Heliyon 9 (2023) e22544

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Time series based road traffic accidents forecasting via SARIMA and Facebook Prophet model with potential changepoints

Edmund F. Agyemang [a,b,c,*], Joseph A. Mensah [c], Eric Ocran [a], Enock Opoku [a,c], Ezekiel N.N. Nortey [a]

[a] Department of Statistics and Actuarial Science, College of Basic and Applied Sciences, University of Ghana, Ghana
[b] School of Mathematical and Statistical Science, College of Sciences, University of Texas Rio Grande Valley, USA
[c] Department of Computer Science, Ashesi University, No. 1 University Avenue, Berekuso, Accra, Eastern Region, Ghana

ARTICLE INFO

Keywords:
Road traffic accident
SARIMA
Facebook Prophet
Potential changepoints
Ghana

ABSTRACT

Road traffic accident (RTA) is a critical global public health concern, particularly in developing countries. Analyzing past fatalities and predicting future trends is vital for the development of road safety policies and regulations. The main objective of this study is to assess the effectiveness of univariate Seasonal Autoregressive Integrated Moving Average (SARIMA) and Facebook (FB) Prophet models, with potential change points, in handling time-series road accident data involving seasonal patterns in contrast to other statistical methods employed by key governmental agencies such as Ghana's Motor Transport and Traffic Unit (MTTU). The aforementioned models underwent training with monthly RTA data spanning from 2013 to 2018. Their predictive accuracies were then evaluated using the test set, comprising monthly RTA data from 2019. The study employed the Box-Jenkins method on the training set, yielding the development of various tentative time series models to effectively capture the patterns in the monthly RTA data. $SARIMA\,(0,1,1)\times(1,0,0)_{12}$ was found to be the suitable model for forecasting RTAs with a log-likelihood value of −266.28, AIC value of 538.56, AICc value of 538.92, BIC value of 545.35. The findings disclosed that the $SARIMA\,(0,1,1)\times(1,0,0)_{12}$ model developed outperforms FB-Prophet with a forecast accuracy of 93.1025% as clearly depicted by the model's MAPE of 6.8975% and a Theil U1 statistic of 0.0376 compared to the FB-Prophet model's respective forecasted accuracy and Theil U1 statistic of 84.3569% and 0.1071. A Ljung-Box test on the residuals of the estimated $SARIMA\,(0,1,1)\times(1,0,0)_{12}$ model revealed that they are independent and free from auto/serial correlation. A Box-Pierce test for larger lags also revealed that the proposed model is adequate for forecasting. Due to the high forecast accuracy of the proposed SARIMA model, the study recommends the use of the proposed SARIMA model in the analysis of road traffic accidents in Ghana.

* Corresponding author at: Department of Statistics and Actuarial Science, College of Basic and Applied Sciences, University of Ghana, Ghana.
E-mail address: efagyemang@st.ug.edu.gh (E.F. Agyemang).

https://doi.org/10.1016/j.heliyon.2023.e22544
Received 14 June 2023; Received in revised form 5 November 2023; Accepted 14 November 2023

Available online 20 November 2023
2405-8440/

1. Introduction

The invention of automobiles brought great relief to humanity; people had little complaints due to fatigue from walking long distances, lateness to work and goods and services could be procured within the shortest possible time [1–3]. The use of automobiles have increased access to remote places and enhanced livelihood due to related research, promoted social and economic interactions geographically and created jobs [4–6]. This not withstanding the continuous use and misuse of automobiles have negatively impacted lives as well as property [7]. The rampant occurrence of road accidents seem to have no cure especially in developing countries unlike many studies pertaining to the control and elimination of diseases as reported in the scientific literature [8]. Therefore, many reseachers have given considerable amount of time and space to studying models that could predict the occurrence of road accidents over the years. A study by [9] measured the effects of randomness, exposure, weather, and daylight to variations of road accidents by using a generalized Poisson regression model based on data from four countries (Denmark, Norway, Finland, and Sweden). Their study concluded that randomness and exposure account for eighty to ninety per cent of the variation in road traffic accidents. On the contrary, road accidents are viewed as deterministic occurrences according to a study by [10]. However, inadequate information makes it uncertain how accidents happen. As a result, in this work, we uphold the notion that road accidents are more random than deterministic, as argued by [9], though the vehicle's driver may have a hand in road traffic accidents.

A road traffic accident is a significant cause of death, injury and a disadvantage or handicap worldwide, both in high-income, low-middle income and low-income countries [11]. [12] asserted that "Road Traffic Accidents (RTAs) manifest when a motor vehicle collides with another vehicle, pedestrian, animal, geographical features, or architectural barriers, potentially leading to injuries, property damage, and fatalities". A host of researchers have commented on the causes of road traffic accidents. [13], among a ton of researchers, attributed the causes to overspeeding, drunk driving, wrong overtaking, poor road network and poor worthiness of many vehicles in the country. Notable causes of RTAs include but not limited to unnecessary speeding, reckless driving, fatigue [14], inadequate experience, traffic rules violation, road surface defects, wrong overtaking, machine failure and defective light [15,16], overloading, poor vision [17,18] among others. It is a robust superstitious belief in Africa and Ghana that witches also cause road accidents, as many converted witches and wizards confess and attest to this fact [19,20]. However, studies have shown that road traffic accidents result from drivers' unethical behaviours [21]. The force behind this could be strongly linked to the inability of the drunk driver to control the vehicle because of sleeping [22]. In addition to drunk drivers, passengers and other drunk road users may not know what happens before, during, and after a road accident. When passengers and other road users behave in such a way, they are prevented from taking action to avoid serious injuries or death. Drivers do what they want and cause accidents that cost lives. The effects of RTAs, apart from causing injuries and death, have also brought about other consequences. Road traffic accidents have been known to cause traumas [23], reduce family or persons involved in the accident's financial position [24], disabilities to people [25,26] and psychological effects [27] among others. The impact of road accidents has gone far to the extent that some people refuse to drive a vehicle again. RTAs are classified as fatal, serious, or minor based on the damage they inflict on human lives and property [28,29].

Ghana, classified as a low-middle income earning country, suffers the most regarding road traffic accidents. RTAs are rising each year on Ghana's roads and have become a significant concern for all and sundry in recent times. Even though the road system network in Ghana is terrible, some drivers have also considerably contributed to the number of fatalities registered on Ghana's roadways each year [30]. Although numerous African nations have made strides in reducing road fatalities, some have encountered challenges in this effort. Ghana, like several others, faces difficulties in effectively addressing this issue. Over the last three decades, traffic injuries and deaths in Africa have increased [31]. In 2005, South Africa, one of the most industrialized countries on the African continent, had seventeen (17) allowed automobiles per 100 citizens and no sign of a decrease in road traffic accident deaths as of now [32]. Forecasting future RTA-related deaths worldwide is difficult, although past patterns might be thought to give a realistic picture of what may occur later. However, a few nations thoroughly veer off from these expectations. Moreover, drifts in numerous parts of the world are inconsistent, and there is a confirmation of an increment in deaths in Africa and Asia/Pacific. Statistics from developing nations are consistent with changes in the total number of road traffic accidents recorded year after year. Annually, an estimated 1.2 million individuals succumb to fatalities resulting from RTAs, and another fifty (50) million are injured. Statistical analysts have forecasted that these descriptive statistics will increase by about 65% over the next 20 years if care is not taken [33–35]. 75% of road traffic deaths came about because vehicles collided with each other in low-earning income countries despite owning only 32% of yearly fatalities for every 10,000 cars around the world, making this claim unbelievable. Globally, an estimated cost of US$ 518 billion is spent on RTAs [36]. The share of the developing countries is about US $100 billion, representing 1 to 3 per cent of their gross national product [37]. These stunning numbers indicate that road traffic accidents happen on all landmasses and in every nation. Numerical and computational approaches such as those suggested by [38–42] can be employed as baseline mathematical models in conjunction with optimization algorithms to help reduce the alarming rate of the occurrence RTAs.

Recognizing the limitations of traditional regression techniques, particularly in handling road traffic accident-related cases due to their reliance on independence assumptions, numerous studies have shifted toward time series methodologies. Approaches such as ARMA, ARIMA, DRAG, state space models, and structural models are favoured for their ability to enhance the forecasting of factors related to RTAs. Models have additionally been utilized to examine injuries and deaths caused by RTAs. Various models have been used to model road accident data. [43] estimated the influence of speed limit modifications on the number of road crashes in metropolitan and provincial interstate thruways in the United States using a structural equation of stochastic modelling technique. [44] analyzed RTAs in Kuwait using an Autoregressive Integrated Moving Averages (Box Jenkins) model and compared it to Artificial Neural Networks (ANN) to predict RTA deaths in Kuwait. The study found that ANN was superior if there should arise an occurrence of long-term series without regular variations of accidents. Several researchers have used collision prediction models to model RTAs in various regions of the world. However, due to differences in numerous parameters in different geographical locations, it becomes challenging to apply models that have worked elsewhere in the globe to data gathered from other parts of the world [45]. In Ghana, there has been minimal statistical modelling of RTAs. This problem stems from the unavailability of data acquisition on road accident cases by the authorities in charge of road accident data. Considering the rate at which RTAs are increasing annually in Ghana, there is a need for this study. As a result, statistical analysis of the Madina-Adenta highway RTAs is required to determine the validity or falsity of current literature on RTAs in Ghana. When seasonal patterns in the road traffic accident data are validated, statistical models such as SARIMA and FB Prophet would be employed to fit a model to the RTAs data for improved prediction and decision-making. The education and research department of the National Motor Transport and Traffic Unit (MTTU) has used descriptive statistical techniques and charts for reporting road traffic accidents in Ghana over the past few years. This method's notable drawback lies in its failure to provide essential estimates of road accident occurrences, injuries, and fatalities in Ghana, hampering the ability of the National Road Safety Commission (NRSC) and Motor Transport and Traffic Unit (MTTU) stakeholders to make informed projections. Utilizing time series models like SARIMA and FB Prophet is, therefore, deemed crucial in addressing this knowledge gap.

A significant limitation to researchers in RTA research is the inability to obtain data on people who suffer from road accidents. In most parts of the world, people at an accident scene may fail to report the incident to the police for records to be taken, or if it is reported to the police, they fail to keep the records. Additionally, the emergency unit of various hospitals refuses to keep records of RTA victims once they are admitted [46]. In Ghana, accurate data on RTA cases are usually hard to come by. The information is inadequate even if acquired, mainly because not all accidents are reported to the police for records to be kept [47]. Furthermore, the police may have neglected to complete some of the accident report forms on RTAs submitted to them. However, other researchers who have utilized data on RTAs from MTTU in Ghana have provided adequate proof that their data is credible. This research makes several key contributions. Firstly, it showcases the application of the SARIMA model for capturing temporal patterns in accident data. Additionally, it introduces the implementation of the Facebook Prophet model, which adeptly handles holidays, special events, and outliers. As a result of incorporating potential changepoints into the study, the models are more capable of adapting to shifts in accident patterns than traditional methods. A rigorous comparative analysis of SARIMA and Facebook Prophet models evaluates their predictive capabilities, collectively providing an innovative and practical framework for accurate road traffic accident forecasting. Due to the increasing rate of RTAs in Ghana, undertaking this research is helpful. The results and findings of this study would be beneficial for road safety planning to help minimize road traffic accidents and fatalities in Ghana. The time series model developed in this study is recommended for use by the MTTU, NRSC, and relevant stakeholders to help monitor the efficacy of diverse road safety policies. Additionally, the study's findings will contribute to the body of academic literature concerning RTAs

The remainder of the paper is organized as follows: Section 2 discusses the data and methods used for the study, including ARMA, ARIMA, SARIMA, FB Prophet models, SARIMA model building process, model identification tools, model diagnostics, and model accuracy. Section 3 presents and discusses the results of the investigation. Section 4 concludes the research and provides recommendations.

2. Data and methods

Secondary RTAs data on the Madina-Adenta Highway were retrieved from police reports from 2013 to 2019 and analyzed using univariate SARIMA and FB Prophet time series model with potential changepoints. The recorded number of RTAs data used in the study include rear-end collisions, head-on collisions, side impact collisions, rollovers, pedestrian or cyclist Accidents, multi-vehicle pileups and run-off accidents. Monthly RTA data from 2013 to 2018 (72 months) were used in building the two models, while the monthly RTAs for 2019 (12 months) were used in testing the accuracy of the two models under consideration. Data from 2020 to the first half of 2023 of the number of Road Traffic Accidents (RTAs) per month were regrettably excluded from this study due to unforeseen circumstances resulting from the COVID-19 pandemic. These circumstances led to non-representative and incomplete RTA data specific to the Madina-Adenta Highway. The study focused on analyzing RTA data from this highway, revealing a conspicuous seasonal pattern. This observation prompted the application of SARIMA and FB Prophet models. The study's analysis was conducted using the R programming language. The study's data and codes are publicly accessible on GitHub via the repository located at github.com/Agyemang1z/Road-Accidents.

2.1. AutoRegressive moving average (ARMA) model

The autoregressive moving average ARMA (p, q) model is formulated by the combination of autoregressive AR (p) and moving average MA (q) model, which is a suitable model for univariate time series data. The AR (p) model is given mathematically by (1):

$$x_t = \vartheta_0 + \vartheta_1 x_{t-1} + \vartheta_2 x_{t-2} + \ldots + \vartheta_p x_{t-p} + \varepsilon_t = \vartheta_0 + \sum_{i=1}^{p} \vartheta_i x_{t-i} + \varepsilon_t \tag{1}$$

where x_t are the observed values, ε_t is random shocks at time t, ϑ_i ($i = 1, 2, \ldots, k$) are the parameters of the $AR(p)$ model, ϑ_0 is the constant term, and p is the order of the time series model.

The $MA(q)$ model is likewise given by (2):

$$x_t = \mu + \varepsilon_t - \alpha_1\varepsilon_{t-1} - \alpha\varepsilon_{t-2} - \ . \ . \ . \ - \alpha_q\varepsilon_{t-q} = \mu + \varepsilon_t - \sum_{i=1}^{q}\alpha_i\varepsilon_{t-i} \quad (2)$$

where μ is the mean of the series, α_i $(i = 1, 2, \ . \ . \ . \ , q)$ represents parameters of the model with order q, with random errors ε_t are assumed as a white noise process.
The mixed autoregressive moving average $ARMA(p,q)$ model is also expressed mathematically in (3) by:

$$y_t = \mu + \sum_{i=1}^{p}\vartheta_i x_{t-i} + \varepsilon_t - \sum_{i=1}^{q}\alpha_{1i}\varepsilon_{t-i} \quad (3)$$

where the order (p,q) represents p order for autoregressive $AR(p)$ and q for the moving average $MA(q)$ terms.

2.2. AutoRegressive integrated moving average (ARIMA) model

The $ARIMA(p,d,q)$ model using the lag operator is mathematically expressed in (4) as:

$$\vartheta(L)(1-L)^d x_t = \mu + \alpha(L)\varepsilon_t$$
$$\left(1 - \sum_{i=1}^{p}\vartheta_i L^i\right)(1-L)^d x_t = \mu + \left(1 - \sum_{i=1}^{p}\alpha_i L^i\right)\varepsilon_t \quad (4)$$

The order of autoregressive, integrated, and moving average terms of the model are given respectively by p, d and q; d is the differencing required to achieve series stationarity.

2.3. Box-Jenkins seasonal ARIMA (SARIMA) model

Many real-world time series datasets feature a seasonal component that repeats after every S observation. For example, consider utilizing a monthly observation time series dataset, where $S = 12$. We can generally anticipate that X_t to a large extent rely on X_{t-12} and probably X_{t-24} in addition to terms such as $X_{t-1}, X_{t-2}, \ldots$. The Box and Jenkins generalization $ARIMA(p,d,q)$ include seasonal components and are often characterized as a general multiplicative Seasonal ARIMA model abbreviated as SARIMA $(p,d,q)\times(P,D,Q)_S$ model and expressed mathematically in the study in (5) by:

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_S{}^D X_t = \theta_q(B)\Theta_Q\left(B^S\right)Z_t \quad (5)$$

Where S denotes the seasonal lag, B denotes the backshift operator and Z_t is the random error component. ϕ_p and Φ_P are the non-seasonal and autoregressive seasonal parameters. Additionally, θ_q and Θ_Q are the non-seasonal and moving average seasonal parameters. p and q are, respectively, the orders of non-seasonal autoregressive and moving average parameters, whilst P and Q are orders of autoregressive and moving average seasonal parameters, respectively. Lastly, d and D respectively represent the non-seasonal and seasonal differences. ∇^d means we apply the ∇ operator d times and similarly for $\nabla_S{}^D$.

2.4. ARIMA model building process

The ARIMA model uses a three-stage approach to get a suitable model for a forecast. These include:

1. Model Identification: The model identification involves determining if the time series data is stationary or non-stationary [48]. If it is non-stationary, determine the degree of differencing needed to make it stationary. The acquisition of the AR order p and the MA order q follows afterwards. Typically, the non-stationary time series data is frequently shown by an autocorrelation graph with slow decay. For this study, the Augmented Dickey-Fuller Test (ADF) test was used to test for series stationarity.
2. Model Estimation: This entails determining the best feasible estimates for the Box-Jenkins model parameters [49]. Nonlinear least squares and maximum likelihood estimation are the primary methodologies for fitting Box-Jenkins models. The parameters in this study were estimated using Maximum Likelihood Estimation (MLE).
3. Model Diagnostic: This stage checks if the model is adequate or not. If the model is inadequate, it is essential to return to stage one and choose a better model. Once the model has been selected, estimated, validated, and determined to be acceptable, it is utilized to generate forecasts.

2.5. Test of significance of model coefficients

For each coefficient, the estimated t-value is given by (6) as:

$$t = \frac{estimated\ coefficient}{standard\ error} \quad (6)$$

If $|t| \geq 2$, the estimated coefficient is significantly different from zero (0), and the model coefficient is statistically significant. Also, if the p-value of a model coefficient is less than the 5% significance level, the estimated coefficient is adjudged statistically substantial and otherwise.

2.6. Model identification tools

They evaluate the balance between model adequacy and model complexity. Various indicators measuring the quality of fit applied in this study's model identification process include Akaike Information Criterion (AIC), corrected Akaike Information Criterion (AICc), Bayesian Information Criterion (BIC), Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). The AIC uses the MLE approach. The MLE technique is utilized to estimate a variety of feasible SARIMA models for this approach, and each AIC computed using (7)

$$AIC = -2\ln L + 2W \tag{7}$$

where $\ln L$ is the model's log-likelihood, and W is the number of model parameters. In the case of two or more competing models, the one with the lower AIC is superior. The AIC exhibits bias, particularly evident when the ratio of parameters to available data is high. [50] demonstrated that the tendency might be approximated by introducing an additional non-static penalty factor to the AIC, resulting in the development of the corrected AIC, denoted by AICc, and mathematically given by (8) as

$$AIC_c = AIC + \frac{2W(W+1)}{N-W-1} \tag{8}$$

where N is the sample size or the number of time series observations.
The BIC, like AIC, also uses the MLE. It is expressed by (9) as:

$$BIC = -2\ln L + W\ln(N) \tag{9}$$

The BIC penalizes the number of estimated model parameters more severely than the AIC. Applying minimal BIC for model selection results in a model with fewer parameters than that chosen for AIC. According to the concept of parsimony, BIC is considerably superior in model selection over AIC. A lower BIC value indicates that the model fits better.

2.7. Test of model diagnostics for SARIMA model

The Box-Pierce and Ljung-Box tests is employed to check the adequacy of the study's estimated model. The Ljung-Box test fits residual (error term) randomness based on several lags. If the autocorrelations of the residuals are small, the model does not exhibit a significant lack of fit and is thus assumed adequate. The Lilliefors (KS) test is also used in this study to check for the normality of the model's residuals, and it must have a p-value more significant than 0.05; otherwise, the model's residuals are considered not to be normally distributed. The Ljung-Box statistics is a function of the accumulated sample autocorrelation, ρ_h, up to any specified time lag k. It is obtained as a function of h given by (10) as

$$Q \; = \; n(n+2)\sum_{h=1}^{k}\frac{\rho_h^2}{n-h} \tag{10}$$

where $Q \sim \chi^2_{\alpha,n-p-q}$, and n is the number of data points that can be used after any differencing processes. When the calculated value of Q is obtained, the critical region for rejection of the hypothesis of randomness is $Q > \chi^2_{\alpha,n-p-q}$. This means that the model under consideration is inadequate but adequate if otherwise. When the model is insufficient, there arises a need to fit an appropriate model. That is, going back to the model identification and developing a better model.

2.8. Test of model accuracy

Detecting the best-fit model based on accuracy ensures that the chosen model is not over fitted. It is important to note that a high error rate indicates that the model is built poorly, whereas a low error rate indicates that it is built well. The accuracy of the two competitive models were computed using (11), (12), (13), (14) and (15) respectively: $MAPE$, $MAE(MAD)$, MSE, $RMSE$ and Theil U1 Statistic (τ).

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t - \hat{y}_t|}{y_t} \times 100\%. \tag{11}$$

$$MAE(MAD) = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t|. \tag{12}$$

$$MSE = \frac{\sum_{t=1}^{n}\left(y_t - \hat{y}_t\right)^2}{n}. \tag{13}$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{n}}. \tag{14}$$

$$\tau = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2} \div \left(\sqrt{\frac{1}{n}\sum_{t=1}^{n} y_t^2} + \sqrt{\frac{1}{n}\sum_{t=1}^{n}\hat{y}_t^2}\right),. \tag{15}$$

where y_t are the actual values, $\hat{y}_t$ are the forecast values and $y_t - \hat{y}_t = \epsilon_t$ are the forecast errors. $0 \leq \tau \leq 1$, for $\tau \approxeq 0$ implies good fit of model to data and $\tau \approxeq 0$ implies poor fit of model to data.

2.9. Forecasting

After successfully identifying, estimating, diagnosing, and deciding on the appropriate time series model to use, forecasting can be done. If the current time is denoted by t, the forecast for Y_{t+r} is the $r-$ period ahead forecast and denoted by $\hat{Y}_{t+r}$. The infinite MA representation of the forecast is given in (16) by;

$$\hat{Y}_{t+r} = \mu + \sum_{i=1}^{\infty}\psi_i\varepsilon_{t+r-i} \tag{16}$$

and an $ARIMA(p,d,q)$ process at time $t+r$ (that is, a period in the future) is given in (17):

$$\hat{Y}_{t+r} = \sum_{i=1}^{p+d}\emptyset_i y_{t+r-i} + \varepsilon_{t+r} - \sum_{i=1}^{q}\theta_i\varepsilon_{t+r-i} \tag{17}$$

where, ψ_i is the weight (a constant). Once a forecast is obtained for Y_{t+1}, it can be used to obtain a forecast for Y_{t+2} and then, these two generate a forecast for Y_{t+3}. This can be used to acquire forecasts for any point in time.

2.10. Facebook Prophet forecasting model

Developed by Facebook, the FB Prophet, an additive regression model is in high demand for forecasting purposes due to its three main features: trend, seasonality, and holiday. The model is expressed in (18) as

$$y(t) = \alpha(t) + \beta(t) + \eta(t) + \varepsilon(t) \tag{18}$$

where $y(t)$ is the forecast; the model parameters $\alpha(t)$, $\beta(t)$ and $\eta(t)$ are respectively the trend (non-periodic changes), seasonal (periodic changes) and holidays effects, which gives irregular schedules. $\varepsilon(t)$ is the error term of the forecast $y(t)$ which represent any unusual changes. The FB Prophet model adopts a Fourier series to fit models with seasonality effects $s(t)$ represented in (19) as

$$s(t) = \sum_{k=1}^{N}\alpha_k\cos\left(\frac{2\pi kt}{p}\right) + \beta_k\sin\left(\frac{2\pi kt}{p}\right) \tag{19}$$

where p is the period of the seasonal pattern, α_k and β_k are the Fourier coefficients. Employing the data's rising points as a reference, the Prophet model adopts a logistic growth curve trend to discern trends. FB Prophet is adept at managing time series data characterized by significant seasonal fluctuations and a substantial historical data span. Notably, the Prophet model effectively manages outliers, even in scenarios involving missing data or shifts in trends [51,52]. The effective application of a Prophet model necessitates the variables y (target) and ds (Date Time) in the time series. It demonstrates optimal performance when applied to datasets encompassing multiple seasons and featuring notable seasonal impacts [53]. For the purpose of this study, the potential change points were chosen as the months with the major holidays in Ghana. This study chose January, March, April, May and December as our potential changepoints and the number of changepoints was set as 30. We operationalized the trend model using a saturating growth approach, establishing the logistic growth model's carrying capacity at 10. We set the interval width and change point prior Scale to 0.8 and 0.05 respectively.

The introduction and comparison of the FB Prophet model (with potential changepoints) with other competitive time series models (such as SARIMA) to forecast RTAs in Ghana has not been explored and to the best of our knowledge, this is the first study to explore the FB Prophet model in this domain. Figs. 1 and 2 presents the working model of the SARIMA and FB Prophet adopted for the study.

2.11. Comparison between Facebook Prophet and ARIMA models

1. Model Complexity: Facebook Prophet has been designed to be a user-friendly forecasting tool with minimal configuration requirements [54]. It automates several steps involved in time series forecasting, such as handling seasonality, trend detection, and outliers. ARIMA is a more traditional and widely used time series forecasting model. A key aspect of ARIMA models is tuning model parameters, such as order (p,d,q) values for autoregressive, differencing, and moving average components.
2. Seasonality Handling: As part of its integrated functionality, FB Prophet is able to handle a variety of seasonalities, such as daily, weekly, and yearly patterns [55]. It can handle multiple seasonal components and also handle irregular holidays and events.

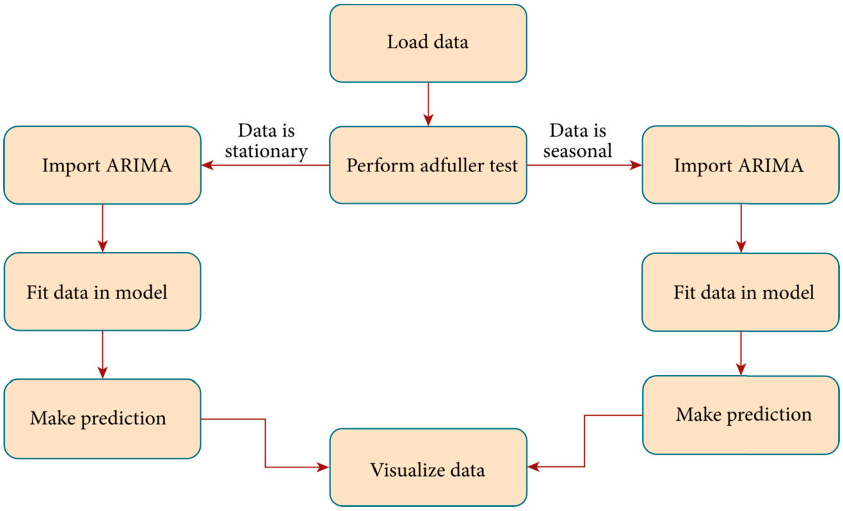


Fig. 1. Working model of ARIMA adopted from [53].

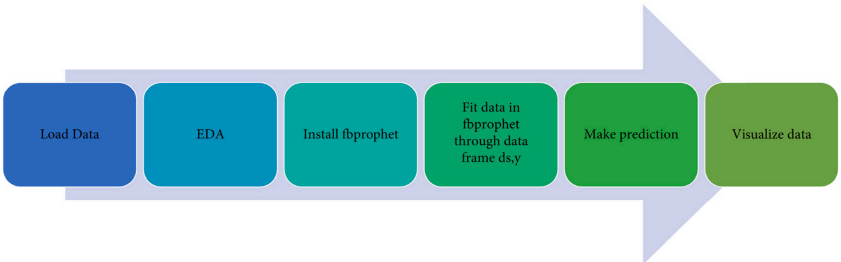


Fig. 2. Working model of Facebook Prophet adopted from [53].

Models that use ARIMA can also handle seasonal data through seasonal differencing or by manually incorporating seasonal components.

3. Trend Detection and Outlier handling: FB Prophet automatically detects and models both linear and non-linear trends and can handle outliers in the data. It can identify and adjust for these outliers, preventing them from overly influencing the forecast and can handle situations where the trend changes over time. ARIMA models can capture linear trends but may not handle non-linear trends effectively.
4. Interpretability: The FB Prophet provides a more interpretable forecast due to its breakdown into the trend, seasonality, and holiday components. In addition, it provides visualizations and diagnostics for evaluating the model's performance. The ARIMA model is less interpretable since it focuses mostly on the statistical properties of the time series. An understanding of the underlying mathematics is necessary to interpret the model parameters and diagnostics of an ARIMA model.

2.12. Examples of solving optimization problems in road traffic accident research

It is worth knowing that solving optimization problems in road accident research is key to reducing toad traffic accidents. Below are three concrete examples of solving optimization problems in road accident related research.

1. Traffic Signal Timing Optimization
   - Problem: Enhancing the synchronization of traffic signals at intersections to alleviate congestion and lower the probability of accidents.
   - Solution: By optimizing the timing of traffic signals, researchers and traffic engineers can work to minimize the likelihood of accidents occurring. An objective function is defined, which could include minimizing the total number of conflict points (locations where accidents are more likely to occur) or maximizing the throughput of vehicles. Constraints are established to ensure that traffic signal timings adhere to safety and operational standards. These constraints may include minimum green time, maximum cycle length, and pedestrian crossing times. Various optimization algorithms, such as those established by [56,57] may be applied to find the optimal signal timing plan that minimizes the objective function while satisfying the constraints. The optimized signal timing plan is then simulated to assess its impact on traffic flow and safety. Once an optimized signal timing plan is validated, it can be implemented at the intersection.
2. Route Planning for Emergency Services

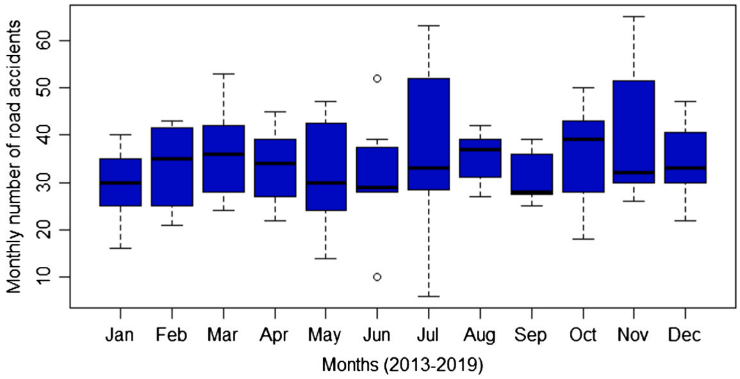


Fig. 3. Box-plot with data on traffic accidents by months (2013–2019).

- Problem: Finding the optimal routes for emergency vehicles (e.g., ambulances, fire trucks) to reach accident scenes quickly while avoiding traffic congestion.
- Solution: Applying optimization algorithms such as that established by [58–60] to take into account real-time traffic data, accident locations, and the urgency of the situation. These algorithms can recommend the fastest and safest routes for emergency responders, potentially saving lives by reducing response times.

3. Vehicle Fleet Optimization for Safety Inspections
   - Problem: Optimizing the scheduling and routing of safety inspection teams to inspect a large number of vehicles efficiently, ensuring compliance with safety regulations and reducing the risk of accidents due to faulty vehicles.
   - Solution: Use vehicle routing optimization algorithms to determine the best inspection routes for a fleet of inspectors, considering factors like the locations of inspection sites, inspection durations, and traffic conditions. The goal is to maximize the number of inspections performed within a given time frame while minimizing travel distance and time.

These examples demonstrate how optimization techniques can be applied in road traffic accident research to enhance safety, improve traffic flow, and allocate resources effectively. They leverage data and modelling to make informed decisions and reduce the risk of accidents on the road.

3. Results and discussion

This section presents the outcomes of the forecast generated by the SARIMA and FB Prophet models.
Over the seven-year span under study, January demonstrated the lowest occurrence of road accidents, with an average of 29 incidents. The first and third quarters each witnessed an average of 33 road accidents per month. Notably, the fourth quarter experienced the highest monthly average of reported RTAs, with an average of 37 incidents, marking this period as the most perilous for drivers, passengers, pedestrians, and other road users. Unexpectedly, the most perilous month of the year is November, despite the festive activities typically associated with December in Ghana. November falls within the fourth quarter and records an average of 41 traffic accidents per month. The ratio between the highest average value (November) and the lowest average value (January) stands at 41.38%. The standard deviations of the monthly road accidents cases depicted in Fig. 3 by year are: 10 (2013), 11 (2014), 8 (2015), 4 (2016), 11 (2017), 12 (2018), and 4 (2019).

3.1. Development of the SARIMA model

Fig. 4 depicts the frequency of RTAs on the Madina-Adenta highway as spikes and troughs increase and decrease. The average number of traffic accidents per year are 26 (2013), 35 (2014), 27 (2015), 29 (2016), 39 (2017), 41 (2018) and 42 (2019). From the reference year 2013, it is evident that the average number of road accidents has risen by 34.62% (2014), 3.85% (2015), 11.54% (2016), 50.00% (2017), 57.69% (2018) and 61.54% (2019). From Fig. 4, we observe some upward and downward surges in the series. Also, Fig. 4 shows that the monthly RTAs time series dataset under examination has a solid and persistent seasonal tendency. This indicates that seasonal components exist in the monthly RTAs. This further suggests that the monthly RTAs data on the Madina-Adenta highway is non-stationary. To effectively apply an appropriate time series model to the RTA data, it's imperative to eliminate both the seasonality and trend present within the dataset. The single exponential smoothing method was employed to deal with the seasonality and trend components as suggested by the ndiffs() function in R.

3.2. Single exponential smoothing method

The road traffic accident data underwent a first-order differencing method to effectively eliminate both the trend and seasonality from the original dataset. Fig. 5 helps us to see that the differenced data exhibits stationarity with constant mean and variance. The Augmented Dickey-Fuller Test (ADF) test was employed to confirm series stationarity test.

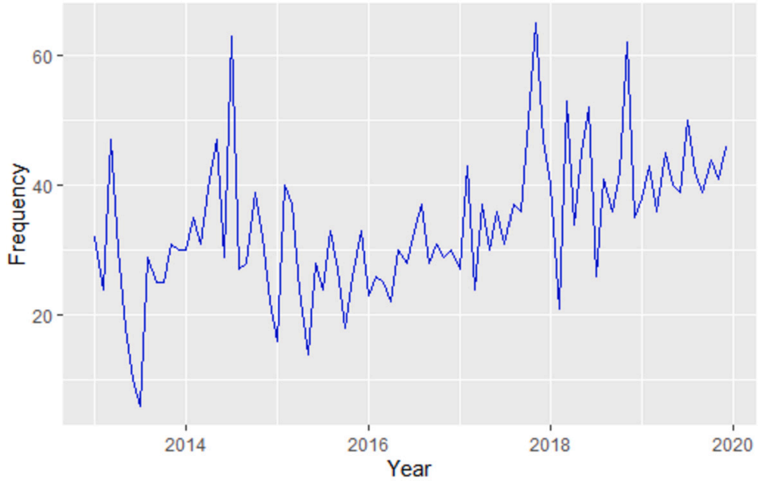


Fig. 4. Time series plot of Road Traffic Accident cases from 2013 to 2019.

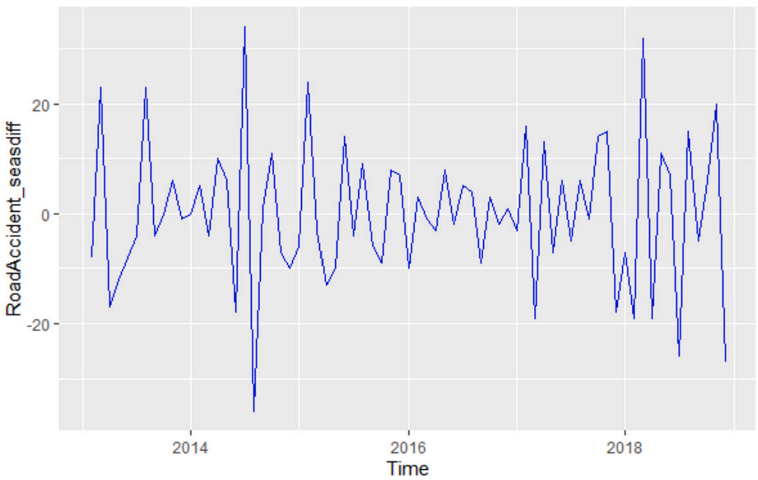


Fig. 5. Single Exponential plot of Road Traffic Accident cases.

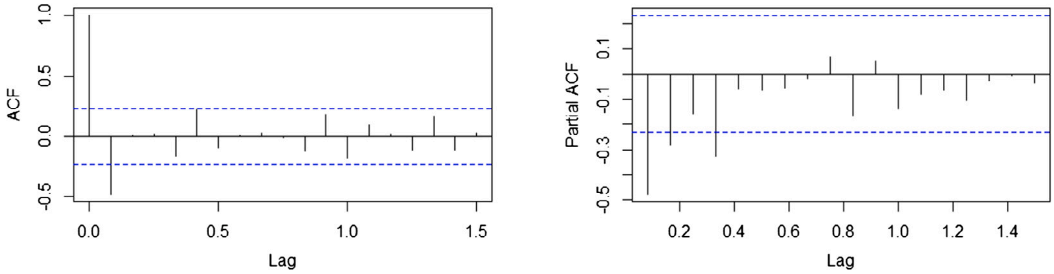
Table 1
ADF test for Level Stationarity.

Test	Statistic	Lag Order	P-value
ADF	−5.8053	4	0.0100

We reject the null hypothesis of non-stationarity and conclude that the first differenced RTAs data is stationary since $0.0100 < 0.05$ as evident in Table 1.

3.3. Autocorrelation and partial autocorrelation function plots of the first difference

From Fig. 6(a), looking at the autocorrelation function with the error limits, autocorrelation at lag one is significant. This depicts a possible behaviour of MA (1). For the plot of ACF, we notice sinusoidal waves shown in the autocorrelation plot.
Fig. 6(b) depicts the partial autocorrelation function plot, which demonstrates that the partial autocorrelation coefficient exceeds the error bounds at lags 1, 2, and 4, which are significant. The error limit at lag 3 is insignificant, and it is the first to do so. Since the PACF cuts off after lags 1, 2 and lag 4, this suggests possible behaviour of SMA (1), SMA (2) or SMA (4). The ACF and PACF plots suggest that $Q = 1, 2$ *or* 4, $p = 1$, $P = 1$, $d = 1, D = 0$ and $q = 1$ would be needed to model the road traffic accident data. The following tentative models; $SARIMA\ (1,1,1)\times(1,0,1)_{12}$, $SARIMA\ (1,1,1)\times(1,0,0)_{12}$, $SARIMA\ (1,1,0)\times(0,0,1)_{12}$, $SARIMA\ (0,1,1)\times(1,0,0)_{12}$, $SARIMA\ (0,1,2)\times(1,0,0)_{12}$, $SARIMA\ (1,1,2)\times(1,0,0)_{12}$ and $SARIMA\ (0,1,3)\times(1,0,2)_{12}$ can be suggested after careful examination of the ACF and PACF plots in Fig. 6.



(a) ACF of diffferenced RTA cases (b) PACF of diffferenced RTA cases

Fig. 6. Plot of the ACF and PACF of First Difference road traffic accident cases.

Table 2
Final parameter estimates of $SARIMA$ $(1,1,1)\times(1,0,1)_{12}$.

Parameter	Coefficient	Standard Error	Z-value	P-value
AR (1)	−0.0057	0.0007	−7.8014	0.000***
MA (1)	−0.7831	0.0069	−1111.2357	0.000***
SAR (1)	−0.5513	NaN	NaN	NaN
SMA (1)	0.3654	0.0028	132.0225	0.000***

Table 3
Final parameter estimates of $SARIMA$ $(1,1,1)\times(1,0,0)_{12}$.

Parameter	Coefficient	Standard Error	Z-value	P-value
AR (1)	0.0422	0.1543	0.2733	0.7846
MA (1)	−0.8080	0.0975	−8.2895	0.000***
SAR (1)	−0.2176	0.1363	−1.5965	0.1104

Table 4
Final parameter estimates of $SARIMA$ $(1,1,0)\times(0,0,1)_{12}$.

Parameter	Coefficient	Standard Error	Z-value	P-value
AR (1)	−0.4929	0.1075	−4.5863	0.000***
SMA (1)	−0.1678	0.1325	−1.2654	0.2057

Table 5
Final parameter estimates of $SARIMA$ $(0,1,1)\times(1,0,0)_{12}$.

Parameter	Coefficient	Standard Error	Z-value	P-value
MA (1)	−0.7905	0.0807	−9.7978	0.000***
SAR (1)	−0.2170	0.1364	−1.5911	0.1116

3.4. Possible seasonal ARIMA models parameter estimation

The tables below display the final parameter estimates of the possible seasonal ARIMA models under consideration with their respective coefficients, standard error, and p-values.

From Table 2, AR(1), MA (1) and SMA (1) are statistically significant at the 5% significance level.

From Table 3, MA (1) with a coefficient of -0.8080 and a Z-value of -8.2895 is statistically significant.

AR (1) with a coefficient of -0.4929 and a standard error of 0.1075 statistically significant as seen in Table 4.

From Table 5, MA (1) is statistically significant with p-value < 0.05.

From Table 6, MA (1) is statistically significant with p-value < 0.05.

Table 6
Final parameter estimates of $SARIMA$ $(0,1,2)\times(1,0,0)_{12}$.

Parameter	Coefficient	Standard Error	Z-value	P-value
MA (1)	−0.7665	0.1186	−6.4636	0.000***
MA (2)	−0.0325	0.1209	−0.2686	0.7883
SAR (1)	−0.2179	0.1363	−1.5983	0.1100

Table 7
Final parameter estimates of $SARIMA\ (1,1,1)\times(1,0,0)_{12}$.

Parameter	Coefficient	Standard Error	Z-value	P-value
AR (1)	0.0422	0.1543	0.2733	0.7846
MA (1)	−0.8080	0.0975	−8.2895	0.000***
SAR (1)	−0.2176	0.1363	−1.5965	0.1104

Table 8
Final parameter estimates of $SARIMA\ (0,1,3)\times(1,0,2)_{12}$.

Parameter	Coefficient	Standard Error	Z-value	P-value
MA (1)	−0.7654	0.1204	−6.3592	0.000***
MA (2)	−0.0072	0.1710	−0.0420	0.9665
MA (3)	−0.0269	0.1639	−0.1645	0.8693
SAR (1)	−0.3745	0.4348	−0.8615	0.3890
SMA (1)	0.1952	0.4297	0.4543	0.6496
SMA (2)	0.1129	0.2598	0.4346	0.6638

Table 9
Potential SARIMA models with their AIC, AIC_C and BIC values.

SARIMA Models	AIC	AIC_C	BIC
SARIMA $(1,1,1)\times(1,0,1)_{12}$	542.08	543.01	553.4
SARIMA $(1,1,1)\times(1,0,0)_{12}$	540.49	541.10	549.54
SARIMA$(1,1,0)\times(0,0,1)_{12}$	552.94	553.30	559.73
SARIMA $(0,1,1)\times(1,0,0)_{12}$	538.56	538.92	545.35
SARIMA $(0,1,2)\times(1,0,0)_{12}$	540.49	541.10	549.54
SARIMA $(1,1,2)\times(1,0,0)_{12}$	542.51	543.43	553.82
SARIMA$(0,1,3)\times(1,0,2)_{12}$	546.10	547.88	561.94

Table 10
Ljung-Box test of SARIMA $(0,1,1)\times(1,0,0)_{12}$.

Test	Statistic	P-value
Ljung-Box	6.5417	0.8864

From Table 7, MA (1) is also statistically significant at alpha level of < 0.05.
MA(1) is also statistically significant in the $SARIMA\ (0,1,3)\times(1,0,2)_{12}$ model as seen in Table 8.

3.5. Fitting the suitable SARIMA model to road accident time series data

Both the initial differenced ACF and PACF plots were rigorously examined when modelling the road traffic accident series data to produce the suitable-fitted model based on their respective AIC, AIC_C and BIC values.

From Table 9, comparing the AIC, AIC_C and BIC values of the candidate models, it can be deduced that the SARIMA $(1,1,0)\times(0,0,1)_{12}$ (highlighted in red) has an AIC value of 538.56, an AIC_C value of 538.92 and a BIC value of 545.35 which are the lowest accuracies of all the tentative SARIMA models constructed. Therefore, per the model selection criteria, SARIMA $(1,1,0)\times(0,0,1)_{12}$ is the suitable SARIMA time series model for modelling the Madina-Adenta highway road traffic accident.

The general multiplicative Seasonal ARIMA model is thus given in the study by equation (5) as:

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_S{}^D X_t = \theta_q(B)\Theta_Q\left(B^S\right)Z_t$$

Hence, the seasonal ARIMA $(0,1,1)\times(1,0,0)$ can be represented in (20) as

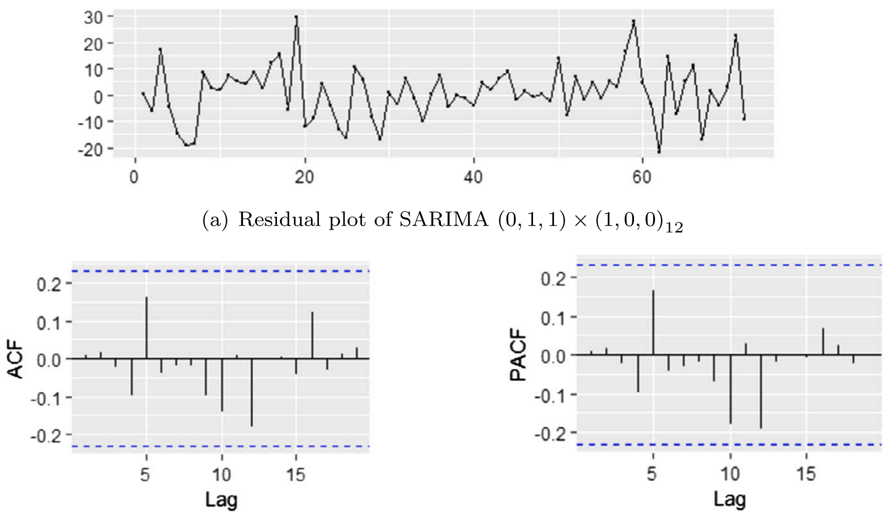
$$\phi_0(B)\Phi_1\left(B^{12}\right)\nabla^1\nabla_{12}{}^0 X_t = \theta_0(B)\Theta_0\left(B^{12}\right)Z_t \tag{20}$$

3.6. Diagnostic checking of estimated model

The suitable-fitted model for modelling the road accident data is then tested further to make theoretical conclusions about the model as a good fit and an optimal model for both estimation and forecasting.

Table 11
Lilliefors (KS) Normality test of SARIMA $(0,1,1) \times (1,0,0)_{12}$.

Test	Statistic	P-value
Lilliefors (KS)	0.0841	0.2364



(a) Residual plot of SARIMA $(0,1,1) \times (1,0,0)_{12}$

(b) ACF of SARIMA $(0,1,1) \times (1,0,0)_{12}$ (c) PACF of SARIMA $(0,1,1) \times (1,0,0)_{12}$

Fig. 7. Plot of ACF and PACF plot of residuals of SARIMA $(0,1,1) \times (1,0,0)_{12}$ model.

From Table 10, Given the p-value of 0.8864, which exceeds the significance threshold of 0.05, we conclude that the residuals of the estimated model demonstrate independence and follow an identically distributed or white noise process. Hence, the model displays no substantial lack of fit.

At a 5% significance level, since the p-value is greater than the alpha level (thus, $0.2364 > 0.05$), the residuals of the estimated SARIMA $(0,1,1) \times (1,0,0)_{12}$ model are normally distributed as evident in Table 11.

3.7. Test of model adequacy of SARIMA

The limits of the 2σ are given by $\pm Z_{\frac{\alpha}{2}} \times \frac{1}{\sqrt{n}}$.

$$-1.96\left(\frac{1}{\sqrt{72}}\right) \leq r \leq 1.96\left(\frac{1}{\sqrt{72}}\right)$$

This is simplified as $-0.2310 \leq r \leq 0.2310$. Hence, the confidence interval of the random error spans across two (2) standard deviations. Fig. 7 shows two blue horizontal lines indicating this claim. It could be seen that the errors are within ± 0.2310 as evident in the ACF/PACF plots. It should also be noted that the plot of these autocorrelations shows no systematic structure, indicating that the residuals are purely random.

3.8. Box-Pierce test for larger lags

The modified Ljung-Box test (Box-Pierce test) for larger lags was employed to ascertain whether the proposed model is adequate for forecasting.

Observing Table 12, it is apparent that the p-values at different lags considerably surpass the 0.05 significance level. As a result, we lack substantial evidence to reject our model. Given its adequacy for lags 12, 24, 36, 48, and 60, it is plausible to assume its adequacy for larger lags as well. Hence, the SARIMA $(0,1,1) \times (1,0,0)_{12}$ model is adequate at a 0.05 level of significance and can be used to forecast future road accident cases optimally.

3.9. Evaluation of SARIMA model

It is clear from Table 13 that the SARIMA $(0,1,1) \times (1,0,0)_{12}$ model developed for RTAs on the Madina-Adenta Highway has a forecast accuracy of 93.1025%, displayed by the model's MAPE of 6.8975% which is indicative of highly accurate forecasting as suggested by [61]. A τ statistic of 0.0376 further indicates good model fit.

Table 12
Modified Ljung-Box (Box-Pierce test) for larger lags of SARIMA $(0,1,1)\times(1,0,0)_{12}$.

Lags	Chi-Square Statistic	P-value
12	5.7045	0.9302
24	8.1798	0.9989
36	14.956	0.9992
48	19.566	0.9999
60	25.717	1.0000

Table 13
Out of sample Validation for $SARIMA$ $(0,1,1)\times(1,0,0)_{12}$ with Forecast Performance Statistics.

Month Year	y_t	$\hat{y}_t$	y_t^2	$\hat{y}_t^2$	ϵ_t	$\lvert\epsilon_t\rvert$	ϵ_t^2	$\lvert\frac{\epsilon_t}{y_t}\rvert$	$\lvert\% \ Error\rvert$
Jan 2019	38	43	1444	1849	5	5	25	0.1316	13.16
Feb 2019	43	47	1849	2209	4	4	16	0.0930	9.30
Mar 2019	36	40	1296	1600	4	4	16	0.1111	11.11
Apr 2019	45	44	2025	1936	−1	1	1	0.0227	2.27
May 2019	40	42	1600	1764	2	2	4	0.0500	5.00
Jun 2019	39	40	1521	1600	1	1	1	0.0256	2.56
Jul 2019	50	46	2500	2116	−4	4	16	0.0800	8.00
Aug 2019	42	43	1764	1849	1	1	1	0.0233	2.33
Sep 2019	39	44	1521	1936	5	5	25	0.1282	12.82
Oct 2019	44	42	1936	1764	−2	2	4	0.0454	4.54
Nov 2019	41	38	1681	1444	−3	3	9	0.0732	7.32
Dec 2019	46	44	2116	1936	−2	2	4	0.0435	4.35
Total	**503**	**513**	**21253**	**22003**		**34**	**122**	**0.8277**	
MAE	2.8333								
MAPE	6.8975								
MSE	10.1667								
RMSE	3.1885								
Theil's U1 Statistic (τ)	0.0376								

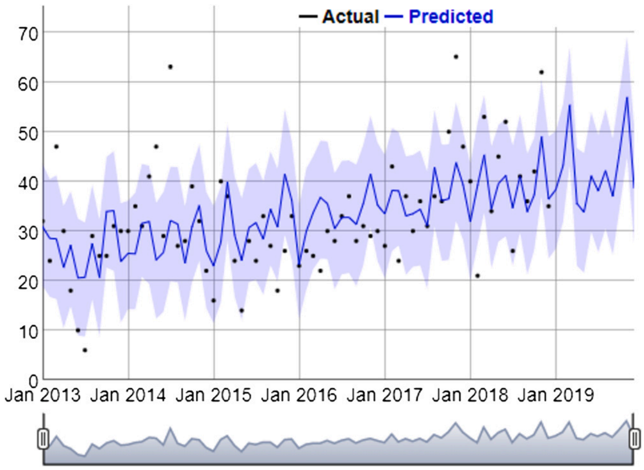


Fig. 8. Plot of Forecasted values of Facebook Prophet model with 95% confidence limits.

From Table 14, it is evident that the best model by the auto.arima function in R programming software confirms that our selected model (ARIMA(0,1,1)(1,0,0)[12]) using the diagnostic of the ACF and PACF plots is the ideal seasonal model to forecast the underlying RTAs data.

3.10. Evaluation of Facebook Prophet model

The performance of the FB Prophet model, with change points being the months with the major holidays in Ghana is assessed. January, March, April, May and December were chosen as potential change points.

It is evident from Table 15 and Fig. 8 that the Facebook Prophet model developed for RTAs on the Madina-Adenta Highway has a forecast accuracy of 84.3569%, displayed by the model's MAPE of 15.6431%.

Table 14
Possible tentative ARIMA models extracted by auto.arima function.

Model	Evaluation Statistic
ARIMA(0,1,0)(0,0,1)[12]	569.2443
ARIMA(0,1,0)(0,0,1)[12] with drift	571.4203
ARIMA(0,1,0)(0,0,2)[12]	570.8983
ARIMA(0,1,0)(0,0,2)[12] with drift	573.145
ARIMA(0,1,0)(1,0,0)[12]	568.8583
ARIMA(0,1,0)(1,0,0)[12] with drift	571.0352
ARIMA(0,1,0)(1,0,1)[12]	570.9207
ARIMA(0,1,0)(1,0,1)[12] with drift	573.1644
ARIMA(0,1,0)(1,0,2)[12]	572.3704
ARIMA(0,1,0)(2,0,0)[12] with drift	573.0677
ARIMA(0,1,0)(2,0,1)[12]	573.0674
ARIMA(0,1,0)(2,0,1)[12] with drift	575.3814
ARIMA(0,1,0)(2,0,2)[12]	Inf
ARIMA(0,1,0)(2,0,2)[12] with drift	Inf
ARIMA(0,1,1)(0,0,1)[12] with drift	540.1289
ARIMA(0,1,1)(0,0,2)[12]	541.2295
ARIMA(0,1,1)(0,0,2)[12] with drift	542.4327
ARIMA(0,1,1)(1,0,0) [12]	**538.9229**
ARIMA(0,1,1)(1,0,0) [12] with drift	**539.9962**
ARIMA(0,1,1)(1,0,1)[12]	540.9437
ARIMA(0,1,3)(0,0,1)[12]	543.619
ARIMA(0,1,3)(0,0,1)[12] with drift	544.4773
ARIMA(0,1,3)(0,0,2)[12]	545.7991
ARIMA(1,1,1)(0,0,1)[12]	541.3203
ARIMA(1,1,1)(0,0,1)[12] with drift	542.2117
ARIMA(1,1,1)(0,0,2)[12]	543.4696
ARIMA(1,1,1)(0,0,2)[12] with drift	544.597
ARIMA(1,1,1)(1,0,0)[12]	541.0963
ARIMA(1,1,1)(1,0,0)[12] with drift	542.1079
ARIMA(1,1,1)(1,0,1)[12]	Inf
ARIMA(1,1,1)(1,0,1)[12] with drift	Inf
ARIMA(1,1,1)(1,0,2)[12]	545.4409
ARIMA(1,1,1)(1,0,2)[12] with drift	546.8489
ARIMA(1,1,1)(2,0,0)[12]	Inf
ARIMA(1,1,1)(2,0,0)[12] with drift	Inf
ARIMA(2,1,1)(0,0,1)[12]	543.6271
ARIMA(2,1,1)(0,0,1)[12] with drift	544.5017
ARIMA(2,1,1)(0,0,2)[12]	545.8267
ARIMA(2,1,1)(0,0,2)[12] with drift	546.9558
ARIMA(2,1,1)(1,0,0)[12]	543.4051
ARIMA(3,1,0)(0,0,1)[12]	549.7264
ARIMA(3,1,0)(0,0,1)[12] with drift	551.9101
ARIMA(3,1,0)(0,0,2)[12]	552.022
ARIMA(3,1,0)(0,0,2)[12] with drift	554.3071
ARIMA(3,1,0)(1,0,0)[12]	549.4729
ARIMA(4,1,0)(0,0,1)[12]	543.9409
ARIMA(4,1,0)(0,0,1)[12] with drift	546.0397
ARIMA(4,1,0)(1,0,0)[12]	543.4893
ARIMA(4,1,0)(1,0,0)[12] with drift	545.5827
ARIMA(4,1,1) with drift	548.1061
ARIMA(5,1,0) with drift	548.3584

Best model: ARIMA(0,1,1)(1,0,0) [12] .

The SARIMA $(0,1,1)\times(1,0,0)_{12}$ model developed in this study outperforms the Facebook Prophet model per their respective forecast performance statistics. It is then ideal to use the SARIMA model to make forecasts.

It is observed that road accidents on the Madina-Adenta Highway will show decreasing and increasing spikes from the start of January 2019 to the last quarter of 2019. There will be an anticipated steady pattern in RTA cases from the first quarter of 2020 and afterwards, as depicted by Fig. 9.

4. Conclusion and recommendation

Introducing and producing several vehicles has led to increased traffic accidents and negative consequences. Therefore, constantly reviewing, analyzing, and evaluating the existing situation is necessary. This will help identify the main causes of accidents and the most effective prevention strategies. Creating awareness campaigns to educate people on safe driving practices is also imperative. An integral facet of traffic management within a jurisdiction involves forecasting the frequency of RTAs during specific periods of

Table 15
Out of sample Validation for Facebook Prophet model with Forecast Performance Statistics.

Month Year	y_t	$\hat{y}_t$	y_t^2	$\hat{y}_t^2$	ϵ_t	$\lvert\epsilon_t\rvert$	ϵ_t^2	$\lvert\frac{\epsilon_t}{y_t}\rvert$	$\lvert\% \ Error\rvert$
Jan 2019	38	38	1444	1444	0	0	0	0.0000	0.00
Feb 2019	43	43	0	0	0	0	0	0.0000	0.00
Mar 2019	36	55	1296	3025	−19	19	361	0.5278	52.78
Apr 2019	45	35	2025	1225	10	10	100	0.2222	22.22
May 2019	40	33	1600	1089	7	7	49	0.1750	17.50
Jun 2019	39	41	1521	1681	−2	2	4	0.0513	5.13
Jul 2019	50	38	2500	1444	12	12	144	0.2400	24.00
Aug 2019	42	42	0	0	0	0	0	0.0000	0.00
Sep 2019	39	37	1521	1369	2	2	4	0.0513	5.13
Oct 2019	44	46	1936	2116	−2	2	4	0.0455	4.54
Nov 2019	41	57	1681	3249	−16	16	256	0.3902	39.02
Dec 2019	46	38	2116	1444	8	8	64	0.1739	17.39
Total	**503**	**503**	**21253**	**21699**		**78**	**986**	**1.8772**	
MAE	6.5000								
MAPE	15.6431								
MSE	82.1667								
RMSE	9.0646								
Theil's U1 Statistic	0.1071								

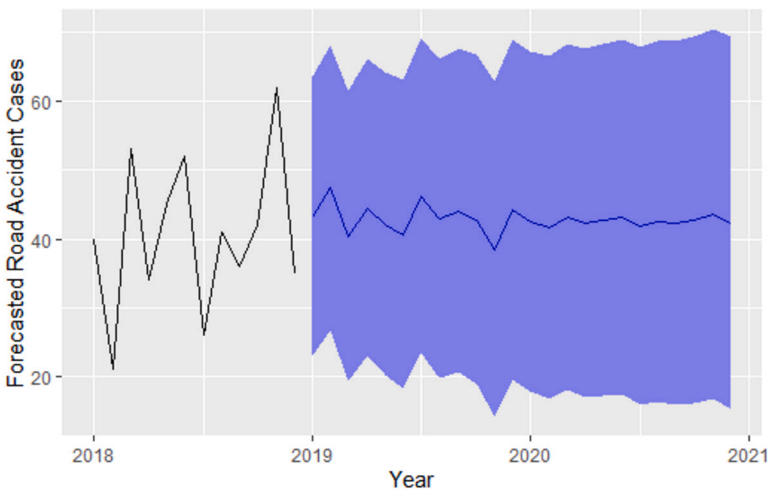


Fig. 9. Plot of Forecasted values of $(0,1,1)\times(1,0,0)_{12}$ model with 95% confidence limits.

the year [61]. This is mainly to get citizens informed about road safety status to understand the issue better, improve their attitudes, and improve their driving habits. Our primary objective for this study is to compare how well SARIMA and Facebook Prophet models with potential change points handle time series data with seasonal components. The study aimed to identify a suitable model that fits the Madina-Adenta RTA data and use it to forecast. The time series plot of the RTA data showed both increasing and decreasing spikes with sharp upward and downward surges, indicative of seasonality in the data. This gives a suspicion of the data being non-stationary. The RTA data was then differenced once to make it stationary. An estimated seasonal ARIMA and FB Prophet models were developed using the monthly RTA time series data from 2013 to 2018 (72 months) as the training set. The monthly RTA time series data for 2019 (12 months) were used to test the accuracy of candidate models. A comparison of all the possible model accuracy metrics (AIC, AIC_C and BIC) of the suggested tentative models for the Madina-Adenta monthly RTAs time series data revealed that $SARIMA$ $(0,1,1)\times(1,0,0)_{12}$ which has an AIC value of 538.56, AIC_C value of 538.92 and a BIC value of 545.35, which were respectively the lowest among all the possible SARIMA models formulated was chosen as the suitable model in modelling the RTAs on the Madina-Adenta highway, Ghana per the selection criteria approaches. $SARIMA$ $(0,1,1)\times(1,0,0)_{12}$ was then subjected to model diagnostic tests. The diagnostic test was conducted on the normality of the residuals, independence of the residuals and the test for model accuracy. The Lilliefors test for normality proved that the residuals were normally distributed. In contrast, the Ljung-Box test shows that the residuals were free of serial or autocorrelation and followed a white noise process. The modified Ljung-Box, also known as the Box-Pierce test for larger lags, proved significant at the 5% significance level for lags 12, 24, 36, 48 and 60, which indicates that it will be significant for larger lags, too, making the estimated model considered ideal for forecasting. A comparative analysis was then made between $SARIMA$ $(0,1,1)\times(1,0,0)_{12}$ and FB Prophet model, where the former provided a high forecast accuracy of 93.1025% relative to the latter's forecast accuracy of 84.3569%. A Theil U1 statistic of 0.0376 for the SARIMA model compared to 0.1071 for the FB Prophet model further indicates a good model fit of the SARIMA model to the RTA data. Even though the FB Prophet model has outperformed SARIMA models in most domains regarding RTA modelling, its applicability to the Ghanaian setting is missing in action. For example, forecasting daily time series of passenger demand for urban rail transit [62], road traffic injury prediction in Northeast China [63], road traffic forecasting in Bangladesh [64]. The study, therefore, recommended that the MTTU and National Safety Road Commission (NRSC) should adopt $SARIMA\,(0,1,1)\times(1,0,0)_{12}$ model in their RTAs safety intervention planning activities due to its high forecast accuracy. Several avenues for further research can be pursued. First, use model comparison and ensemble techniques to compare SARIMA and Prophet against advanced counterparts such as machine learning techniques as in [65–67]. Researchers can also explore more intricate time series models, such as state space models, VARs, and Bayesian structural models, for enhanced predictive accuracy. Anomaly detection within accident data can be explored, identifying unusual patterns and sudden spikes, which can help inform the development of early warning systems and targeted intervention strategies for more proactive accident prevention. The government of Ghana is also urged to support institutions like the MTTU and NRSC in terms of recruiting qualified personnel and providing logistics and quality education on road traffic accident prevention for the citizenry. The fight against RTA is the responsibility of every citizen. This research should be used to develop effective road safety strategies and policies, evaluate and monitor road safety initiatives' progress, and identify improvement areas. To enhance road safety and mitigate accidents, multifaceted approaches must be employed in road accident-related research in Ghana. These interventions span diverse domains, including infrastructure enhancement through improved road design and pedestrian/cyclist facilities; stringent enforcement of traffic regulations encompassing speed limits, seat belt and helmet laws, and measures against impaired driving; fostering public awareness through road safety campaigns and educational programs integrated into school curricula; elevation of vehicle safety standards through mandatory incorporation of advanced safety features and rigorous crash testing; establishment of comprehensive accident databases for data-driven decision-making and identification of accident patterns; legal and regulatory reforms about liability, insurance, and penalty structures, to bolster road safety. By integrating these measures, we aim to reduce the burden of road accidents on society and create a safer road environment in a low-middle income country such as Ghana.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Additional information

No additional information is available for this paper.

CRediT authorship contribution statement

Edmund F. Agyemang: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Joseph A. Mensah:** Writing – original draft, Supervision, Project administration, Methodology, Investigation, Formal analysis. **Eric Ocran:** Writing – original draft, Visualization, Software, Methodology, Formal analysis. **Enock Opoku:** Writing – original draft, Software, Project administration, Investigation, Formal analysis. **Ezekiel N.N. Nortey:** Writing – review & editing, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data used to support the findings of this study are available from the corresponding author upon request and can also be assessed from the R codes available on GitHub repository at github.com/Agyemang1z/Road-Accidents.

Acknowledgement

The authors express their gratitude to the Motor Traffic and Transport Unit for granting access to the dataset utilized in this research. Additionally, they acknowledge the invaluable contributions of the anonymous reviewers and editors, whose insightful comments greatly enriched this work. The first and corresponding author also acknowledges the enormous support of the University of Texas Rio Grande Valley Presidential Research Fellowship fund.

References

[1] Ryan Collin, Yu Miao, Alex Yokochi, Prasad Enjeti, Annette Von Jouanne, Advanced electric vehicle fast-charging technologies, Energies 12 (10) (2019) 1839.
[2] Brian Caulfield, James Kehoe, Usage patterns and preference for car sharing: a case study of Dublin, Case Stud. Transp. Policy 9 (1) (2021) 253–259.
[3] Wenqi Zhou, Lisheng Fan, Fasheng Zhou, Feng Li, Xianfu Lei, Wei Xu, Arumugam Nallanathan, Priority-aware resource scheduling for uav-mounted mobile edge computing networks, IEEE Trans. Veh. Technol. (2023).

[4] Rodolfo I. Meneguette, R. De Grande, A.A. Loureiro, Intelligent Transport System in Smart Cities, Springer International Publishing, Cham, 2018.
[5] R. Kanthavel, S.K.B. Sangeetha, K.P. Keerthana, Design of smart public transport assist system for metropolitan city Chennai, Int. J. Intell. Netw. 2 (2021) 57–63.
[6] Wenqi Zhou, Junjuan Xia, Fasheng Zhou, Lisheng Fan, Xianfu Lei, Arumugam Nallanathan, George K. Karagiannidis, Profit maximization for cache-enabled vehicular mobile edge computing networks, IEEE Trans. Veh. Technol. (2023).
[7] Sampurna Mandal, Swagatam Biswas, Valentina E. Balas, Rabindra Nath Shaw, Ankush Ghosh, Motion prediction for autonomous vehicles from lyft dataset using deep learning, in: 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), IEEE, 2020, pp. 768–773.
[8] Ahmad M. Khalil, The genome editing revolution, J. Genet. Eng. Biotechnol. 18 (1) (2020) 1–16.
[9] Lasse Fridstrøm, Jan Ifver, Siv Ingebrigtsen, Risto Kulmala, Lars Krogsgård Thomsen, Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts, Accid. Anal. Prev. 27 (1) (1995) 1–20.
[10] Jingru Gao, Gary A. Davis, Using naturalistic driving study data to investigate the impact of driver distraction on driver's brake reaction time in freeway rear-end events in car-following situation, J. Saf. Res. 63 (2017) 195–204.
[11] Amrit Banstola, Julie Mytton, Cost-effectiveness of interventions to prevent road traffic injuries in low- and middle-income countries: a literature review, Traffic Inj. Prev. 18 (4) (2017) 357–362.
[12] Wubetie Habtamu Tilaye, Human Injury Causing Road Traffic Accident at Debre Markos Town, 2020.
[13] Thiri Ko, Analysis and forecasting of road traffic accidents in Yangon municipal area (2014-2018), PhD thesis, MERAL Portal, 2019.
[14] Debela Deme, Review on factors causes road traffic accident in Africa, J. Civ. Eng. Res. Technol. 1 (1) (2019) 1–8.
[15] P. Shunmuga Perumal, M. Sujasree, Suresh Chavhan, Deepak Gupta, Venkat Mukthineni, Soorya Ram Shimgekar, Ashish Khanna, Giancarlo Fortino, An insight into crash avoidance and overtaking advice systems for autonomous vehicles: a review, challenges and solutions, Eng. Appl. Artif. Intell. 104 (2021) 104406.
[16] Muhammad Hussain, Jing Shi, Zahara Batool, An investigation of the effects of motorcycle-riding experience on aberrant driving behaviors and road traffic accidents-a case study of Pakistan, Int. J. Crashworthiness 27 (1) (2022) 70–79.
[17] Huiying Wen, Yingxin Du, Zheng Chen, Sheng Zhao, Analysis of factors contributing to the injury severity of overloaded-truck-related crashes on mountainous highways in China, Int. J. Environ. Res. Public Health 19 (7) (2022) 4244.
[18] Williams Ackaah, Benjamin Aprimah Apuseyine, Francis K. Afukaar, Road traffic crashes at night-time: characteristics and risk factors, Int. J. Inj. Control Saf. Promot. 27 (3) (2020) 392–399.
[19] Gabriel Klaeger, Stories of the road: perceptions of power, progress and perils on the Accra-Kumasi road, Ghana, in: The Making of the African Road, Brill, 2017, pp. 86–115.
[20] Irenius Konkor, Moses Kansanga, Yujiro Sano, Roger Antabe, Isaac Luginaah, Community perceptions and misconceptions of motorcycle accident risks in the upper West region of Ghana, Travel Behav. Soc. 15 (2019) 157–165.
[21] Hilde Iversen, Risk-taking attitudes and risky driving behaviour, Transp. Res., Part F Traffic Psychol. Behav. 7 (3) (2004) 135–150.
[22] Hai-peng Shao, Juan Yin, Wen-hao Yu, Qiu-ling Wang, Aberrant driving behaviours on risk involvement among drivers in China, J. Adv. Transp. 2020 (2020) 1–8.
[23] Bekir Nihat Dogrul, Ibrahim Kiliccalan, Ekrem Samet Asci, Selim Can Peker, Blunt trauma related chest wall and pulmonary injuries: an overview, Chin. J. Traumatol. 23 (03) (2020) 125–138.
[24] T.L. Gunaruwan, P.C.J. Nayanalochana, Economic cost of human fatalities due to road traffic accidents in Sri Lanka: an estimation based on the human capital approach, J. South Asian Logist. Transp. 3 (1) (2023).
[25] Rannveig Svendby, Becoming the 'Other'a Qualitative Study of Power, Masculinities and Disabilities in the Lives of Young Drivers After Road Traffic Accidents, 2019.
[26] Kudzai Mwapaura, Witness Chikoko, Kudzai Nyabeze, Kwashirai Zvokuomba, Socio-economic challenges faced by persons with disabilities induced by road traffic accidents in Zimbabwe: the case of st giles rehabilitation centre, Harare, AfriFuture Res. Bull. (2021) 117.
[27] Geleta Mussa Yimer, Yonas Fissha Adem, Yosef Haile, Determinants of post-traumatic stress disorder among survivors of road traffic accidents in dessie comprehensive specialized hospital North-East Ethiopia, BMC Psychiatry 23 (1) (2023) 1–11.
[28] Sharaf Alkheder, Madhar Taamneh, Salah Taamneh, Severity prediction of traffic accident using an artificial neural network, J. Forecast. 36 (1) (2017) 100–108.
[29] C. Cabrera-Arnau, R. Prieto Curiel, S.R. Bishop, Uncovering the behaviour of road accidents in urban areas, R. Soc. Open Sci. 7 (4) (2020) 191739.
[30] Frances Agyapong, Thomas Kolawole Ojo, Managing traffic congestion in the Accra Central market, Ghana, J. Urban Manag. 7 (2) (2018) 85–96.
[31] Davies Adeloye, Jacqueline Y. Thompson, Moses A. Akanbi, Dominic Azuh, Victoria Samuel, Nicholas Omoregbe, Charles K. Ayo, The burden of road traffic crashes, injuries and deaths in Africa: a systematic review and meta-analysis, Bull. World Health Organ. 94 (7) (2016) 510.
[32] Pali Lehohla, Road Traffic Accident Deaths in South Africa, 2001-2006: Evidence from Death Notification, Statistics South Africa Pretoria, 2009.
[33] Paul J. Moroz, David A. Spiegel, The world health organization's action plan on the road traffic injury pandemic: is there any action for orthopaedic trauma surgeons?, J. Orthop. Trauma 28 (2014) S11–S14.
[34] Adnan A. Hyder, Nino Paichadze, Tamitza Toroyan, Margaret M. Peden, Monitoring the decade of action for global road safety 2011–2020: an update, Global Public Health 12 (12) (2017) 1492–1505.
[35] Margaret M. Peden, Prasanthi Puvanachandra, Looking back on 10 years of global road safety, Int. Health 11 (5) (2019) 327–330.
[36] Ahmad Salisu Aliyu, Nuru Yakubu Umar, Funmilayo Abubakar Sani, Hussaini Mohammed, Epidemiological study on the prevalence of road traffic accident and associated risk factors among drivers in bauchi state, Nigeria, Am. J. Surg. Clin. Case Rep. 3 (8) (2021) 1–8.
[37] Charles G. Manyara, Combating road traffic accidents in Kenya: a challenge for an emerging economy, in: Kenya After 50: Reconfiguring Education, Gender, and Policy, 2016, pp. 101–122.
[38] Muhammad Abdul Basit, Umar Farooq, Muhammad Imran, Nahid Fatima, Abdullah Alhushaybari, Sobia Noreen, Sayed M. Eldin, Ali Akgül, Comprehensive investigations of (Au-Ag/Blood and Cu-Fe3O4/Blood) hybrid nanofluid over two rotating disks: numerical and computational approach, Alex. Eng. J. 72 (2023) 19–36.
[39] Umar Farooq, Muhammad Abdul Basit, Sobia Noreen, Nahid Fatima, Abdullah Alhushaybari, Sayed M. El Din, Muhammad Imran, Ali Akgül, Recent progress in Cattaneo-Christov heat and mass fluxes for bioconvectional carreau nanofluid with motile microorganisms and activation energy passing through a nonlinear stretching cylinder, Ain Shams Eng. J. (2023) 102316.
[40] Muhammad Abdul Basit, Muhammad Imran, Shan Ali Khan, Abdullah Alhushaybari, R. Sadat, Mohamed R. Ali, Partial differential equations modeling of bio-convective sutterby nanofluid flow through paraboloid surface, Sci. Rep. 13 (1) (2023) 6152.
[41] M.A. Basit, Madeeha Tahir, Ayesha Riasat, S.A. Khan, Muhammad Imran, Ali Akgül, Numerical simulation of bioconvective Casson nanofluid through an exponentially permeable stretching surface, Int. J. Mod. Phys. B (2023) 2450128.
[42] Muhammad Imran, Sumeira Yasmin, Hassan Waqas, Shan Ali Khan, Taseer Muhammad, Nawa Alshammari, Nawaf N. Hamadneh, Ilyas Khan, Computational analysis of nanoparticle shapes on hybrid nanofluid flow due to flat horizontal plate via solar collector, Nanomaterials 12 (4) (2022) 663.
[43] Sandy Balkin, J. Keith Ord, Assessing the impact of speed-limit increases on fatal interstate crashes, J. Transp. Stat. 4 (1) (2001) 1–26.
[44] Abdul Hameed, Al Abbasi, Comparison Between Neural Networks and Traditional Statistical Methods to Predict the Number of Deaths Due to Traffic Accidents in Kuwait, 2005.
[45] Luke Fletcher, Alexander Zelinsky, Driver inattention detection based on eye gaze—road event correlation, Int. J. Robot. Res. 28 (6) (2009) 774–801.
[46] Angela Watson, Barry Watson, Kirsten Vallmuur, Estimating under-reporting of road crash injuries to police using multiple linked data collections, Accid. Anal. Prev. 83 (2015) 18–25.

[47] Mohammed Salifu, Williams Ackaah, Under-reporting of road traffic crash data in Ghana, Int. J. Inj. Control Saf. Promot. 19 (4) (2012) 331–339.
[48] Guy P. Nason, Stationary and non-stationary time series, Stat. Volcanol. 60 (2006).
[49] Keith William Hipel, Angus Ian McLeod, William C. Lennox, Advances in Box-Jenkins modeling: 1. Model construction, Water Resour. Res. 13 (3) (1977) 567–575.
[50] Clifford M. Hurvich, Chih-Ling Tsai, Regression and time series model selection in small samples, Biometrika 76 (2) (1989) 297–307.
[51] Mohan Mahanty, K. Swathi, K. Sasi Teja, P. Hemanth Kumar, A. Sravani, Forecasting the spread of Covid-19 pandemic with prophet, Rev. Intell. Artif. 35 (2) (2021) 115–122.
[52] Rajat Kumar Rathore, Deepti Mishra, Pawan Singh Mehra, Om Pal, Ahmad Sobri Hashim, Azrulhizam Shapi'i, T. Ciano, Meshal Shutaywi, Real-world model for bitcoin price prediction, Inf. Process. Manag. 59 (4) (2022) 102968.
[53] Sweeti Sah, B. Surendiran, R. Dhanalakshmi, Sachi Nandan Mohanty, Fayadh Alenezi, Kemal Polat, Forecasting Covid-19 pandemic using prophet, arima, and hybrid stacked LSTM-GRU models in India, Comput. Math. Methods Med. (2022) 2022.
[54] Karuna Gull, Suvarna Kanakaraddi, Ashok Chikaraddi, Covid-19 outbreak prediction using additive time series forecasting model, Trends Sci. 19 (22) (2022) 1919.
[55] Pousali Chakraborty, Marius Corici, Thomas Magedanz, A comparative study for time series forecasting within software 5g networks, in: 2020 14th International Conference on Signal Processing and Communication Systems (ICSPCS), IEEE, 2020, pp. 1–7.
[56] V.A. Turchenko, S.V. Trukhanov, V.G. Kostishin, F. Damay, F. Porcher, D.S. Klygach, M.G. Vakhitov, Dmitry Lyakhov, D. Michels, Bernat Bozzo, et al., Features of structure, magnetic state and electrodynamic performance of $SrFe_{12-x}In_xO_{19}$, Sci. Rep. 11 (1) (2021) 18342.
[57] Munirah A. Almessiere, Yassine Slimani, Norah A. Algarou, Maksim G. Vakhitov, Denis S. Klygach, Abdulhadi Baykal, Tatyana I. Zubar, Sergei V. Trukhanov, Alex V. Trukhanov, Hussein Attia, et al., Tuning the structure, magnetic, and high frequency properties of Sc-doped $Sr_{0.5}Ba_{0.5}Sc_xFe_{12-x}O_{19}/NiFe_2O_4$ hard/soft nanocomposites, Adv. Electron. Mater. 8 (2) (2022) 2101124.
[58] Lunyuan Chen, Lisheng Fan, Xianfu Lei, Trung Q. Duong, Arumugam Nallanathan, George K. Karagiannidis, Relay-assisted federated edge learning: performance analysis and system optimization, IEEE Trans. Commun. (2023).
[59] Sihui Zheng, Cong Shen, Xiang Chen, Design and analysis of uplink and downlink communications for federated learning, IEEE J. Sel. Areas Commun. 39 (7) (2020) 2150–2167.
[60] Junjuan Xia, Lisheng Fan, Wei Xu, Xianfu Lei, Xiang Chen, George K. Karagiannidis, Arumugam Nallanathan, Secure cache-aided multi-relay networks in the presence of multiple eavesdroppers, IEEE Trans. Commun. 67 (11) (2019) 7672–7685.
[61] Nemanja Deretić, Dragan Stanimirović, Mohammed Al Awadh, Nikola Vujanović, Aleksandar Djukić, Sarima modelling approach for forecasting of traffic accidents, Sustainability 14 (8) (2022) 4403.
[62] Dung David Chuwang, Weiya Chen, Forecasting daily and weekly passenger demand for urban rail transit stations based on a time series model approach, Forecasting 4 (4) (2022) 904–924.
[63] Tianyu Feng, Zhou Zheng, Jiaying Xu, Minghui Liu, Ming Li, Huanhuan Jia, Xihe Yu, The comparative analysis of sarima, Facebook prophet, and lstm for road traffic injury prediction in northeast China, Front. Public Health 10 (2022).
[64] Mohammmad Asifur Rahman Shuvo, Muhtadi Zubair, Afsara Tahsin Purnota, Sarowar Hossain, Muhammad Iqbal Hossain, Traffic forecasting using time-series analysis, in: 2021 6th International Conference on Inventive Computation Technologies (ICICT), IEEE, 2021, pp. 269–274.
[65] Yinghao Guo, Rui Zhao, Shiwei Lai, Lisheng Fan, Xianfu Lei, George K. Karagiannidis, Distributed machine learning for multiuser mobile edge computing systems, IEEE J. Sel. Top. Signal Process. 16 (3) (2022) 460–473.
[66] Le He, Lisheng Fan, Xianfu Lei, Xiaohu Tang, Pingzhi Fan, Arumugam Nallanathan, Learning-based mimo detection with dynamic spatial modulation, IEEE Trans. Cogn. Commun. Netw. (2023).
[67] Shunpu Tang, Lunyuan Chen, Ke He, Junjuan Xia, Lisheng Fan, Arumugam Nallanathan, Computational intelligence and deep learning for next-generation edge-enabled industrial IoT, IEEE Trans. Netw. Sci. Eng. (2022).