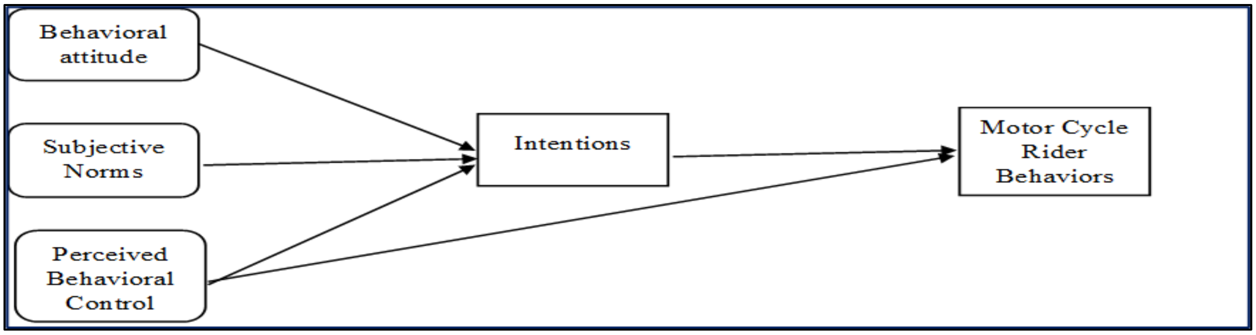


Figure 2.1: Theory of Planned Behaviour

2.1.1 Time series analysis

Brockwell & Davis (2016), defined time series as a sequence of data points measured at regular time intervals. These observations can be represented as X_1,X_2, X_3 ... X_t where t denotes the time period and X represents the corresponding value. Similarly, Ramasubramanian (2015) describes a series of data points recoded over time arranged chronologically. Univariate time series have records for a single variable. They can be discrete or continuous, with four components which are trend(T), seasonal variations(S), cyclical variations(C), and irregular variations(I).

2.1.2 Components of Time Series

The above components can be represented by the equation, $Y = S \times C \times I \times T$ where y is time series (Hyndman & Athanasopoulos, 2018).

The Trend

Stock and Watson, (2018) defined a trend as the long-term component of a time series that captures the gradual, persistent, and predictable changes in the data point. A trend can either be upward or downward and can be influenced by factors like population growth, price fluctuations, and economic changes.

Seasonality

Seasonality in time series refers to recurring patterns with consistent timing, direction, and magnitude, influenced by seasonal fluctuations, calendar changes, weather variations, business cycles, and cultural traditions (Brockwell & Davis, 2016).

Irregular (Unsystematic)

An unsystematic component is caused by random shocks like wars, strikes, earthquakes, floods, and revolutions, lacking a defined statistical technique for measurement. Figure 2.2 shows aspects of time series (Brockwell & Davis, 2016).

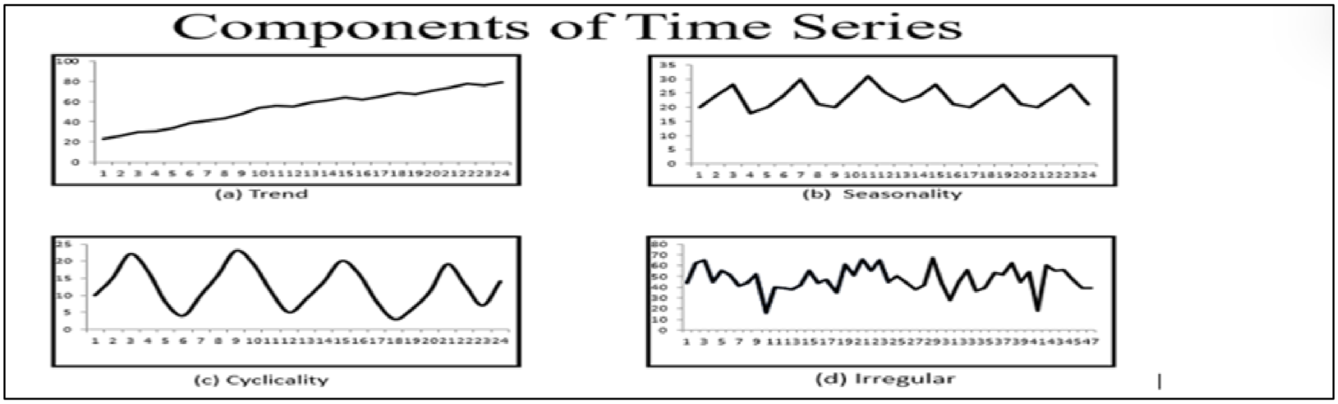


Figure 2.1.2: Components of Time Series

2.1.3 Assumptions of Time Series Analysis

Stationarity

Any sequence with mean and variance which are constant is considered to be stationary. (Hyndman & Athanasopoulos, 2018). Stationarity exists in three forms which are strict, weak and second-order stationarity. Strict stationarity implies that all moments and joint distribution of the series remain constants. Augmented Dickey-Fuller test is used to determine the stationarity of any given time series data. It is a popular choice for analysing the stationarity of a series. If the p-value from an Augmented Dickey-Fuller test is greater than the level of significance (alpha value), we reject the null hypothesis (H0) and take appropriate measures to address non-stationarity (Ramasubramanian, 2015).

Independence

Autocorrelation must not exist, and residuals must be independent. To determine if the residuals have positive autocorrelation, apply the Durbin-Watson test. Plotting residuals against the fitted values is another technique that is suggested. The plot should be unstructured if the model is accurate. Partial auto corelation function (PACF) and auto-corelation function (ACF) plots are also used to test for (Mapuwei et al, 2022).

Normality

Residuals data should follow a normal distribution. Normality can be assessed through the use of density plot or histogram, box plots, percent-percent (P-P) plots, quantile-quantile (Q-Q) plots, Shapiro-Wilk test and the Kolmogorov-Smirnov test. The Shapiro-Wilk test is a commonly used test that assesses the hypothesis that the data are from a normal distribution (Mapuwei et al, 2022.

Homoskedasticity

Plotting the residuals scatter plot demonstrates that the residual's variance remains consistent, with a rectangular shape enveloping zero horizontal levels and no discernible trends (Mapuwei et al, 2022).

2.1.4 Models in Time Series Analysis

Autoregressive (AR) Model

A sequence is said to be autoregressive if its current value is affected by previous values plus a random shock (Da Hye et al, 2021). Thus

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + a_t \text{ ………………………… (2.1)}$$

where Y_t– Current Value, Y_{t-p} – a value at lag p

a_t – White noise error

$\phi_1, \phi_2 \ldots \phi_p$– Parameter of the model which is estimated from the data.

Moving Averages (MA) Model

A moving average model make use of error lags in a forecasting process. Schaffer et al, (2021) discovered that there is a large surge in the ACF. Order q of MA model is established based on the number of ACF significant spikes, and PACF decreases sequentially. This model, the subsequent event is determined as the average of the previous event, considering the short-term autocorrelation of the time series. ACF plots are used to estimate the order q of a model. The formula for the MA (q) model can be expressed as follows:

$$Y_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q} \text{ ……………………….. (2.2)}$$

With a_t as the white noise and a_{t-q} as the white noise at lag q

Autoregressive Moving Averages (ARMA) Model

A combination of AR and MA gives us the ARMA model. When an equation of first-order autoregressive (AR) model reaches the initial point, it leads to an infinite moving average. To utilize the ARMA model effectively, we determine p and q values. The value of p corresponds to significant terms in the autocorrelation function (ACF), while q represents the number of significant terms in the partial autocorrelation function (PACF). If a time series conforms to an ARMA (p, q) model, it is considered to exist (Box, Jenkins, Reinsel, & Ljung, 2015).

$$Y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} \ldots + \phi_q \varepsilon_{t-q} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.3)$$

Where ε_t is the white noise process.

Seasonal Autoregressive Integrated Moving Average (SARIMA) Model

A SARIMA model is applicable to both periodic and non-periodic data. It utilizes the appropriate order of seasonal fluctuation to make the series stationary. The model is defined as follows

$$(1-\phi_1\mathrm{B})(1-\phi_1\mathrm{B}^4)(1-\mathrm{B})(1-\mathrm{B}^4)y_t = (1+\phi_1\mathrm{B})(1+\phi_1\mathrm{B}^4)\varepsilon_t \ldots\ldots\ldots \text{(2.4)}$$

B backshift operator with '4' the seasonal lag (Ramasubramanian, 2015)

2.1.5 Forecasting Approaches

2.1.5.1 Box-Jenkins Data Analysis Procedures (ARIMA)

An ARIMA model is formed from Autoregressive Moving Averages. The first step is to make the dataset stationary (Mboso, 2022). It is an iterative process involving four phases which are model identification, estimation, diagnostic checking, and time series forecasting. The main goal is to find the most parsimonious model from a broad range of ARIMA models. An iterative process is repeated multiple times until a satisfactory model is ultimately chosen. The time series' future values can then be predicted using this model, as stated by Khan & Alghulaiakh (2020). This provides a general framework for describing the ARIMA model.

$$\phi\ (B)\ Zt = \Theta\ (B)\ a\mathrm{t} \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. \quad (2.5)$$

where Zt is the variable being modelled, the unknown model parameter ϕ and Θ for a white noise process are estimated using the method of estimation which either least-squares or maximum likelihood, and the backward difference operator B is also employed.

2.1.5.2 Artificial Neural Networks

The concept of ANN is based on the real biological systems found in human brains, as explained by Schmidhuber (2015). Neurophysiology and cognitive theory are the two branches of brain science that inform neural network techniques. ANNs consist of connection patterns, neuron number, the learning algorithm and the activation function.

Artificial Neural Networks Architecture

Erguzel, et al., (2019) classified ANN as a semiparametric method. Complex functions are produced when a large group of neurons are connected in a suitable way. Signals are sent and received by the neurons from one another. The neurons are simple processors of information, and generally observe signals that other neurons are sending along the connections.

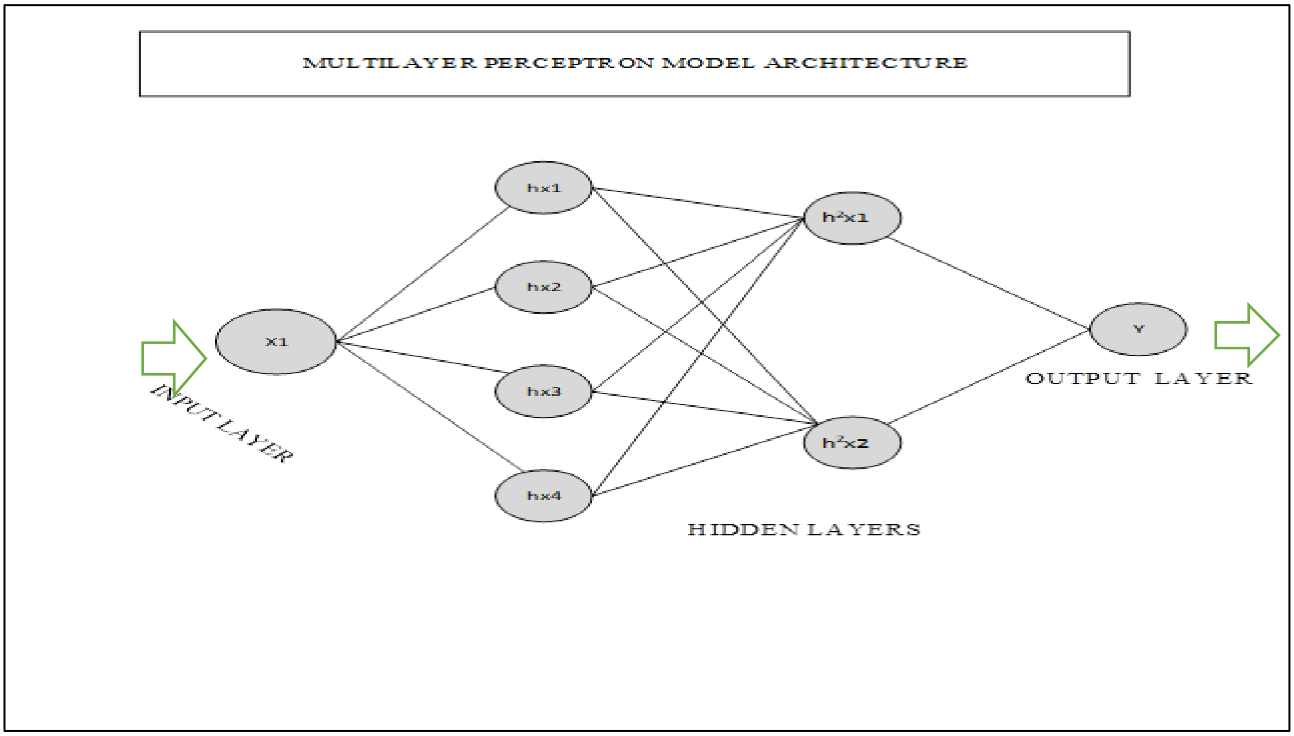


Figure 2.1.5: Multilayer Perceptron Architecture

Figure 2.2 above is a multilayer perceptron which is basically a model created from feedforward neural networks (FNN). The architecture above shows a model structure of 1-(4,2)-1. It has one input layer, two hidden layers with four and two neurons respectively and one output layer. The first layer involves *M* linear combinations of the *d*-dimensional inputs as summarized in the general equation below:

$$b_j = \sum_{i=0}^{D} w^{(1)}{}_{ji} x_{i\,,} \qquad {\scriptstyle j=1,2,\;.\;.\;.\;M} \qquad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (2.6)$$

Where w_{ji} are weights and the quantities b_j are the activations. Each of the activations is then transformed by a nonlinear activation function g, typically a sigmoid:

$$z_j = h(b_j) = \frac{1}{1 + exp(-b_j)} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (2.7)$$

However, in a multiclass problem, a softmax action function is used and it can be generalized as follows

$$h(b_j) = \frac{exp(b_j)}{\sum_{t=1}^{k} exp(b_j)} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (2.8)$$

Then by so doing we can arrive at the general equation for MLP and it can be written as follows

$$y_k = h\left(\sum_{j=0}^{M} w^{(2)}{}_{kj}\, b\left(\sum_{i=0}^{D} w^{(1)}{}_{ji}\, x_i\right)\right) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots.. (2.9)$$

Where subscript (1) and (2) represents first and second layers. Alternatively, the following MLP equation can be used.

$$y = f_s\left[\sum_{k=0}^{K} w^{0}{}_{1k}\left(f\sum_{n=0}^{N} w^{i}{}_{kn}\, u_n + b_n\right)\right] \dots\dots\dots\dots\dots\dots\dots..\dots (2.10)$$

Where the network inputs u_n are the bias of the network b_n, f is the activation function of the intermediate layers, and f_s is the activation function in the output layer, y is the output signal, $w^{i}{}_{kn}$ is the weight of the intermediate layer and $w^{0}{}_{1k}$ is the connection of the output neurons (Qureshi et al, 2022).

MLP Training: Back-propagation of Error

The process of training a neural network model using the gradient descent starting from output layer can be referred to as backpropagation. This stage involves the idea of chain rule differentiation method when defining error function E and evaluating the derivatives $\partial E \big/ \partial {w^{(2)}}_{kj}$ and $\partial E \big/ \partial {w^{(1)}}_{kj}$. Through a training process, the error gradients are the product of the derivative of the error at the output of the weights and the value at the input to the weight (Shimodaira, 2015). Calculations of the gradient backward should be done independently and recomputing of the weights should be attached between each node. Weights are optimized so as to accurately plot arbitrary inputs to output.

2.1.6 Other Neural Network Models

Long Short-term Memory (LSTM)

The LSTM is a class of recurrent neural networks (RNN). It is a dynamic system with temporal states (Du and Swamy, 2014). Rehan, Swarna and Dipayan (2020) described the movement of data in RNNs. The data travels in a way that, at each node, the network learns from both the current and previous inputs, and is therefore able to share the weights over time.

Time lagged Neural Networks (TLNN)

It is a network in which temporal dependence in time series data is captured by supplying the network with present value of the input, x_t in addition to p past values of the input x_{t-1}, x_{t-2} $\dots x_{t-p}$. The relationship between output $y_t$and the input is assumed to be of the form $y_t = f\left(x_t, x_{t-1}, x_{t-2} \dots, x_{t-p}\right) + e_t$ where e_t is a zero mean Gaussian variable with variance and $f(.)$ is a non-linear function in its arguments (Kihoro et al, 2004).

2.2 Empirical Literature

Mutangi (2015), utilized the Box-Jenkins ARIMA method to model yearly road accident in Zimbabwe from 1997 to 2013 and forecasted up to 2018. The time series data was not stationary and differenced once to make it stationary. Three ARIMA models that were proposed by the study are ARIMA (0,1,0), ARIMA (1,1,0), and ARIMA (1,1,1) based on the ACF and PACF plots of the differenced series. The model with the lowest values was chosen as the best fit, based on a comparison of AICc and BIC values. From the three models tested, ARIMA (0,1,0) was found to be the most appropriate in analysing Zimbabwe Road accidents. The forecasting process retained the value at the forecast origin. However, Mutangi did not do a comparative study of ANN and ARIMA.

Wannuraw et al (2023), conducted a comparative analysis of SARIMA and ANN in modelling Selangor Road traffic from January 2011 up to December 2021. The study's aim was to forecast monthly road accident occurrences on federal and state roads in Selangor, Malaysia. The traditional univariate SARIMA and ANN models were employed and their performances were assessed. The results showed that ANN model outperformed the SARIMA model in both training and validation sets. This study demonstrated the ability of neural networks in forecasting road accidents, giving more flexibility and assumption-free methodology.

In their study conducted in 2021 Lind and Ridhagen compared ANN and AR baseline using M4 competition dataset. Twenty (20) observations were drawn from the sample and the two models were employed on the data set. The ANN models performed much better than the Autoregressive baseline giving lower values of performance evaluation metrics.

A separate study by Emenike and Kanu in 2017, investigated the relationship between distracted driving and road traffic collisions specifically in Port Harcourt. The study found that the use of mobile phones or gadgets by commercial drivers contributed to the majority of accidents.

West Arsi (2023), wrote their research paper on road traffic accidents in Ethiopia. The research used monthly data from January 2016 to December 2020. Statistical packages and Box-Jenkins approach were used in the data analysis. After applying first-order differencing to the data, a (2, 1, 3) model was the best model for the data. The model was then used for forecasting.

2.3 Research Gap

The research gap lies in the application of time series analysis to motor vehicle accidents in Zimbabwe from 1993 to 2023. Existing studies have primarily focused on the ARIMA model while neglecting the use of ANN. The gap arises from the need for a comparative analysis to evaluate the effectiveness of ANN as a complementary method for complex accident data. Furthermore, there is limited research specific to the Zimbabwean context, where accident dynamics may differ. Comparing the performance of ARIMA and ANN models would provide insights into their strengths and limitations in terms of accuracy, forecasting capabilities, interpretability, and computational efficiency.

2.4 Proposed Conceptual Model

Figure 2.4 is a proposed conceptual model of the study. The diagram shows elements associated with traffic crash and injuries. Travel behaviour and crash risk are the two fundamental theories which results in road accidents. Figure 2.5 show the flow chat depicting ARIMA and ANN methodologies used in this study. It shows stages followed when developing ARIMA and ANN as well as their comparison. The Figure also indicates performance metrics that were used for comparison of the two time series models.

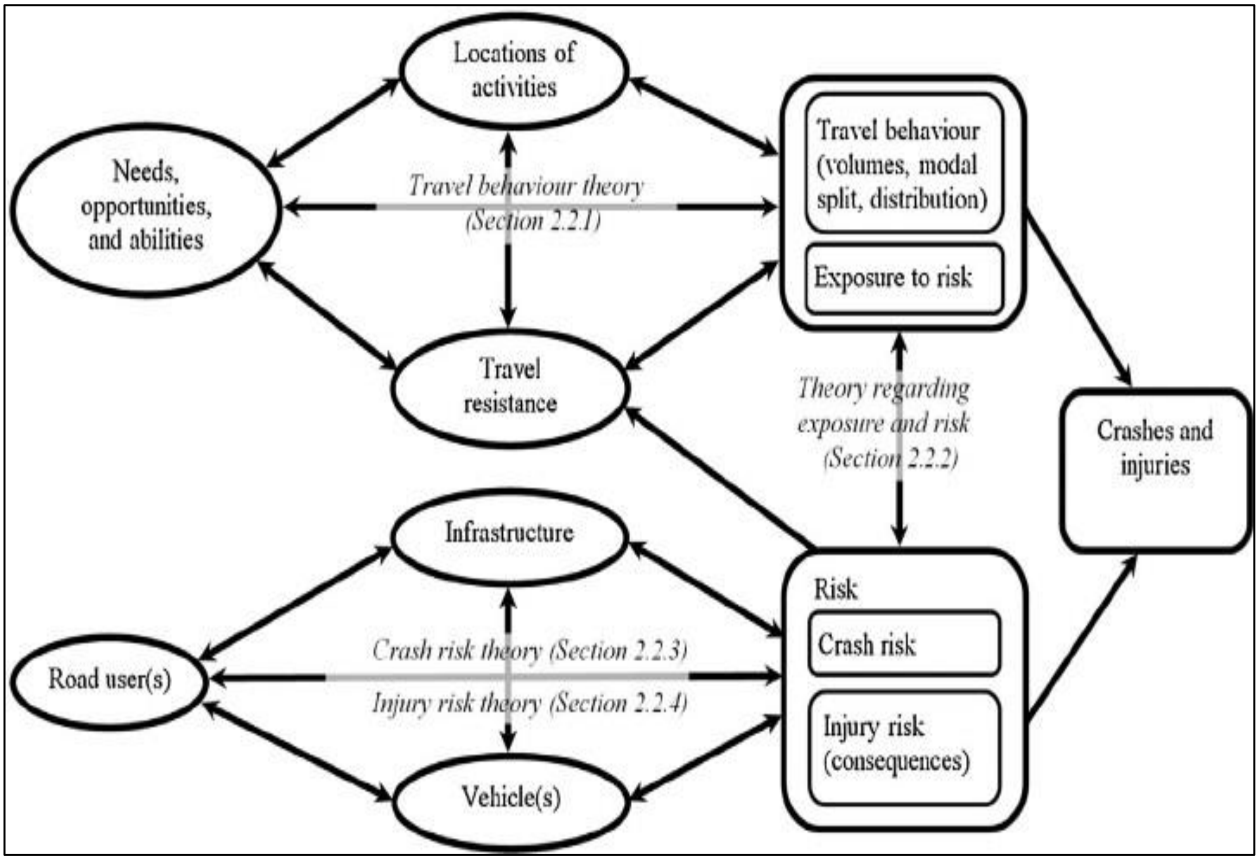

Locations of activities
Needs, opportunities, and abilities
Travel behaviour theory (Section 2.2.1)
Travel behaviour (volumes, modal split, distribution)
Exposure to risk
Travel resistance
Theory regarding exposure and risk (Section 2.2.2)
Crashes and injuries
Infrastructure
Risk
Crash risk
Road user(s)
Crash risk theory (Section 2.2.3)
Injury risk theory (Section 2.2.4)
Injury risk (consequences)
Vehicle(s)

Figure 2.4.1 Proposed Conceptual Model

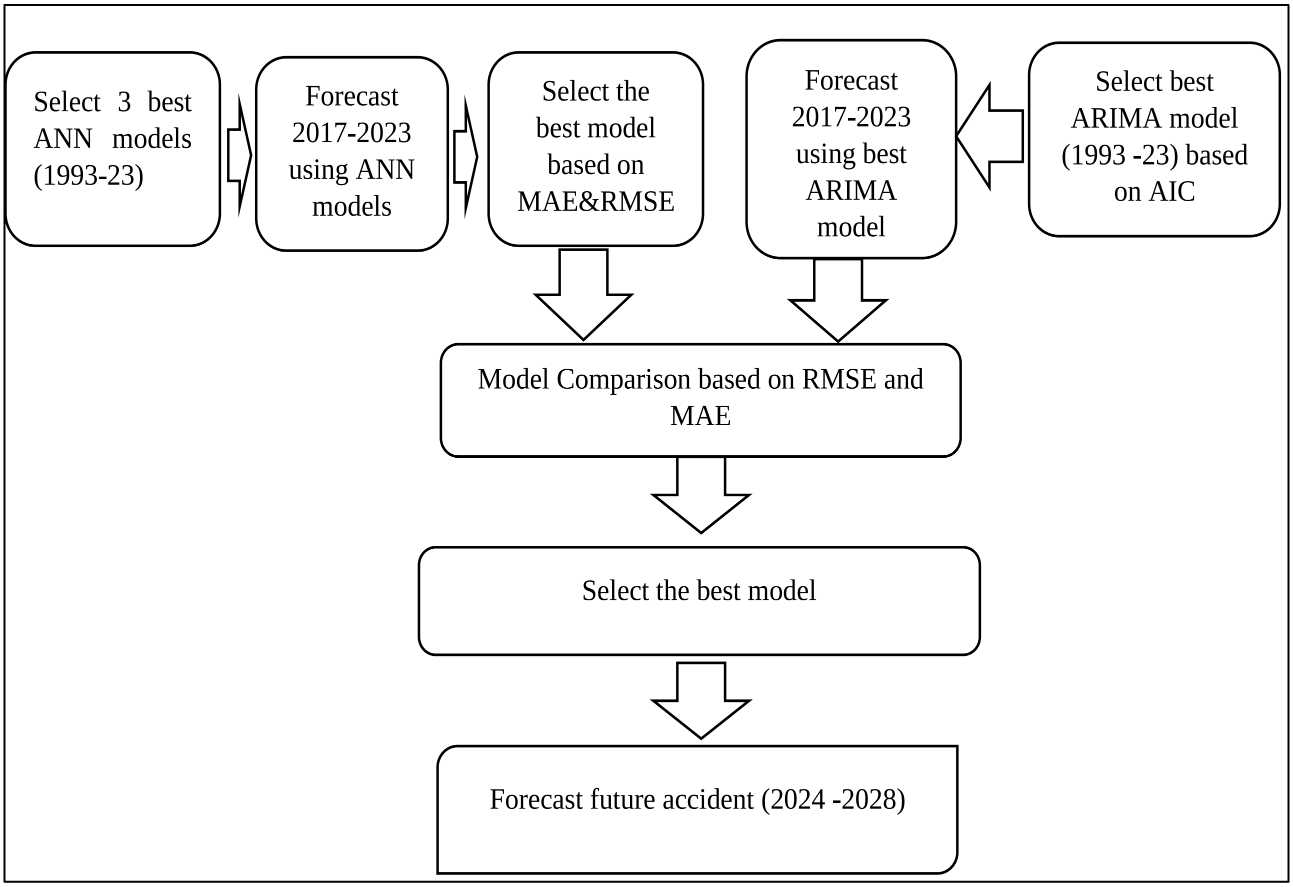

Select 3 best ANN models (1993-23)
Forecast 2017-2023 using ANN models
Select the best model based on MAE&RMSE
Forecast 2017-2023 using best ARIMA model
Select best ARIMA model (1993 -23) based on AIC
Model Comparison based on RMSE and MAE
Select the best model
Forecast future accident (2024 -2028)

Figure 2.4.2: Model building and selection procedure.

2.5 Chapter Summary

In conclusion, the chapter reviewed the theoretical and empirical literature on motor vehicle accidents. The research gap in terms of time series analysis was identified and a proposed conceptual model was presented to guide the analysis and interpretation of the data. The research methodology and data collection techniques will be discussed in the next chapter.

CHAPTER 3: RESEARCH METHODOLOGY

3.0 Introduction

Strategies employed in achieving the study's goals are the main topics of this chapter. This methodology serves as a framework for the tactics employed in this study in addition to research methods. It highlights the research design, methods for data collection, sampling techniques, targeted population and data sources for the study. The methodology leads and navigates how this study was carried.

3.1 Research Design

A quantitative research design was employed in this study. The researcher selected predictive research design, which is a quantitative research technique that makes predictions about future events based on past data. It is a crucial instrument for investigators as it furnishes directives for carrying out the investigation and delineates the strategy for collecting, quantifying, and interpreting data. Quantitative research involves gathering and interpretation of numerical data to establish relationships and make generalizations. This study focused on one variable which is the number of accidents (Y_t) at time t .The used quantitative research design is crucial because it permits systematic and rigorous data collection and analysis, offers a framework for testing hypotheses and drawing conclusions from the data, and permits results to be replicated by other researchers and compared across studies.

3.2 Data Sources and Data Collection

The Zimbabwe Traffic Safety Department, which keeps track of car accidents in the nation, is the study's main source of data. ZIMSTAT and Zimbabwe Parliamentary report of the Portfolio Committee on Transport and Infrastructure Development on The Causes of Road Carnage are the alternative data sources. The Traffic Safety Council of Zimbabwe' (TSCZ) secondary data for all recorded motor vehicle accidents from 1993 to 2023 was used in the study**.**

Justification of Data Source: Secondary Data (TSCZ)

Since TSCZ is an established government body tasked with maintaining road safety in Zimbabwe, the data was gathered from them as they have the authority and experience to collect data that is authentic and trustworthy. Secondary data was chosen because it saves time and easy to access as compared to primary data. So as a result, the researcher just downloaded the data from TSCZ website. However, secondary data may be outdated or incomplete.

3.3 Targeted Population and Sampling Procedures

The study included all motor vehicle accidents that occurred in Zimbabwe from 1993 to 2023. This study's sample size was 31 yearly data points. The study applied purposive sampling technique due to its ability in identifying long-term trends and patterns of motor vehicle accidents (trend analysis).

3.4 Research instruments

Methods or instruments are created to collect and analyse data as part of the research process. In this study a laptop was used to download motor vehicle accidents data from 1993 to 2023 from the TSCZ website. Quantitative research instruments such as Microsoft packages (excel) to input and view data and R-Programming 4.3.2 for the analysis and creating graphs of collected data were used.

3.5 Description of Variables and Expected Relationships

WHO (2018), defined road traffic accidents as an event or occurrence that involves at least one moving road vehicle. It can be described as an unanticipated incident of a car crash that has the potential to cause injuries, fatalities, and property damage. The following symbols and their meaning for easy data analysis were used.

Table 3.5: Description of variables and expected relationships

Variable	Variable symbol	Expected relationship
The number of motor vehicle accidents at time t	Y_t	It was expected that Y_t has a positive relation with the occurrence of motor accidents at time t.
		As proposed by Mutangi (2015).

3.6 Data Analysis Procedures

Data cleaning is an essential stage of data analysis as it seeks to organize unorganized data. Wickham (2014) defined a tidy dataset as a dataset that contains characteristics like, each observation comprises of a row, a variable from the column and an observational unit produces a table. The accidents data was presented in a tabular format with each observation arranged in rows and columns. The data was cleaned and duplicate records were eliminated hence making the data ready for analysis. The cleaning process was done through the following stages, data importation, merging data sets, rebuilding missing data, deduplication, verification and data enrichment. After data preparation, the Box-Jenkins methodology was applied to the data. The model was identified, selected and parameters were estimated. Model diagnosis was done as well as forecasting. On Neural Network methodology, the processed data was normalised, split into training and testing sets and model was identified (architecture of a model). The model was trained, selected and used for forecasting. The models were then compared based on performance metrics.

3.7 The Box Jenkins Methodology

It is a systematic approach to finding, evaluating and using the ARIMA model. This model performs better for a minimum of thirty observations. Steps required to complete ARIMA model include model recognition, parameter calculation and diagnostic checking. This was done in order to determine model adequacy based on historical data analysis (Montgomery, Jennings, and Kulahci, 2015).

3.7.2 Model Identification.

The model was first checked for stationarity using a run sequence plot. Time series data in a run sequence display should be consistent. Box and Jenkins recommended a differencing process to render non-stationary data into stationary. The model was identified through the use of PACF and ACF plots. A function in R programming called auto-Arima automatically chose the best among a collection of ARIMA models. After that, the Akaike information criterion with a correction (AICc) and the maximum likelihood function were to look for the optimal ARIMA model. Autocorrelation function (ACF) can be calculated using the formular below

$$P_k = {}^{Y_k}/_{Y_0} = \frac{Covariance\ at\ lag\ k}{variance} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (3.1)$$

PACF is calculated by the formular below

$$\varphi_{kk} = Corr(Y_t, Y_{t-k}|Y_{t-1}, Y_{t-2}, \ldots\ Y_{t-k+1}. \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.2)$$

Differencing

Converting a nonstationary series into stationary known as differencing. The researcher differenced data once to make it stationary. The process of taking sequential data differences is of prime importance in frequent differencing. Subtracting the values of two consecutive observations on a time series is the easiest way to figure out the initial data difference. If the initial data has n observations $(Y_1, Y_2, Y_3 \ldots Y_n)$, the first differenced data became n-1 observations $(X_2, X_3 \ldots X_n)$ where $X_2 = Y_2 - Y_1,\ X_3 = Y_3 - Y_2 \ldots X_n = Y_n - Y_{n-1}$

Generally, $X_t = \Delta Y_t = Y_t - Y_{t-1}$

$$Z_t = \Delta X_t = \Delta^2 Y_t = \Delta(\Delta Y_t) = \Delta(Y_t - Y_{t-1}) = \Delta Y_t - \Delta Y_{t-1} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.3)$$

$$= (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}).$$

3.7.3 Model Selection Criteria

In time series analysis, the best-fitting model can be chosen using a variety of factors. Performance evaluation metrics like the Mean Squared Error (MSE), Bayesian Information Criterion (BIC), and Akaike Information Criterion (AIC) were used. These standards aid in evaluating the complexity and goodness of fit of various models. The following formula is be used to determine the AIC.

$$AIC = 2K - 2In(L) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. \quad (3.4)$$

Where K represents a parameter count, and L is the maximum value of the likelihood function for the model. The decision rule selects a model with a minimum value of AIC. (West Arsi, 2023).

Bayesian Information Criterion is given by the following formular

$$BIC = -2Log(L) + klog(n) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (3.5)$$

3.7.4 Model / Parameter Estimation

According to Montgomery, Jennings, and Kulahci (2015), parameters of a model that has been roughly identified is estimated using the least squares method. Two fundamental methods for fitting Box-Jenkins models are likelihood maximisation and nonlinear least squares. Once the values of p, d and q are obtained, the greatest likelihood estimate is frequently the recommended course of action and back casting can apply when estimating the initial residuals.

3.7.5 Diagnostic Checking.

The appropriateness of the fitted model was evaluated using residual analysis from the AR and MA models. Residuals of an ARIMA model were characterized by the white noise and a stationary distribution. A residual scatter plot with a rectangular shape and no trends indicates that the model is adequate. For the residuals to provide the best feasible fit for the data, certain presumptions must be true. If the residuals of the Box-Jenkins model satisfy the assumptions, this can be ascertained by statistical visualizations of the residuals. Another choice is to examine the Box-Ljung statistic's value.

The model was diagnosed for the following:

1. **Normality:** normality was assessed through the use of histograms, normal quantile-quantile plot and density plot. Alternative methods include Shapiro-Wilk and Kolmogorov-Smirnov test.
2. **Independence**: ACF and PACF analysis was used to assess for independence. The plots aid in determining the order of AR and MA components. The aforementioned plots display the relation between a time series and its lag values. The Durbin-Watson test is an alternative for determining whether the residuals have positive autocorrelation.

An ARIMA model for motor vehicle accidents in Zimbabwe was represented by the equation below

$$Y_t = \mu + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \theta_1 Y_{t-p} + \theta_1 \varepsilon_{t-1} \ldots\ldots (3.6)$$

The model assumes that a time series Y_t follows a stationary process after differencing. The coefficients φ and θ represent the strength and direction of the dependencies, respectively. The error term ε(t) accounts for the unexplained variation in the time series.

3.8 Artificial Neural Networks Methodology

1. **Data processing**

Data pre-processing comes first in ANN design, before model fitting. Pre-processing is the process of data coding, enrichment, and cleaning that includes handling missing data and compensating for noise. Prior to feeding data into the neural network, the data was normalized to avoid the saturation of hidden nodes. Data normalization techniques like min-max, sigmoid and Z- score were applied to data pre-processing. Min-max normalization technique was applied first on input value so as maximize the neural network model's convergence rate (Mapuwei et al, 2020). One can generalize the minimum-maximum criterion formula by using:

$$y'_t = \frac{Y_t - Min\ (y_t)}{Max\ (y_t) - Min\ (Y_t)} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots.. (3.7).$$

where y_t is the number of accidents at given t, y'_t represents normalized, Min (y_t) and Max (y_t) the highest and lowest values, respectively, of the variable y_t over the range of data.

2. **Training and Testing Set**

A process of creating a network model starts once the data has been processed. Training and testing sets were created from processed data. The network model was developed using the first set, also known as the model-building set or training set. The model's forecasting accuracy is assessed using the second set, often known as the testing or prediction set. Typically, the training set receives higher percentages whereas the testing set receives lower percentages (Yaseen et al, 2016). The training and testing sets consisted of twenty-four (24) and seven (7) observations, respectively which is the 80:20 % ratio. It is of prime importance to note that the forecasts are more accurate when the test set is smaller in length**.**

3. **Model Architecture**

The number of hidden, output and the number of neurons in all layers were used to determine the model architecture. Trial and error process was used to determine the optimal number of layers (Mai et al, 2021). The architecture is written as follows:

$$I - (H_1, H_2, H_3 \dots, H_n) - 0 \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3.8)$$

Where I denotes input size, H_n the hidden unit count, and O neuron count in output layer. 1-(10,5)-1 is an ANN architecture with 1 input layer, 2 hidden layers (with 10 neurons in first layer and 5 5 neurons in second hidden layer) and 1 output layer respectively.

4. **Training a Neural Network**

The process involves determining weights and quantity of neurons in each layer of the network. The most well-liked and frequently applied network learning technique is backpropagation, sometimes known as backward error propagation or backprop. A rule called backpropagation extends the gradient descent technique and was used to modify the weights in the hidden layer of an ANN. It gives the change Δw_{jk}in the weight of the connection between neurons j and k at iteration i as

$$\Delta w_{jk}(i) = -\alpha \frac{\partial E}{\partial w_{jk(i)}} + \mu \Delta w_{jk}(i-1) \text{………………..……… (3.9)}$$

where α is called the learning coefficient, $\Delta w_{jk}(i-1)$the weight change from the iteration before it (Mai et al, 2021).

By guaranteeing a maximum drop in the error function, the learning coefficient quickens the rate of convergence. Convergence is incredibly slow if it is too tiny, particularly if the error function is too big, it won't converge. Because it reduces abrupt fluctuations, the momentum coefficient acts as a lowpass filter, which tends to aid in convergence. The weight shift is subjected to smoothed averaging while avoiding local minima. The training data logarithm determined the starting number of hidden neurons, which were raised during the neural network's training process. Logistic and linear activation functions were used in the hidden and output layers.

Neural Network Model Selection

According to Cui et al (2022), the number of neurons and hidden layers in the selection process was systematically changed to produce the most accurate models. Neural networks without hidden units are the same as linear techniques for statistical forecasting. Hidden units map the input and the output variable, find patterns in the dataset, and provide neural networks with the nonlinearity feature. Top three models were selected through the use of performance evaluation metrics like the RMSE and MAE. Performance measures were computed using values in the test set and predicted values from the training set. The best model is used to compute forecasts from 2024 to 2028.

Multilayer Perceptron artificial neural model for the number of motor vehicle accidents in Zimbabwe is written as follows:

$$y = f\big(w_{kj} f\big(w_{ji} \dots f(w_{21} f(w_{10} x_1 + \beta_{10}) + \beta_{21}) + \dots + \beta_{kj}\big) + \beta_k\big) \dots \dots \text{(3.10)}$$

Where:

- $x_1, x_2 \dots, x_n$ are the input features.
- $w_{10}, w_{21}, \dots, w_{kj}$ are the weights for neuron connections.

- $\beta_{10}, \beta_{21,} \ldots \beta_k$ represents biases for each neuron in the network.

- $f(\,)$ is the activation function.

3.9 Model Comparison / Forecasting Accuracy

A far more complex item, like the economy, can have its future predicted via forecasting, or a simpler entity, like a traffic accident, might have its future predicted for next year or five years. Road traffic accident forecasting analyses both past and present movements in a time series statistically to obtain information about possible trends for future movements. The two models were compared using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to check their performance. Alternative metrics include Mean Absolute Percentage Error (MAPE) and forecast plots. MAPE (Mean Absolute Percentage Error) measures the accuracy of fitted model and the accuracy is expressed as a percentage. It can be calculated using the following formular

$$MAPE = \sum_{t=0}^{n} \frac{\left|\frac{Y_t - \widehat{Y_t}}{Y_t}\right|}{n} * 100 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3.11)$$

Where$Y_t = the\ actual\ value, \widehat{Y_t}\ the\ forecasted\ value\ and\ n\ the\ number\ of\ observations$

The Root Mean Square is calculated using the following formular

$$RMSE = \sqrt{mean({\varepsilon_t}^2)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3.12)$$

Also, the mean absolute error is calculated by the following formular

3.10 Ethical Considerations.

- Transparency in data sources
- Harmful representations were avoided

3.11 Chapter Summary

The research methodology used was described in this chapter. This was then followed by data analysis discussions, which will be covered in the next chapter. The next section focuses on how the data was presented, interpreted, and analysed

CHAPTER 4: DATA PRESENTATION, ANALYSIS AND DISCUSSION

4.0 Introduction

The procedure for analysing data and data interpretation of the study are the main topics of this chapter. The study evaluates the effectiveness of Artificial Neural Networks (ANN) and ARIMA in estimating motor vehicle accidents in Zimbabwe from 1993 to 2023. R programming software was used for data analysis and visualization in this study.

4.1 Summary Statistics

Table 4.1: Summary Statistics

Minimum	16904
Maximum	78481
Range	61577
Sum	1188650
1st Quartile	27541
Median	39841
3rd Quartile	46684
Mean	38343.55
Sample Variance	2.09E+08
Standard Deviation	14463.59
Kurtosis	0.432149
Skewness	0.408696
Count	31

Table 4.1 shows that overall observations were 31 from 1993 to 2023. It was observed that the minimum number of motor accidents recorded is 16904 occurred in 2008. A maximum number of motor vehicle accidents is 78481 recorded in 2003. Data distribution is measured in quartiles. The standard deviation (14463) measures dispersion from the mean (38343). Since the kurtosis and the skewness are positive, it implies that the number of accidents is increasing.

4.2 Pre-tests /Diagnostic tests

The time series plot for the total number of motor vehicle accidents from 1993 -2023 presented in Figure 4.2.1 below was carried out to determine whether the data was stationary or not prior to conducting any statistical tests. The plot shows a consistent pattern from (2000 -2002) and 2014 -2017, a sudden rise in 2003 and a trickle around 2006 and once more a steady rise in the number of accidents was recorded from (2008 – 2017) until a sharp increase in 2018 followed by a drop in 2020 and an uptrend from 2021 to 2023. The time series is non-stationary because the dataset did not exhibit any consistent variation.

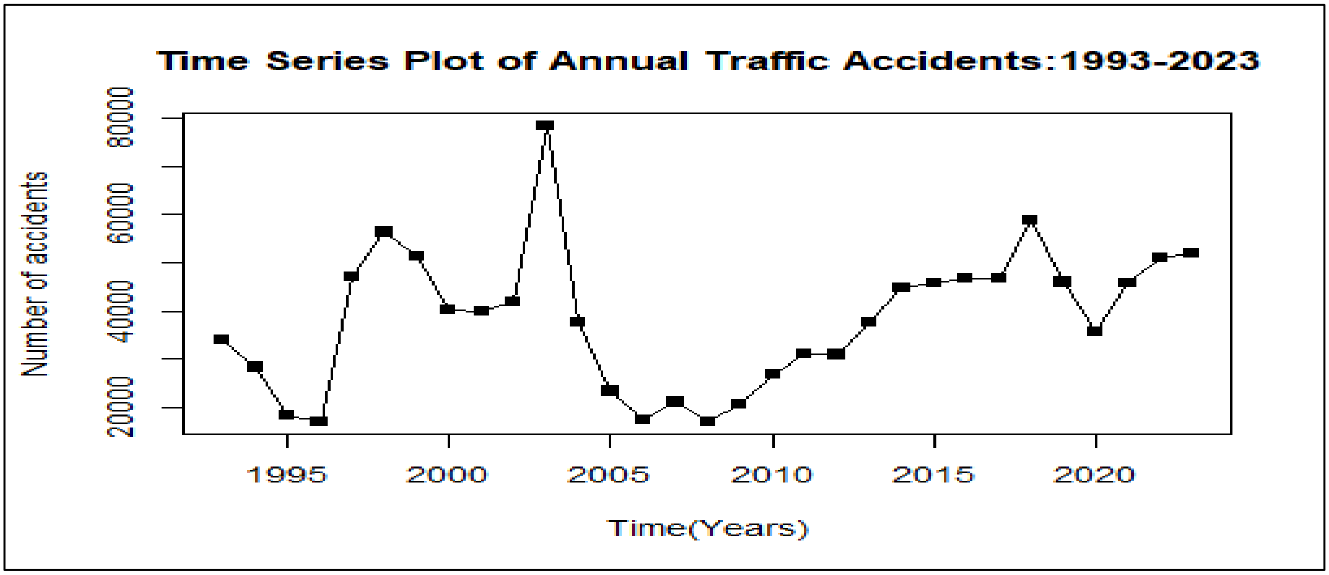


Figure 4.2.1: Time series Plot of Road Accidents from 1993 to 2023

After the visualization of the plot above data was divided in an 80:20 ratio into training and testing sets, yielding 24 training observations and 7 testing observation. The researcher then moved forward with the time series data pre-tests.

4.2.1 ADF Test for Trend Stationarity on Training Data

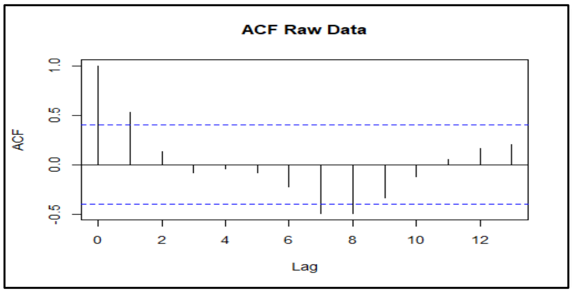
Table 4.2.1: Augmented Dickey-Fuller test

RTA: Data		
Dickey Fuller = -2.3502	lag order = 2	p-value = 0.439

Table 4.2.1 show results from ADF test. This was done to show if the data was stationary. Since p-value is greater than 0.05 level of significance, we can conclude that the data is not stationary. ACF and PACF plots were also used in Figure 4.2.2 to examine stationarity of the data.

4.2.2 Autocorrelation function (ACF) and Partial autocorrelation function (PACF)

If an autocorrelation function plot does not die off quickly, it implies non stationarity in the data. Figure 4.2.2 below shows that the data was highly correlated since some lags has been noticed crossing the dotted blue line hence strengthening the idea of non-stationarity.



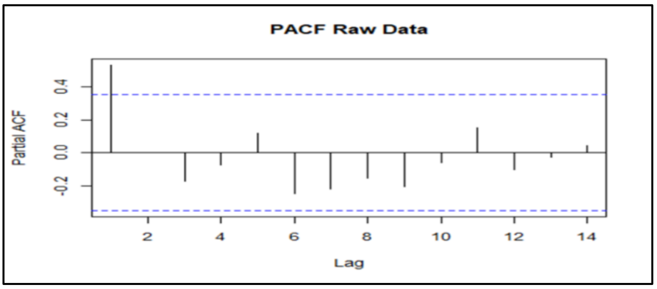


Figure 4.2.2: ACF and PACF Plot for Raw Data

The study proceeded to the next step which involves the process of differencing the time series data. As reflected in previous chapters, non-stationary data has to be differenced as a way of making it stationary and ready for further analysis.

4.2.3 First Difference

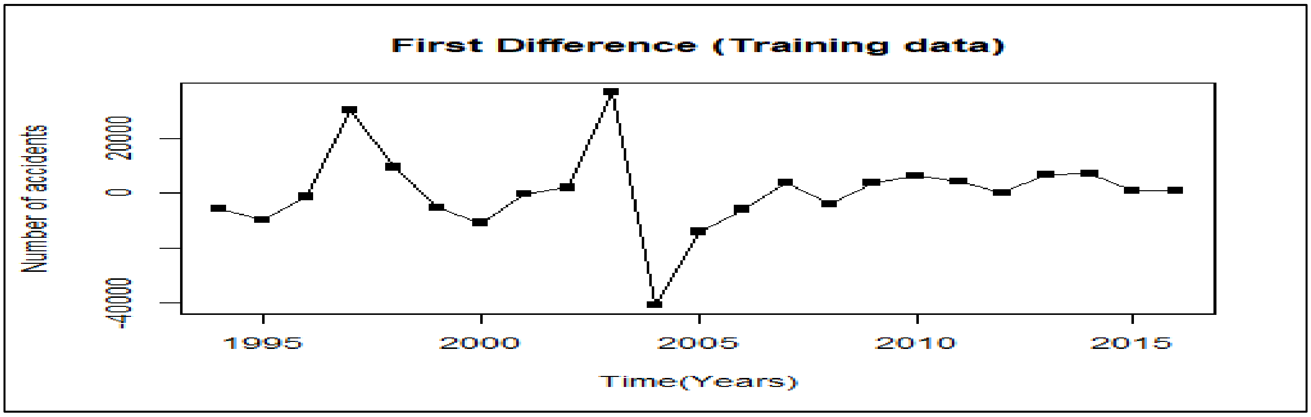


Figure 4.1.3: First Difference Plot for training data

Figure 4.2.3 above is a time series plot for differenced data. It is evident that fluctuations are around zero (0) which implies that over time, the data's average remains unchanged. This simply indicates that following the initial differencing, the data became stationary. Consequently, there was no need for additional differentiation for the ARIMA (p, d, q). A further test for stationarity was conducted in the following table (Table 4.3). Outcomes were that, the p value was less than 0.05 (0.03418 < 0.05) indicating stationarity.

Table 4.2.2: Augmented Dickey- Fuller Test on Differenced Data

	RTA: DATA Frame_d1	
Dickey-Fuller = -3.82	Lag order = 2	p-value = 0.03418

4.3 Model Output/ Results

4.3.1 Model identification

This stage's primary objective is to identify the moving average terms and autoregressive model so that the recognized model, ARIMA (p, d, q), can be obtained. The correlogram of the differenced data was examined and plotted as indicated in Figure 4.3.1.

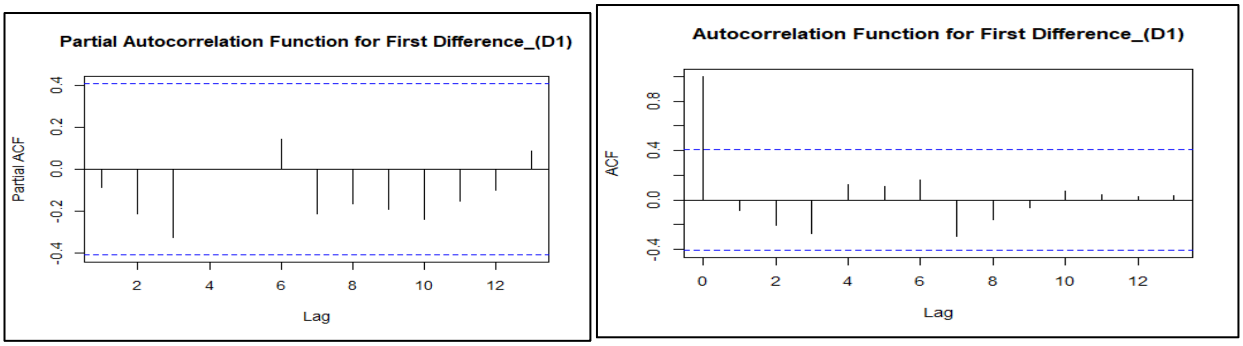


Figure 4.3.1: Results of PACF and ACF from Differenced series

It is evident that both PACF and ACF plots in Figure 4.3.1 did not cut off, p = 0 and q = 0 respectively. Since the data was differenced once, the suggested model is ARIMA (0,1,0) without variations in the seasons. An auto-ARIMA function in R programming was used and a model with minimum AIC value was selected as the best model, which is the AR (1) from Table 4.3.1. Selection of AR (1) model or ARIMA (1.0.0) was due to the underlying data properties such as autocorrelation structure which was adequately captured by first-order autoregressive term or might be due to Occam's razor principle which leads to the selection of a simpler model if it adequately explains the data, and in this case, an ARIMA (1,0,0) model would suffice.

Table 4.3.1 Model Identification Using Auto-Arima

MODEL	AIC VALUES
ARIMA (2,0,2)	531.36
ARIMA (0.0.0)	532.45
ARIMA (1.00)	526.75**
ARIMA (0,0,1)	527.44
ARIMA (2.00)	527.83

4.3.2 Parameter Estimation

This stage involves determining the parameters of the moving average and autoregressive components of the selected model.

Table 4.3.2: Estimated Model Parameters

Series: RTA DATA Frame							
Model: ARIMA (1,0,0)							
AR (1) = 0.5183			Standard error = 0.1689				
Estimated Variance = 167116245			Log likelihood = -260.38				
AIC = 526.75				AICc = 527.95		BIC = 530.29	
Training set error measures							
	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	55.78	12376.99	8982.24	-11.93	28.26	0.981	0.123

Table 4.3.2's AR (1) value demonstrates how the series' previous value(s) affects its current value with an autoregressive coefficient of 0.5183. The model also includes a non-zero mean term of 35929.63, which represents a constant offset added to each value in the series. The standard errors for the AR coefficient (0.1689) represent the average amount of variability or uncertainty in the estimated value of the coefficient and mean term (5030.627) represents uncertainty in the estimated mean value.

4.3.2 .1 Model Structure

$$Y_t = 35929.63 + 0.5183y_{t-1} + \varepsilon_t \qquad \dots\dots\dots\dots\dots.(4.1)$$

With the inclusion of a baseline accident rate, past year's accidents, and random fluctuations, this equation offers a framework for examining and simulating Zimbabwe's time series of motor vehicle accidents. It allows for the investigation of trends, projections, and comprehension of the dynamics of auto accidents in the nation over the specified time frame.

4.4 Model Diagnostic Checking

4.4.1 Stationarity

The residuals' time series plot shows that there is no pattern that is being followed. This means that the residuals are random or the variance remained constant. Therefore, the structure of the residuals indicates that the fitted model is stationary. Therefore, this model can be used for prediction and policy making.

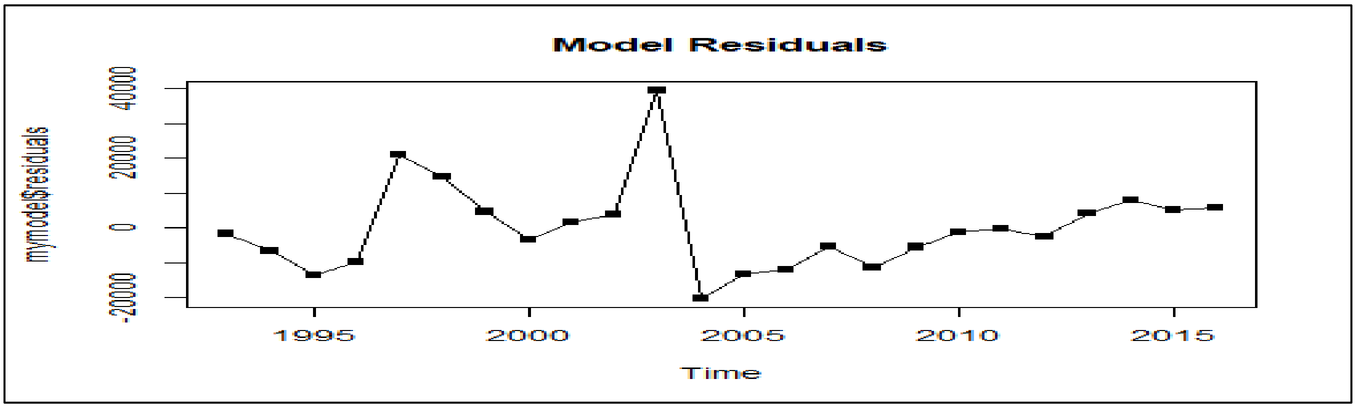


Figure 4.4.1: Model Residuals

4.4.2 Test for Independence

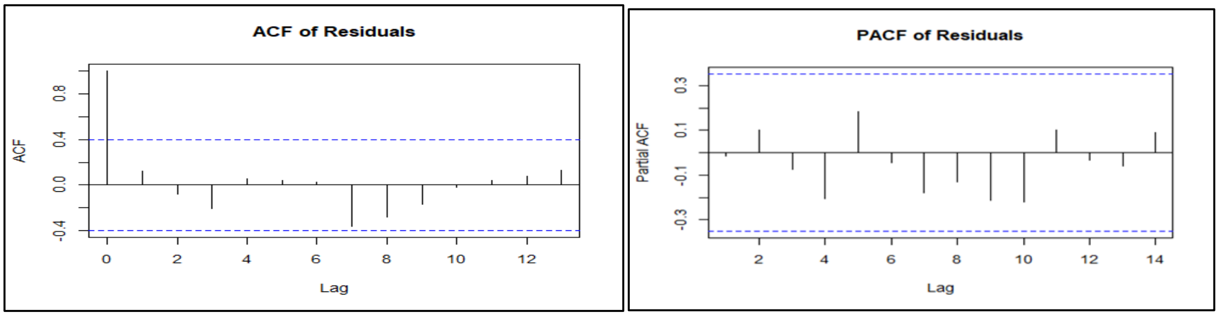


Figure 4.4.2: Test for Independence

The correlograms in Figure 4.4.2 shows no structural pattern hence indicating that there is no serial autocorrelation in the data for residuals. Between lag 1 and lag 14, there are no unique lags that are greater than the threshold. The fitted model's residuals have no association with the variable and are precisely reliant on it.

4.4.3 Test for Normality

Figure 4.4.3 is a histogram of residuals. From its shape we can tell that the residuals follow a normal distribution.

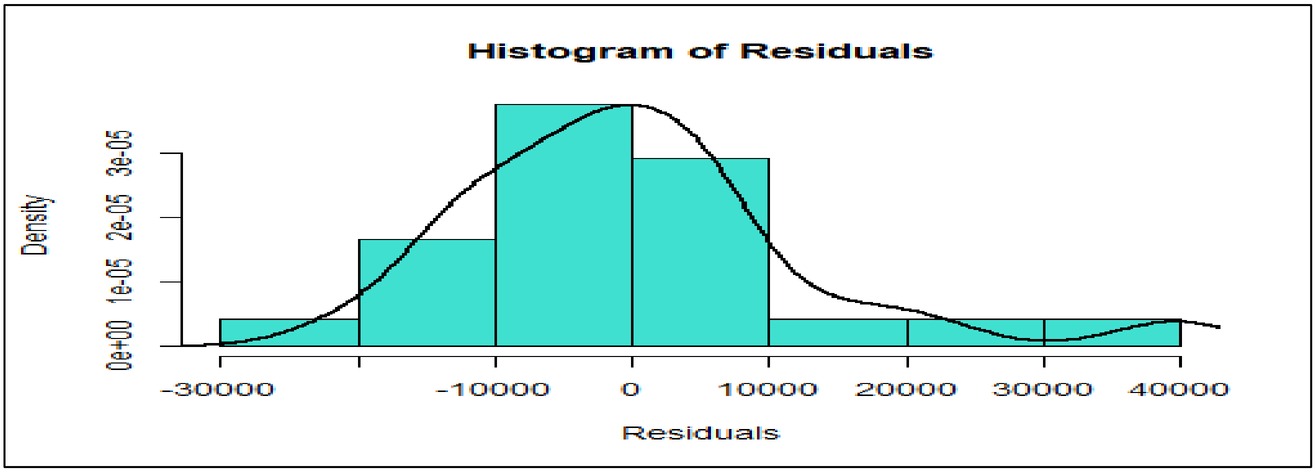


Figure 4.4.3: Histogram of Residuals

Plotting the quantiles yields a normal Q-Q plot which ascertains whether the variables are normally distributed or not. Since most of the plotted points in Figure 4.4.4 below are straight lines, the distribution is shown to follow a normal distribution.

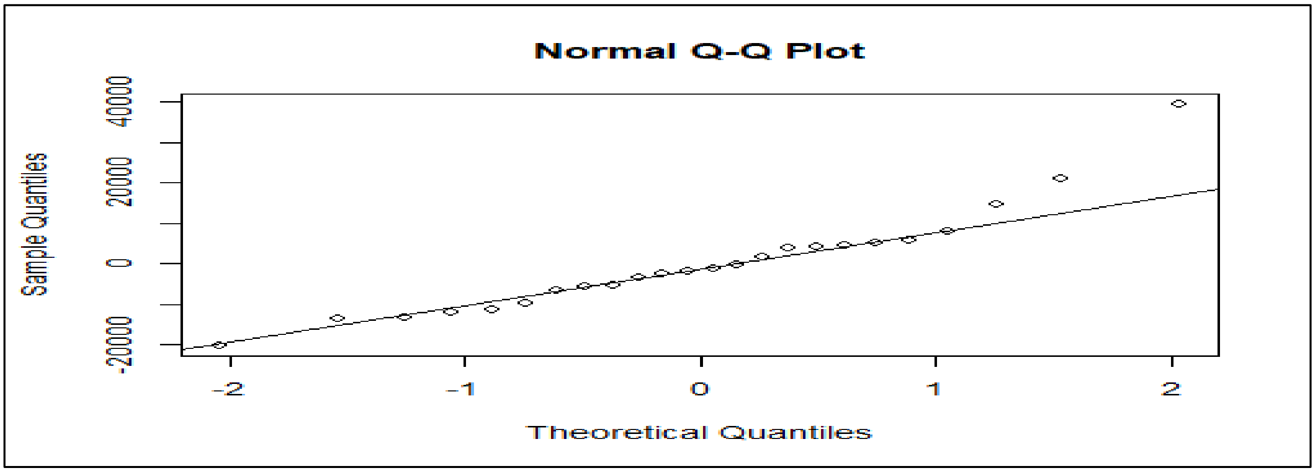


Figure 4.4.4: Q-Q Plot for Normality Test

4.5 Model Validation Tests

4.5.1 ARIMA Training Set

Table 4.5.1: ARIMA Training Set Metrics Table

Training set error measures

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	55.78	12376.99	8982.24	-11.93	28.26	0.981	0.123

The effectiveness of a model in forecasting motor vehicle accidents between 1993 and 2023 is evaluated based on training set error measures, the mean error (ME) is 55.78, suggesting that average predictions are marginally higher than actual values. The Root Mean Square Error (RMSE) of the model presented above is 12376.99 indicating overall accuracy. Additionally, the Mean Absolute Error (MAE), MPE and MAPE of 898.24, -11.93 and 28.26 respectively suggest the model's average predictions below actual values.

4.5.2 Modified Ljung-Box

ARIMA model was validated using the Ljung test as follows

H0: The time series does not exhibit serial autocorrelation.

H1: The time series exhibits serial autocorrelation.

Table 4.5.2: Ljung-Box Results for Serial Autocorrelation Test

data: RTAModel$residuals

Chi-Square	1.9927	11.06	12.56	13.87
DF	5	10	15	20
P-Value	0.8502	0.3529	0.6365	0.8373

Table 4.5.2 shows that p-values are greater than 0.05 which implies lack of serial autocorrelation in the model and this results in the acceptance of H0 on the fitted model.

4.5.3 Forecasting

Table 4.5.3: Predicted ARIMA values for annual forecasts (Training data)

Year	Forecast	Lo 95	Hi 95
2017	41502.6	16165.472	66839.73
2018	38818.46	10279.903	67357.01
2019	37427.17	8087.961	66766.38
2020	36706.02	7155.392	66256.64
2021	36332.22	6725.049	65939.39
2022	36138.04	6516.123	65760.81
2023	36038.04	6411.619	65664.46

In sample forecasting using ARIMA model was computed in Table 4.6.1 The forecasts show a decreasing trend in motor vehicle accidents from 2017 to 2023 based on the training data which is inconsistent with the actual data.

4.6 Multilayer Perceptron Neural Network Model.

4.6.1 Selecting the Best Neural Network Model.

The study systematically selected several models with different architectures beginning with the one having fewer hidden units. MLP models were separated according to the number of hidden layers. Trial and error method was employed in the selection of inputs. Three MLP models in table 4-8 were generated and performance metrics were used to evaluate the models.

Table 4.6.1: MLP Neural Network Model Identification

Model	Structure	Testing set	
		RMSE	MAE
1	1-(10)-1	7631.466	6165.056
2	1-(10,5)-1	7428.460*	5762.852*
3	1-(10,9)-1	7459.050	5820.136

The table above shows the results from Multilayer Perceptron model identification. Of the three selected MLP models 1-(10)-1, 1-(10,5)-1 and 1-(10,9)-1, the one with minimum MAE and RMSE values has been considered the best model for the data. The model with architecture of 1-(10,5)-1 having MAE of 5762.852 and RMSE of 7428.460 was considered the best ANN model for the number of motor vehicle accidents data.

Table 4.6.2: In sample Forecasts of MLP models

Year	**Actual**	**MLP 1**	**MLP 2**	**MLP3**
		1-(10)-1	**1-(10,5)-1**	**1-(10,9)-1**
2017	46681	48912.12	48452.72	48699.78
2018	58739	48408.63	47591.67	47870.12
2019	45920	49008.87	48117.07	48522.15
2020	35560	48463.42	47872.27	47957.39
2021	45791	48906.33	48013.95	48483.13
2022	51107	48572.21	47664.62	48030.29
2023	51924	48906.22	48116.89	48452.90
	MAE	6165.056	5762.852*	5820.136
	RMSE	7631.466	7428.460*	7459.050

The table above shows the results of different forecasts or predictions from the selected MLP models versus the actual values of the testing data set from 2017 to 2023.It was observed that the same model that has the lowest MAE and RMSE, produced forecasts which are quite closer to the actual values as at the period thereby enhancing the suitability of MLP 1-(10,5)-1 in forecasting and simulations.

4.6.2 MLP Model Architecture

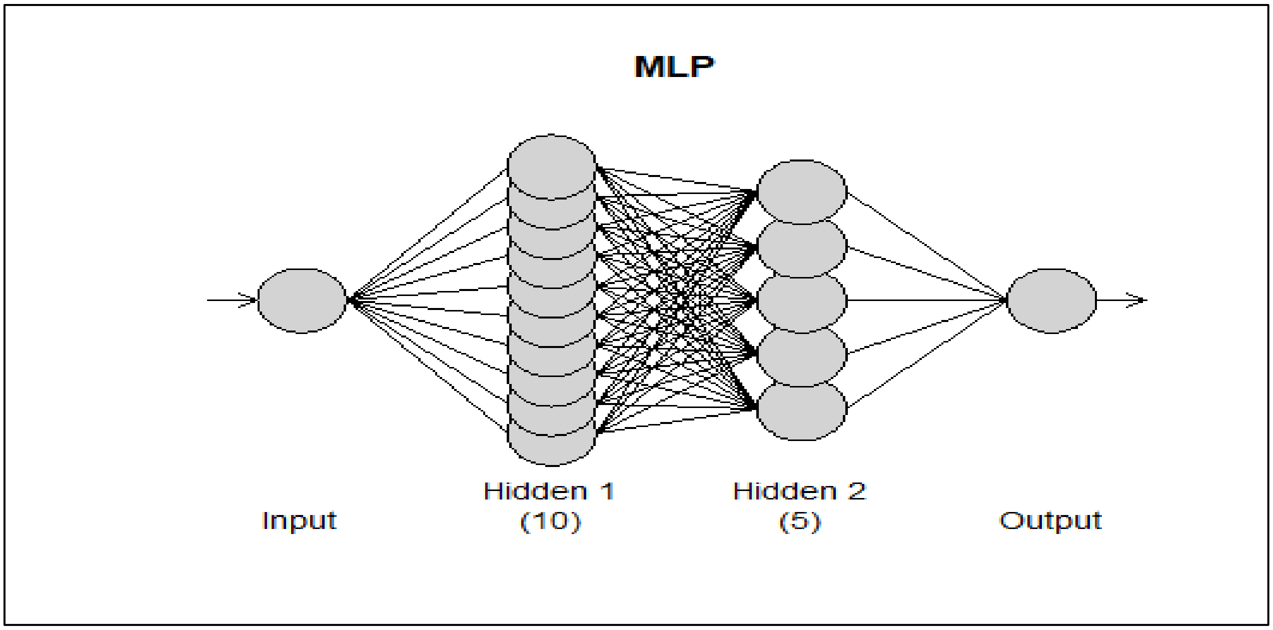


Figure 4.6.2: MLP Model Architecture

Figure 4.6.2 above shows the architecture of the best MLP model from only three selected MLP architectures for motor vehicle accidents data in Zimbabwe from 1993 to 2023. The architecture contains a single input layer, two hidden layers with ten and five neurons respectively and lastly one output layer.

4.6.3 Comparison of MLP versus ARIMA Model

Yearly number of motor vehicle accidents data was compared with the forecasts of ARIMA and MLP as shown in table 4.6.3 below. The graph shows that the pattern of MLP is tending to favour estimated values other than ARIMA which has a specific direction since it projected a linear plot as shown in Figure 4.6.3. Performance measurement metrics like MAE and RMSE were also used to evaluate the models and MLP had MAE and RMSE which were superior than that of ARIMA. This means that MLP performed better than ARIMA model. The MLP was then used to forecast the number of motor vehicles from 2024 to 2028 as shown in Figure 4.6.4.

Table 4.6.3: In Sample Forecasts of motor vehicle accidents by the MLP and ARIMA models

Year	Actual	ARIMA	MLP
2017	46681	41502.6	48452.72
2018	58739	38818.46	47591.67
2019	45920	37427.17	48117.07
2020	35560	36706.02	47872.27
2021	45791	36332.22	48013.95
2022	51107	36138.04	47664.62
2023	51924	36038.04	48116.89
	MAE	8982.242	5762.852*
	RMSE	12376.99	7428.460*

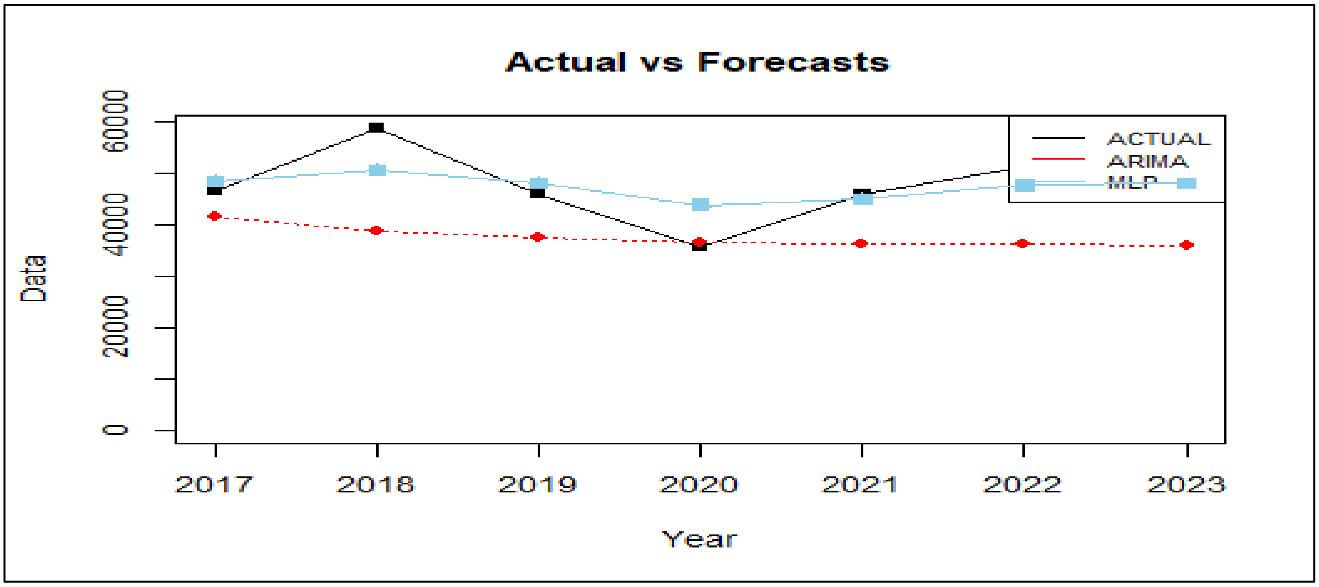


Figure 4.6.3: Performance of ARIMA and MLP Versus Actual Testing Data

Forecasting using the best model MLP 2

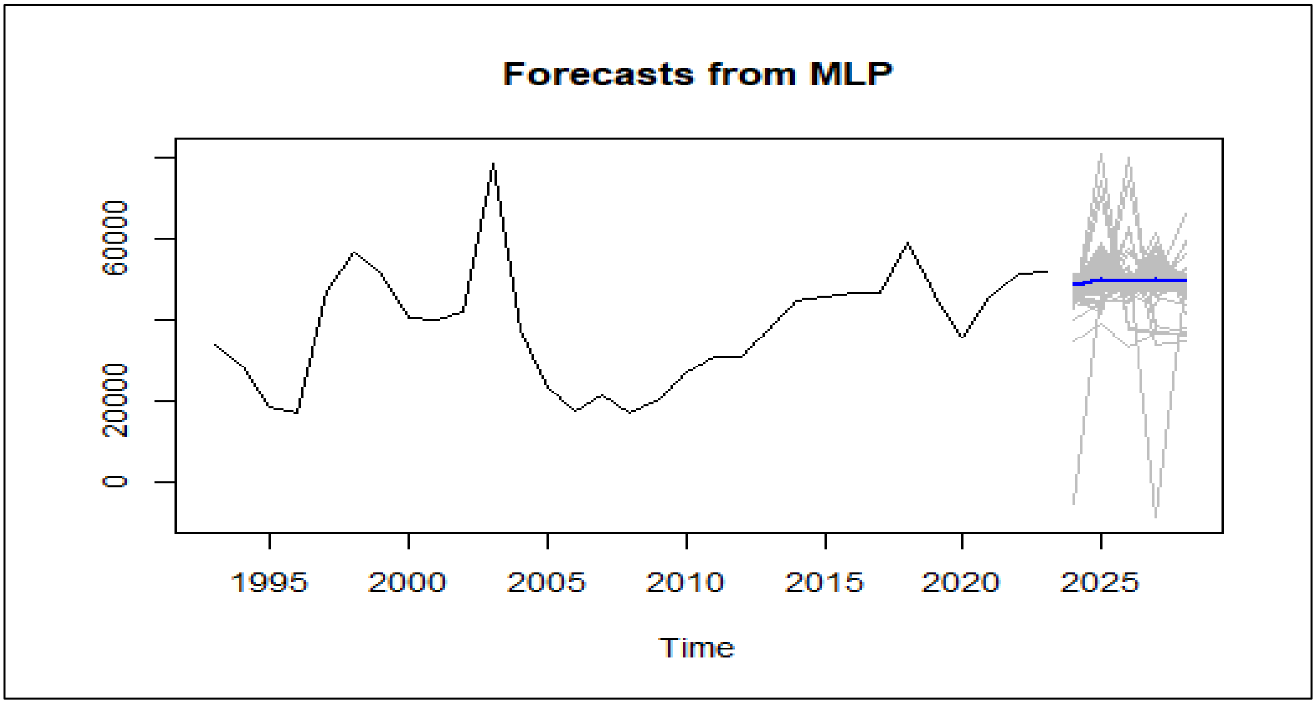


Figure 4.6.4: Forecasting using MLP from 2024 – 2028

Figure 4.6.4 above, resembles the forecasted number of motor accidents from 2024 to 2028 using the best MLP model, the blue line and grey shades indicating confidence interval for the number of motor vehicle accidents. The plot shows a decreasing trend on future motor accidents. Below is a table for MLP annual forecasts.

Table 4.6.4 Out of Sample MLP Forecasts from 2024 - 2028

Year	Point Forecast
2024	48819.75
2025	50012.30
2026	49536.89
2027	49720.53
2028	49630.95

4.7 Discussion of Findings

It can be observed that MLP model with architecture 1-(10,5)-1 performed much better than ARIMA through the use of performance measurement metrics like MAE and RMSE. MLP model's predictions were slightly accurate (more inclined to value estimation) than that of ARIMA which were directional. The MLP plot in forecasting the number of motor vehicle accidents from

2024 up to 2028 shows a decreasing trend in the number future accidents. Overall, there was an increase in the number of motor vehicle accidents from 1993 to 2023 as shown in Figure 4.2.1 However, the forecast curve gives a downward trend from 2024 to 2028. As forecasted by Mutangi, the data for this study shows an increase of motor vehicle accidents from 2015 to 2023. MLP 1-(10,5)-1 performed better than AR (1). The results are in line with what was said in literature by Wannuraw et al in 2023 that ANN performs better than ARIMA.

4.8 Chapter Summary

The chapter outlined presentation and analysis of data using the two proposed models, ARIMA and ANN (MLP). The best performing model was found to be Multilayer Perceptron a type of Artificial Neural Networks. Visualizations done in this chapter shows the trends and forecasts in the number of motor vehicle accidents in Zimbabwe hence providing insights to the stakeholder. The following chapter will dwell much on conclusion and provide detailed recommendations for different stakeholders.

CHAPTER 5: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.0 Introduction

This chapter provides a comprehensive summary, recommendation and conclusions based on the findings of the study comparing ARIMA and Artificial Neural Networks (ANN) for time series analysis of motor vehicle accidents in Zimbabwe from 1993 to 2023. The chapter also highlights the areas for further research that were not mentioned by the researcher in the study.

5.1 Summary of the Study and Findings

The study looked at the comparative analysis of ARIMA and ANN in predicting motor vehicle accidents in Zimbabwe from 1993 to 2023. The study's objectives were to analyse the temporal patterns and trends of motor vehicle accidents, to fit ANN and ARIMA models to the data, compare the performance of the ARIMA and ANN and forecast future accidents from 2024 to 2028. The sample of the study included 31 observations and a quantitative research design was employed. The data was obtained from TSCZ website and other government publication reports. A class of ANN which is the Multilayer Perceptron's performance was compared against ARIMA model.

The first objective of the study is clearly shown in Figure 4.2.1 It shows the pattern and trend of motor vehicle accidents from 1993 to 2023 with an increasing trend. The increasing trend is in agreement with what was said in literature by Mutangi in 2015. Table 4.6.3 is a comparison of MLP and ARIMA in sample forecasts from 2017 to 2023 versus the actual data using performance metrics. A study by Wannuraw in 2023, compared the ARIMA and ANN and results that ANN performed better than ARIMA is in line with the results of the present study.

The optimal model was determined to be an ARIMA with the least AICc values and BIC. Additionally, Ljung-Box statistics were employed to evaluate its level of quality. ARIMA (1,0,0) was determined to be the most effective model for simulating and forecasting the data on motor vehicle accidents by a time series package in R (auto-Arima function). Three MLP models with architecture: 1-(10)-1, 1-(10,5)-1 and 1-(10,9)-1 were developed and the best performing model for simulating and forecasting the motor vehicle accidents was found to be MLP (2) with the architecture 1-(10,5)-1. The suitability and validation of the aforementioned approaches were assessed at various stages. The actual and forecasts were slightly close to match. Statistical metrics like the MAE and RMSE were used to evaluate the models.

Furthermore, the study's outcome of the current work coined that forecasting is feasible using statistical analysis of both current and past time series data as one of the researcher's objectives. The two approaches demonstrated their effectiveness in capturing the patterns and trends. ARIMA model showed a strong performance in short-term forecasting while MLP exhibited superior performance in long-term forecast with complex nonlinear relationships. The performance of ARIMA and MLP was compared. The results revealed that MLP outperformed the ARIMA model in making predictions through the use of evaluation metrics and it showed that ARIMA predictions were directional. The ANN model was then used to make final forecasts of future accidents from 2024 up to 2028 and this shows that all the objectives by the researcher have been met.

5.2 Conclusions

From the findings of the research, it can be concluded that the Multilayer Perceptron (MLP) with the architecture of 1-(10,5)-1 was the best model in predicting future accidents in Zimbabwe. Since there was a slight decrease in forecasted data, it shows that for the next 5 years there will be less accidents. However, it is of prime importance to take note that the model does not reduce or increase the number of accidents in future due to uncertainties but only in showing the trends of motor vehicle accidents in the country.

5.4 Recommendations

To the Driver

Due to the surge in motor vehicle accidents, drivers are recommended to be extra cautious during peak accident hours for example, in rush hour, late nights and weekends. They are also recommended to slow down and drive defensively during rainy weather.

To the Government

The governments should maintain and keep on improving infrastructures especially in high accident areas. Traffic enforcement should be strict so as to lower reckless driving.

To Passengers

They are encouraged to wear seatbelt, avoid distractions like talking to the driver, stay aware and aware of the surrounding and avoid travelling with reckless or inexperienced drivers.

5.5 Areas for Further Research

Subsequent research on this topic ought to investigate the possibility of utilizing hybrid linear-non-linear models for precise motor vehicle accident prediction. Some areas that warrant further research includes the investigation of exogeneous factors like weather and economic indicators on motor vehicles and incorporating them into the forecasting models.

REFERENCES

Ajzen, I. (1991). The theory of planned behavior. Organizational behavior and human decision processes, *50*, 179–211.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.

Brockwell, P. J., Davis, R. A., Brockwell, P. J., & Davis, R. A. (2016). ARMA models. Introduction to Time Series and Forecasting, 73–96.

Çuhadar, M. (2014). Building proper forecast model for daily air passenger demand: a study of Antalya International Airport.

Cui, Z., Wang, L., Li, Q., & Wang, K. (2022). A comprehensive review on the state of charge estimation for lithium-ion battery based on neural network. ***International Journal of Energy Research, 46,*** **5423–5440.**

Du, K.-L., Swamy, M. N., Du, K.-L., & Swamy, M. N. (2014). Multilayer perceptrons: Architecture and error backpropagation. *Neural networks and statistical learning*, 83–126.

Emenike, G. C., & Kanu, C. A. (2017). Drivers Distraction and Road Traffic Crashes in Port Harcourt Metropolis, Rivers State, Nigeria. *International Journal of New Tec****hnology and Research****, 3*, 263322.

Erguzel, T. T., Noyan, C. O., Eryilmaz, G., Ünsalver, B. Ö., Cebi, M., Tas, C., . . . Tarhan, N. (2019). Binomial logistic regression and artificial neural network methods to classify opioid-dependent subjects and control group using quantitative EEG power measures. *Clinical EEG and neuroscience, 50*, 303–310.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice.* OTexts.

Jenkins, G. M., & Box, G. E. (1976). Time series analysis: forecasting and control. *(No Title).*

Khan, S., & Alghulaiakh, H. (2020). ARIMA model for accurate time series stocks forecasting. International Journal of Advanced Computer Science and Applications, 11.

Khashei, M., & Bijari, M. (2010). An artificial neural network (p, d, q) model for timeseries forecasting. Expert Systems with applications, *37*, 479–489.

Kihoro, J., Otieno, R. O., & Wafula, C. (2004). Seasonal time series forecasting: A comparative study of ARIMA and ANN models.

Kuikel, J., Aryal, B., Bogati, T., & Sedain, B. (2022). Road traffic deaths and injuries in Kathmandu. ***Journal of Health Promotion***, *10*, 73–88.

Mai, H.-V. T., Nguyen, T.-A., Ly, H.-B., & Tran, V. Q. (2021). Investigation of ann model containing one hidden layer for predicting compressive strength of concrete with blast-furnace slag and fly ash. Advances in Materials Science and Engineering, 2*021*, 1–17.

Makridakis, S., Assimakopoulos, V., & Spiliotis, E. (2018). Objectivity, reproducibility and replicability in forecasting research. ***International Journal of Forecasting***, *34*, 835–838.

Mankiw, N. G. (2014). One Way to Fix the Corporate Tax: Repeal It. *The New York Times*.

Mapuwei, T. W., Ndava, J., Kachaka, M., & Kusotera, B. (2022). An Application of Time Series ARIMA Forecasting Model for Predicting Tobacco Production in Zimbabwe. ***American Journal of Modeling and Optimization,*** *9*, 15–22.

Mapuwei, T. W., Bodhlyera, O., & Mwambi, H. (2020). Univariate time series analysis of short-term forecasting horizons using artificial neural networks: the case of public ambulance emergency preparedness. ***Journal of Applied Mathematics,*** *2020, 1- 11.*

Mboso, J. (2022). The autoregressive integrated moving average models of the classical Box-Jenkins methods of Time Series Analysis. ***American Journal of Statistics and Actuarial Sciences***, *4*, 18–34.

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). Introduction to time series analysis and forecasting. John Wiley & Sons.

Mutangi, K. (2015). Time series analysis of road traffic accidents in Zimbabwe.

Negesa, L., & Dessie, Y. (2017). Characterization of Road Traffic Accidents on the Road between Harar and Dire Dawa, Eastern Ethiopia: A cross-sectional study. ***East African Journal of Health and Biomedical Sciences,*** *1*, 29–35.

Organization, W. H. (2015). Global status report on road safety 2015. World Health Organization.

Organization, W. H. (2019). Global status report on road safety 2018. World Health Organization.

Organization, W. H., & others. (2020). Road safety.

Qureshi, M., Daniyal, M., Tawiah, K., & others. (2022). Comparative evaluation of the Multilayer perceptron approach with conventional ARIMA in modeling and prediction of COVID-19 daily death cases. ***Journal of Healthcare Engineering****, 2022*.

Ramasubramanian. (2015). Time series Analysis. New Delhi: I.A.S.R.I library avenue. This is the Pre-Published Version. (n.d.)

Ramasubramanian, V., Chandra, H., Iquebal, M. A., Paul, R. K., Pal, S., Dash, S., . . . Basak, P. (2019). ICAR-IASRI Annual Report 2018-19.

Ridhagen, M., & Lind, P. (2021). A comparative study of Neural Network Forecasting models on the M4 competition data. A comparative study of Neural Network Forecasting models on the M4 competition data.

Russell, S. J., & Norvig, P. (2016). Artificial intelligence: a modern approach. Pearson.

Schaffer, A. L., Dobbins, T. A., & Pearson, S.-A. (2021). Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large-scale health interventions. BMC medical research methodology, 21, 1–12.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks, 61*, 85–117.

Shimodaira, H. (2015). Single Layer Neural Networks. *Learning and Data Note*.

Slovic, P. (1988). Risk perception. *Carcinogen risk assessment*, 171–181.

Stock, J. H., & Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. ***The Economic Journal****, 128,* 917–948.

Wannuraw, A. Z., Borhan, N. U., & Zahari, S. M. (2023). Modelling Road Accidents In Selangor: A Comparative Analysis By Using Seasonal Autoregressive Integrated Moving Average (Sarima) And Artificial Neural Network (Ann). ***Journal Of Sustainability Science And Management,*** *18*, 92–109.

West Arsi, E. (n.d.). Time Series Modelling of Road Traffic Accidents in.

Wickham, H. (2014). Tidy data. *J**ournal of statistical softwar**e*, *59*, 1–23.

Kihoro, J., Otieno, R. O., & Wafula, C. (2004). Seasonal time series forecasting: A comparative study of ARIMA and ANN models.

Russell, S. J., & Norvig, P. (2016). Artificial intelligence: a modern approach. Pearson.

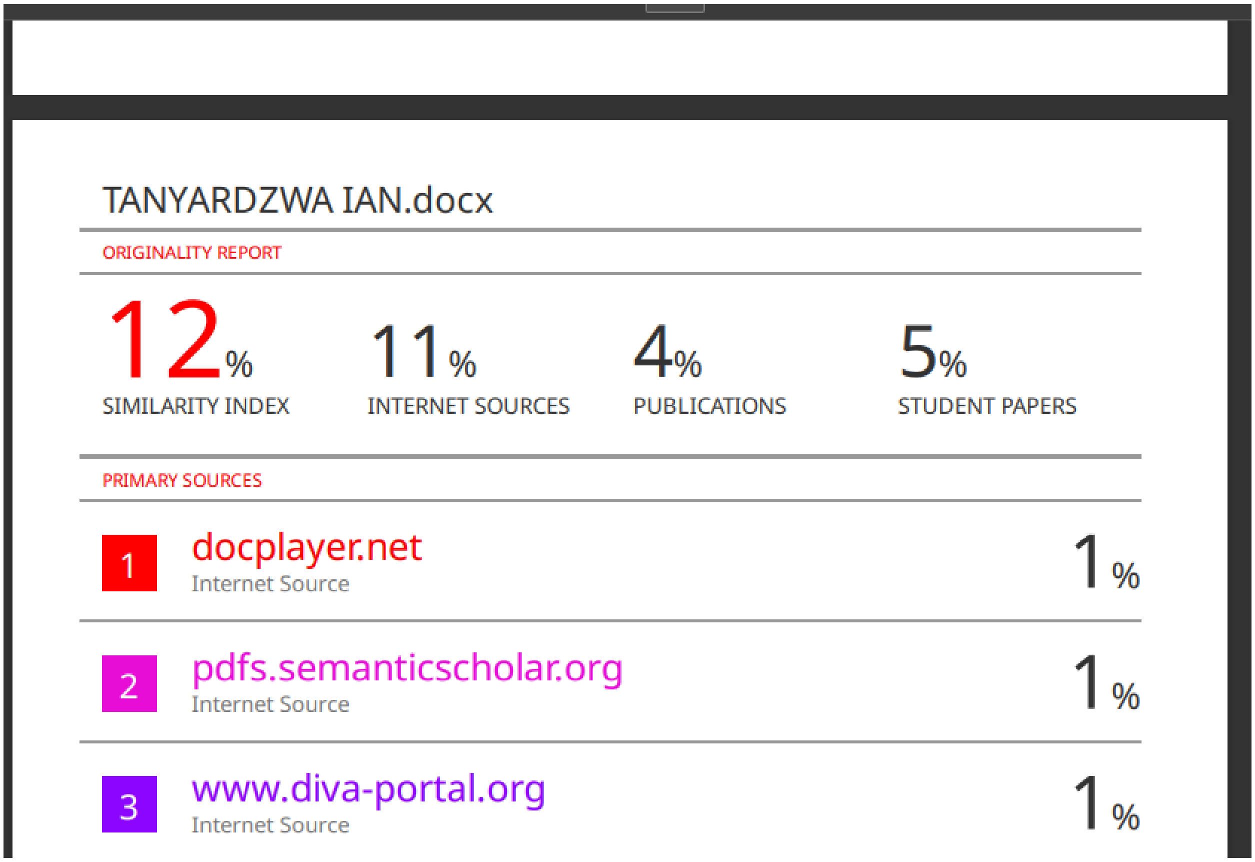
Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J. (2018). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *J**ournal of the American Statistical Associa**tion*.

Xue, D.-m., & Hua, Z.-q. (2016). ARIMA based time series forecasting model. ***Recent Advances in Electrical & Electronic Engineering (Formerly Recent Patents on Electrical & Electronic Engineering)***, *9*, 93–98.

Yaseen, Z. M., El-Shafie, A., Afan, H. A., Hameed, M., Mohtar, W. H., & Hussain, A. (2016). RBFNN versus FFNN for daily river flow forecasting at Johor River, Malaysia. *Neural Computing and Applications, 27*, 1533–1542.

APPENDICES

TURNIT IN REPORT



TANYARDZWA IAN.docx

ORIGINALITY REPORT

12%	11%	4%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	docplayer.net Internet Source	1%
2	pdfs.semanticscholar.org Internet Source	1%
3	www.diva-portal.org Internet Source	1%

ARIMA R CODES

```
#### Step 1: INSTALL REQUIRED PACKAGES

attach(ryt)

library(ggplot2)

library(ggpubr)

library(tseries)

library(forecast)

library(caret)

library(dplyr)
```

```
##### SPLIT THE DATA INTO TRAINING AND TESTING SET USING 80:20 RATIO

train_ratio <- 0.8

train_size <- floor(train_ratio * nrow(ryt))

train_size

w <- ryt[1:train_size, ]

w

print(w,n = 24)

##### COVERT DATA INTO TIME SERIES DATA

class(w)

wt = ts(w$Yt,start = min(w$YEAR),end = max(w$YEAR),frequency = 1)

class(wt)

plot(wt)

##### STATIONARITY TEST

adf.test(wt)

a <- diff(wt,d = 1)

adf.test(a)

plot(a,main = "First Difference (Training data)",type = "o",pch = 15,xlab = "Time(Years)",ylab = "Number of accidents")

adf.test(a)

acf(a,main = "Autocorrelation Function for First Difference_(D1)")

pacf(a,main = "Partial Autocorrelation Function for First Difference_(D1)")

mymodel = auto.arima(wt,ic = "aic",trace = TRUE)

mymodel

summary(mymodel)

####RESIDUAL TESTS / MODEL DIAGONISTIC TESTS

plot(mymodel$residuals,main = "Model Residuals",pch = 15,type = "o")

acf(mymodel$residuals,main = "ACF of Residuals")

pacf(mymodel$residuals,main = "PACF of Residuals")
```

```
hist(mymodel$residuals, freq = FALSE, main = "Histogram of Residuals", xlab = "Residuals", col = "turquoise")

lines(density(mymodel$residuals), col = "black", lwd = 2)

qqnorm(mymodel$residuals)

qqline(mymodel$residuals, col = "black")

####MODEL VALIDATION TESTS

Box.test(mymodel$resid, lag=5, type ="Ljung-Box")

Box.test(mymodel$resid, lag=10, type ="Ljung-Box")

Box.test(mymodel$resid, lag=15, type ="Ljung-Box")

Box.test(mymodel$resid, lag=20, type ="Ljung-Box")

##### FORECASTING

myforecast = forecast (mymodel, level=c (95), h=7)

myforecast

plot(myforecast)
```

MLP CODES

```
#### install required packages

attach(ryt)

library(ggplot2)

library(ggpubr)

library(tseries)

library(forecast)

library(caret)

library(dplyr)

library(nnfor)

library(plotly)

####split data into training and testing sets

train_ratio <- 0.8
```

```
train_size <- floor(train_ratio * nrow(ryt))
train_size
w <- ryt[1:train_size, ]
w
print(w,n = 24)
#### normalize the data
min_value <- min(w$Yt)
max_value <- max(w$Yt)
n <- (w$Yt- min_value) / (max_value - min_value)
class(n)
print(n)
#### convert data into time series data
nt = ts(ryt$Yt,start=1993,end = 2016)
nt
ft <- ts(ryt$Yt,start = 1993,end = 2023)
#### MLP model fitting
mlp.fit = mlp(nt, hd = c(10,5), hd.auto.type = ("valid"), reps = 100, comb =
c("median","mean","mode"))
plot(mlp.fit)
a=forecast(mlp.fit,h=7)
plot(a, main ="Forecasts from 2024 - 2028 by MLP (1-(10,5)-1) Model")
summary(mlp.fit,)
summary(a)
mlp.fit = mlp(ft, hd = c(10,5), hd.auto.type = ("valid"), reps = 100, comb =
c("median","mean","mode"))
plot(mlp.fit,)
summary(mlp.fit,)
b = forecast(mlp.fit,h = 5)
b
```

```
plot(b)

summary(b)
```