Engineering and Technology Journal e-ISSN: 2456-3358

Volume 09 Issue 12 December-2024, Page No.- 5650-5658

DOI: 10.47191/etj/v9i12.13, I.F. – 8.227

A Short-Term Prediction Model for the Number of Registered Motor Vehicles Using Facebook Prophet Forecasting Approach

Paul Boye[1], Yao Yevenyo Ziggah[2], Kofi Agyarko[3]

[1,3]Faculty of Engineering, Department of Mathematical Sciences, University of Mines and Technology (Ghana)

[2]Faculty of Mineral Resources Technology, Department of Geomatic Engineering, University of Mines and Technology (Ghana)

ABSTRACT: Accurate prediction of the total annual number of motor vehicles to be registered in a country is very important to the government because it will help Driver and Vehicle Licensing Authority (DVLA) policymakers to put good road safety measures in place to make their work successful. This study assessed four standard empirical time series methods for the prediction of the total number of motor vehicles to be registered. The evaluated methods were Facebook Prophet, grey, SARIMA, and ARIMA. In developing the time series model, 70% of the data set obtained from DVLA of Ghana was used and the remaining 30% was used to validate the model's forecasting adequacy. The robustness of the Facebook Prophet, grey, SARIMA, and ARIMA was assessed using Normalised Root Mean Square Error (NRMSE), Correlation Coefficient (R), Variance Accounted For (VAF), and Performance Index (PI). The analytical results revealed that the Facebook Prophet time series model had the best results with the lowest NRMSE (0.2176) and highest R (0.8229), VAF (66.6475%), and PI (1.1260) values. The study concluded that the Facebook Prophet forecasting approach could be useful to vehicle licensing authority managers and policymakers to give accurate and timely information on future projections of the number of motor vehicles to be registered.

KEYWORDS: Time Series Models, Accurate Prediction, Driver and Vehicle Licensing Authority

1. INTRODUCTION

With the increase in the world's population, urbanisation has become a dominantly contemporary process. Because of the population rise, urban transportation which originated from both trading social recreational services and goods in urban areas has significantly increased. This affects urban construction and the size of development dearly (Kordani et al., 2019; Jameel and Evdorides, 2019).

Recently, most urban environments are affected by high volumes of motor traffic as well as air and noise pollution. As a result of these congestions, large sums of money are spent annually to cater for these menaces caused by pedestrians and motorists. The loss of people's time in traffic due to these bottlenecks which is causing productivity losses to the country and travel stress, the excessive consumption of fuel in long queues resulting in a significant contribution to global climate change, and the burden of diseases affect the efficient operation of the economy. Thus, the annual environmental impacts of pedestrian and motor traffic factors are having a great tow on nations and damaging their economies of societies as well (Shomar, 2019).

It must be noted that as the urban population keeps rising, the percentage of people using various methods of transportation, such as motor vehicles, has increased greatly, but roads are the same in terms of infrastructure improvements (Qasim et al., 2019). Due to these rapid increases in the volume of motor vehicles on urban streets, accurate determination of the quantity of these vehicles needed to be registered annually is very crucial. Although knowing the number of these vehicles will help organisations with capacity planning and goal setting, determining such numbers in advance is still a major research challenge. Despite its importance to management and policymakers, numerous problems are encountered in modeling high-quality and reliable prediction models when dealing with a diversity of researchers who have numerous skills in time series modeling and time series methods (Huang et al., 2018; Taylor and Letham 2018; Fullerton et al., 2015; Saini 2008; Wei et al., 2019).

Facebook Prophet Methodology is a relatively new time series forecasting tool that provides features from both generalised linear models and additive models. The Prophet is mainly an extension of the generalised linear model using nonlinear smoothing functions. The advantage of the Prophet method over the diversity of statistical and time series forecasting methods such as autoregressive integrated moving average (ARIMA), seasonal autoregressive integrated moving average (SARIMA), fractional autoregressive integrated moving average (FARIMA), and grey is the approach of the analyst-in-the-loop which allows users to apply their domain knowledge about the dataset to the forecasting algorithm without having any knowledge of the statistical methods working from within. Though it is simple to use, yet robust for estimation due to its structure of adjusting parameters without investigating the details of the original model (Weytjens et al., 2019; Yenidoğan et al., 2019). This Prophet approach takes advantage of both statistical and judgmental forecasting; where the latter is the forecasting method based on human expert decisions. In addition, the easily implemented Prophet can generate reasonable quality forecasts at scale; and it works efficiently with time series having several seasons of historical datasets with strong seasonal effects. The robustness of the Prophet to shifts in trends in a dataset and missing data cannot be overlooked. Furthermore, as a completely automated prediction technique, it can adequately handle large outliers in data analysis (Stefenon et al. 2023).

In line with its versatility and accurate forecasting ability have led to its application by researchers in a great number of scientific areas such as medicine (Hyun et al., 202), energy (Almazrouee et al., 2020), hydrology (Aguilera et al., 2019), meteorology (Shen et al., 2020).

In view of the literature reviewed relating to this study, the authors explored for the first time the Prophet forecasting approach to assess its predictive strength in the number of registered motor vehicles which has been given no scholarly attention. The proposed Prophet forecasting method was tested using an acquired dataset from the Driver and Vehicle Licensing Authority (DVLA) of Ghana. Analytical results showed that the Prophet model performance is good and managers and policymakers can rely on it to make informed decisions.

2. DATA USED

In developing the model, a monthly dataset of registered vehicles of all categories was obtained from DVLA, Accra, Ghana. The five years data spans from 2010 to 2014. A descriptive record of the Total Registered Motor Vehicles (TRMV) is presented in Table 1.

Table 1 Summary Statistical Records of Registered Motor Vehicles

Parameter	Data Size	Max	Min	Mean	SD
TRMV	60	27 101	2 020	11 870	5 150.428

2.1. Data Pretreatment

R statistical software version 3.6.1 was used for the preliminary investigation of the TRMV data (Table 1). In Table 1, the standard deviation (SD) is a measure of how far the data set values are from their mean value ($\bar{y}$). Here, the standard deviation is 5150.428 (Equation (1)) and it shows a spread out of the registered motor vehicles data set from its mean value of 11 870 (Equation (2)). This is an indication that the TRMV data set exhibits a larger variation or spread.

Figure 1 shows the entire data set used for the statistical analysis.

$$SD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \qquad (1)$$

$$\bar{y} = \sum_{i=1}^{n}\frac{y_i}{n} \qquad (2)$$

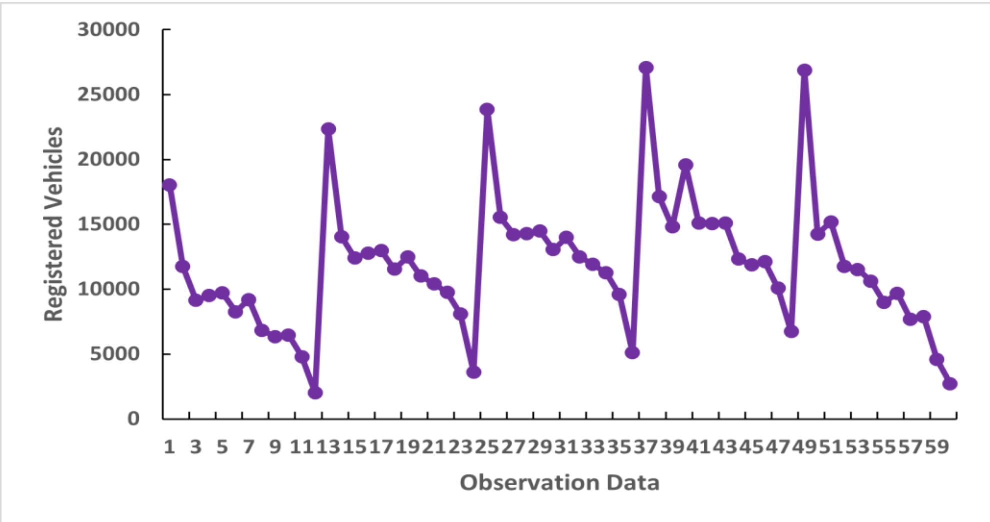

Figure 1: Graphical representation of data set for the statistical analysis

3. OVERVIEW OF METHODS APPLIED

3.1 Facebook Prophet Prediction Approach

The Facebook Prophet approach is a time series model that can be decomposed into three key components comprising: seasonality, holidays, and trend. As a regression model with explainable parameters that function efficiently with its default values, it allows the user to automatically select the components that concern the prediction problem and apply the required adjustments easily (Almazrouee et al., 2020; Papastefanopoulos et al., 2020). The basic Prophet model (Taylor and Letham, 2018) is given in Equation (3) as

$$J(\tau) = s(\tau) + g(\tau) + h(\tau) + \varepsilon_\tau \quad (3)$$

Here, $s(\tau)$ is the seasonality which includes the periodic changes such as weekly, yearly, seasonal, etc., $g(\tau)$ is the trend function used to fit aperiodic changes in the time series, $h(\tau)$ is the additional regressor that can be included within specific points of the dataset, and ε_τ represent the noise term. That is, any distinctive features of the dataset that the model did not fit.

The trend function is mathematically given in Equation (4) as

$$g(\tau) = \left(k + a(\tau)^T \delta\right)\tau + \left(m + a(\tau)^T \gamma\right) \quad (4)$$

where δ is the adjustment rate, k is the growth rate, m is the offset parameter, and γ is the trend change points (s_j), conventionally set as $-s_j\delta_j$ and $a(\tau)$ is given in Equation (5).

$$a_j(\tau) = \begin{cases} 1, \ if \ t \geq s_j \\ 0, \ otherwise \end{cases} \quad (5)$$

Based on previous experience, the change points permit the analyst-in-the-loop to modify the forecast value. Consequently, the forecast trend is fine-tuned and the forecast results become improved. The adjustment at change point j can be computed using Equation (6)

$$\gamma_j = \left(s_j - m - \sum_{l<j} \gamma_l\right)\left(1 - \frac{k + \sum_{l<j} \delta_l}{k + \sum_{l \leq j} \delta_l}\right) \quad (6)$$

If P is the regular seasonal period, the seasonality function $s(\tau)$, which is modeled by the Fourier series to indicate yearly, weekly, and daily seasonality is expressed in Equation (7).

$$s(\tau) = \sum\left(a_n \cos\left(\frac{2\pi n\tau}{P}\right) + b_n \sin\left(\frac{2\pi n\tau}{P}\right)\right) \quad (7)$$

For fitting $s(\tau)$, seasonality vectors matrix is created for parameter estimation for each τ of the series as shown in Equation (8).

$$s(\tau) = X(\tau)\beta \quad (8)$$

where $\beta = [a_1 b_1, \ldots, a_N b_N]^T$ is 2N parameters to be estimated and $\beta \sim Normal(0, \sigma^2)$ for seasonality smoothing parameter.

The Prophet methodology permits the addition of other regressors to enhance the forecast results. For example, defining $Z(\tau)$ (Equation (9)) as a matrix of regressors made up of a list of holiday dates that can be incorporated into the Prophet model by using the function $h(\tau)$ as shown in Equation (10).

$$Z(\tau) = \left[1(\tau \in D_1), \ldots, 1(\tau \in D_L)\right] \quad (9)$$

$$h(\tau) = Z(\tau)k \quad (10)$$

Where $k \sim \text{Normal}(0, \nu^2)$ and ν is an extra regressor smoothing parameter.

3.2 ARIMA Model

The ARIMA model, is a popular complex stochastic linear model for time series prediction with enormous flexibility. The model comprises three parts: autoregression (AR) (Equation (11)) and moving average (MA) (Equation (12)) models with integration (I). In studies, the model is generally known as ARIMA (p, d, q) where p is the AR part order, d is the nonseasonal differences order, and q is the MA part order (Ghimire, 2017; Zhang et al. 2014).

$$y_\tau = a + \sum_{i=1}^{p} \phi_i y_{\tau-i} + e_\tau \quad (11)$$

$$y_\tau = b + e_\tau + \sum_{i=1}^{q} \varphi_i e_{\tau-i} \quad (12)$$

Here, ϕ_i and φ_i are the coefficients of the AR and MA models, respectively. The a and e_τ (Equations (11 and 12)) are a constant for the AR model and a white noise for the models. The b and $e_{\tau-i}$ (Equation (12)) are the expectation which is often assumed to be zero and white noise error terms for the MA model.

But studies have shown that the combination of the AR and MA models is considered as the generalisation of autoregression moving average model as shown in Equation (13).

$$y_\tau = a + e_\tau + \sum_{i=1}^{p} \phi_i y_{\tau-i} + \sum_{i=1}^{q} \varphi_i e_{\tau-i} \quad (13)$$

3.3 SARIMA Prediction Model

In Hyndman and Athanasopoulos (2018), Equation (13) gives rise to models of SARIMA (Equation (14)) when non-seasonal terms are incorporated into it.

$$\text{SARIMA}\ (p,\ d,\ q)\ (P, D, Q)\ [o] \quad (14)$$

where the observations per year is denoted as o. This is generated as follows:

Let the Backshift operator be K. $Kx_\tau \equiv x_{\tau-1}$

$K^2 x_t = KKx_\tau = Kx_{\tau-1} = x_{\tau-2} \vdots\; K^d x_\tau = x_{\tau-d}$

Let ∇ be non-seasonal differencing operator. Define $\nabla \equiv (1-K)$

$$\nabla x_\tau = (1-K)x_\tau = x_\tau - x_{\tau-1}$$

$$\nabla^2 x_\tau = (1-K)^2 x_\tau = (1-2K+K^2)x_\tau = y_\tau - 2x_{\tau-1} + x_{\tau-2}$$

$$\vdots\;\nabla^d x_\tau = (1-K)^d x_\tau = \left(\sum_{i=0}^{d}\binom{d}{i}(-K)^{d-i}\right)x_\tau;$$

from the Binomial Theorem.

Let the seasonal differencing operator be ∇_s (s is season period). Define $\nabla_s \equiv (1-K^s)$

$$\nabla_s x_\tau = (1-K^s)x_\tau = x_\tau - x_{\tau-s}$$

$$\nabla_s{}^2 x_\tau = (1-K^s)^2 x_\tau = (1-2K^s+K^{2s})x_\tau = x_\tau - 2x_{\tau-s} + x_{\tau-2s}$$

$$\vdots\;\nabla_s{}^D x_\tau = (1-K^s)^D x_\tau = \left(\sum_{i=0}^{D}\binom{D}{i}(-K^s)^{D-i}\right)x_\tau$$

Non-seasonal autoregressive model, AR (p), is given in Equation (15).

$$x_\tau = a_0 + \sum_{i=1}^{p} a_i x_{\tau-i} + \varepsilon_\tau \tag{15}$$

$$(1-a_1K-a_2K^2-\ldots-a_pK^p)x_\tau = a_0 + \varepsilon_\tau$$

$$\phi_p(B)x_\tau = a_0 + \varepsilon_\tau \tag{16}$$

where $\phi_p(K) = (1-a_1K-a_2K^2-\ldots-a_pK^p)$

Seasonal autoregressive model, SAR (p), is given in Equation (17).

$$x_\tau = A_0 + \sum_{i=1}^{p} A_i x_{\tau-i} + \varepsilon_\tau \tag{17}$$

$$(1-A_1K^s-A_2K^{2s}-\ldots-A_pK^{ps})x_t = A_0 + \varepsilon_\tau$$

$$\Theta_p(K^s)x_\tau = A_0 + \varepsilon_\tau \tag{18}$$

Where $\Theta_p(K^s) = (1-A_1K^s-A_2K^{2s}-\ldots-A_pK^p)$

Non-seasonal moving average model, MA (q), is given in Equation (19).

$$x_\tau = b_0 + \sum_{j=1}^{q} b_j\varepsilon_{\tau-j} + \varepsilon_\tau \tag{19}$$

$$= b_0 + (1-b_1K-b_2K^2-\ldots-b_qK^q)\varepsilon_\tau$$

$$= b_0 + b_q(B)\varepsilon_\tau \tag{20}$$

Seasonal moving average model, SMA (Q), is given in Equation (21).

$$x_\tau = K^s{}_0 + \sum_{j=1}^{Q} K^s{}_j\varepsilon_{\tau-j} + \varepsilon_\tau \tag{21}$$

$$= K^s{}_0 + (1-K_1K^s-K_2K^{2s}-\ldots-K_QK^{Qs})\varepsilon_\tau$$

$$= K_0 + K_Q(K^s)\varepsilon_\tau \tag{22}$$

Where

$\Phi_Q(K^s) = (1-K_1K^s-K_2K^{2s}-\ldots-K_QK^{Qs})$

Let I(d) be the non-seasonal differencing model.

$$\nabla^d x_\tau = a_0 + \varepsilon_\tau$$

$$(1-K)^d x_\tau = a_0 + \varepsilon_\tau \tag{23}$$

Seasonal differencing model, SI(D), is given in Equation (24).

$$\nabla_s{}^D x_\tau = b_0 + \varepsilon_\tau$$

$$(1-K^s)^D x_\tau = b_0 + \varepsilon_\tau \tag{24}$$

Therefore SARIMA $(p,\ d,\ q)\ (P, D, Q)$ model is given as Equation (25).

$$\phi_p(K)\ \Theta_p(K^s)\ \nabla^d\ \nabla_s{}^D x_\tau = b_0 + \varphi_q(K)\ \Phi_Q(K^s)\varepsilon_\tau \tag{25}$$

3.4 Grey Prediction Model

The GM (1, 1) predicting model is one of the essential models in the grey theory system employed in time series data analysis with fewer sample (Baloochian and Balochian, 2020). One of the grey system theory's most important features is the use of accumulative generation operation (AGO) to minimise randomness of data. The AGO approach efficiently removes noise by transforming random time series data into a monotonically increasing sequence that quickly assess systematic regularity (Islam et al., 2021).

By considering an input dataset, the construction of the GM (1,1) model is as follows (Ma and Li, 2016; Yuan and Chen, 2016).

Step 1: Consider the initial sequence of dataset with n entries as expressed in Equation (26).

$$U^{(0)}(1,n)=\left\{u^{(0)}(1),\dots,u^{(0)}(j),\dots,u^{(0)}(n)\right\} \tag{26}$$

where $u^{(0)}(j)$ is time j time series dataset; and n is greater than or equal to 4. $U^{(0)}$ represents the non-negative initial time series dataset.

Step 2: Construct $U^{(1)}$ (Equation (27)) using a one-time accumulated generation operator (1-AGO), namely

$$U^{(1)}(1,n)=\left\{u^{(1)}(1),u^{(1)}(2),\dots,u^{(1)}(n)\right\} \tag{27}$$

where $u^{(1)}(k)=\sum_{j=1}^{k}u^{(0)}(j),\ k=1,\ \dots\ ,\ n.$ (28)

and $u^{(1)}(1)=u^{(0)}(1)$ (29)

Step 3:

The grey model GM (1,1) first-order differential equation is given as in Equation (30).

$$\frac{d\hat{u}^{(1)}}{dt}+x\hat{u}^{(1)}=y \tag{30}$$

where $\hat{u}^{(1)}(1)=u^{(0)}(1)$ is the initial condition, $\hat{u}$ denotes the predicted grey value, y is the grey input, and x is the developing coefficient.

Step 4: By discretising Equation (31), model parameters are computed as

$$\frac{d\hat{u}^{(1)}}{dt}=\lim_{\Delta t\to 0}\frac{\hat{u}^{(1)}(t+\Delta t)-\hat{u}^{(1)}(t)}{\Delta t} \tag{31}$$

As $\Delta t\to 1$, the approximated forecast value is

$$\frac{d\hat{u}^{(1)}}{dt}\cong u^{(1)}(k+1)-u^{(1)}(k) \tag{32}$$

$$=u^{(0)}(k+1),\ \ k=1,\ \dots\ ,\ n. \tag{33}$$

And $u^{(1)}(t)\cong pu^{(1)}(k)+(1-p)u^{(1)}(k+1)$ (34)

$$=z^{(0)}(k+1),\ \ k=1,\dots \tag{35}$$

where p, the production coefficient, ranges between zero and one. Conventionally, it is considered to be 1/2. The source model is given as Equation (36).

$$u^{(0)}(k)+xz^{(1)}(k)=y,\ \ k=1,\ \dots\ ,\ n. \tag{36}$$

By the least squares method, the model parameters x and y can be solved using Equation (37).

$$[x,y]^{\mathrm{T}}=\left(\mathrm{Z}^{\mathrm{T}}Z\right)^{-1}\mathrm{Z}^{\mathrm{T}}\mathrm{Y} \tag{37}$$

where Z and Y are defined as Equation (38).

$$Z=\begin{bmatrix}-z^{(1)}(2) & 1\\ -z^{(1)}(3) & 1\\ \vdots & \vdots\\ -z^{(1)}(n) & 1\end{bmatrix},\qquad \mathrm{Y}=\begin{bmatrix}u^{(0)}(2)\\ u^{(0)}(3)\\ \vdots\\ u^{(0)}(n)\end{bmatrix} \tag{38}$$

The generated mean sequence $Z^{(1)}$ of $U^{(1)}$ is defined as Equation (39).

$$Z^{(1)}=\left\{z^{(1)}(1),\ \dots\ ,\ z^{(1)}(n)\right\} \tag{39}$$

where $z^{(1)}(k)$ is the mean value of adjacent data of $u^{(1)}(k)$. That is

$$z^{(1)}(k)=\frac{1}{2}\left(u^{(1)}(k-1)+u^{(1)}(k)\right);k=2,\dots,n. \tag{40}$$

Step 5: Substituting Equation (40) into Equation (36), together with an initial condition result in Equation (41).

$$\hat{U}^{(1)}(k+1)=\left(u^{(0)}(1)-\frac{y}{x}\right)e^{-xk}+\frac{y}{x} \tag{41}$$

where $\hat{U}^{(1)}(k+1)$ denotes the prediction of U at time point $(k+1)$.

As a result, the prediction output at step k can be estimated as Equation (42) using IAGO.

$$\hat{u}^{(0)}(j+1)=\left(1-e^{x}\right)\left[u^{(0)}(1)-\frac{y}{x}\right]e^{-xj} \tag{42}$$

Therefore, the fitted GM (1, 1) sequence is given as in Equation (43).

$$\hat{u}^{(0)}=\left\{\hat{u}^{(0)}(1),\hat{u}^{(0)}(2),\dots,\hat{u}^{(0)}(n),\dots\right\} \tag{43}$$

and $\hat{u}^{(0)}(1)=u^{(0)}(1)$ for $k\geq 0$

The forecasted sequence is given as in Equation (44).

$$\left\{\hat{u}^{(0)}(n+1),\hat{u}^{(0)}(n+2),\dots\right\} \tag{44}$$

4. PERFORMANCE ASSESSMENT EVALUATORS

To determine the efficiency and robustness of the Facebook Prophet prediction approach, the method was compared with grey, SARIMA and ARIMA using statistical performance indicators. These applied indicators include variance accounted for (VAF), Performance Index (PI), normalised root mean square error (NRMSE), and correlation coefficient (R). The respective definitions of VAF, PI, NRMSE, and R are given in Equations (45) to (49) (Shen et al., 2020; Wei et al., 2019; Huang et al., 2018; Akoglu, 2018; Tabachnick and Fidell, 2007):

$$VAF=\left[1-\frac{Var\left(y_i-\hat{y}_i\right)}{Var\left(y_i\right)}\right]\times 100\% \qquad (45)$$

$$PI=\left[R^2+\left(\frac{VAF}{100}\right)\right]-NRMSE \qquad (46)$$

with

$$R^2=1-\frac{\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i\right)^2-\frac{\left(\sum_{i=1}^{n}y_i\right)^2}{n}} \qquad (47)$$

$$NRMSE=\frac{RMSE}{y_{Max}-y_{Min}} \qquad (48)$$

$$R=\frac{n\sum(ty)-\left(\sum t\right)\left(\sum y\right)}{\sqrt{\left[n\sum t^2-\left(\sum t\right)^2\right]\left[n\sum y^2-\left(\sum y\right)^2\right]}} \qquad (49)$$

Where y is TRMV, $\bar{y}$ is the TRMV mean value, t is duration, y_i and t_i are the ith bivariate data set, and n is the data size.

5. RESULTS AND DISCUSSION

5.1. Developed Model Efficiency Test

During the model development for the Prophet, ARIMA, SARIMA and grey, the acquired 60 data set of vehicles of all categories was partitioned by the extensively used hold-out cross-validation method into 70% training set (42 data points) for model development and the remaining 30% as testing set (18 data point) served as an independent data to verify the forecasting robustness of the methods. Figure 2 shows the training and the testing datasets employed to develop and validate the model, respectively.

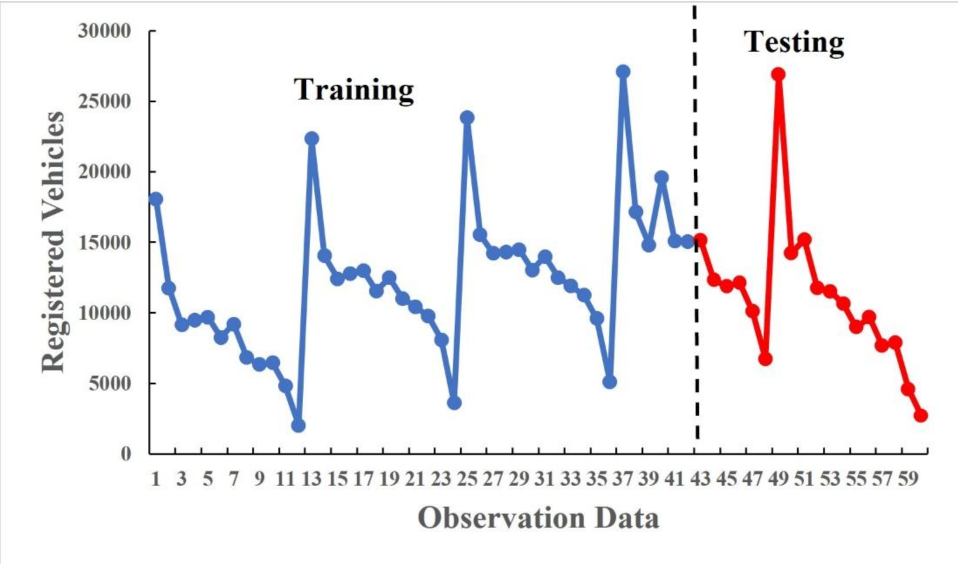

Figure 2: Graphical representation of the training and testing data

The prediction of the registered motor vehicles using each competing model was analysed by considering the errors between the actual and predicted values which was accurately Quantified by using the following statistical performance metrics: NRMSE, R, VAF, and PI. Table 2 shows the test data results for NRMSE, R, VAF, and PI for Prophet, grey, SARIMA and ARIMA respectively.

Table 2 Optimal Model Test Prediction Accuracy

Method	NRMSE	R	VAF (%)	PI
Prophet	0.2176	0.8229	66.6475	1.1260
Grey	0.4226	0.5538	45.1461	0.3356
SARIMA	0.2353	0.7915	60.9491	1.0007
ARIMA	0.2558	0.5404	2.9862	0.0661

In Table 2, the NRMSE values portray the error quantity a developed model could not explain. From Table 2, it can clearly be seen that the Facebook Prophet approach had the least NRMSE value of 0.2176 as compared with the other methods. This means that, the Prophet approach possessed a predictive strength of 99.7824 %.

R is a statistic employed to measure the strength of linear relationship that exists between the actual and the predicted registered motor vehicles. Invariably, it portrays the model prediction accuracy and the larger its value, the better the agreement between the actual and the predicted outputs. From Table 2, the Prophet approach reported a strong positive correlation (R = 0.8229) between the actual and the predicted registered motor vehicles. Thus, it reflects to the degree of closeness of the model predictions to the actual data set; and this is statistically significant.

From Table 2, the percentage of variance accounted for by the Prophet approach (66.6475%) was higher than any of the other methods employed. Thus, the VAF achieved by the Prophet approach (66.6475%) is an indicator of the quality or explanatory power of the Prophet. The 66.6475% is acceptable since the goal is to explain as much of the variance as possible in the data set.

Interpretation of the PI values in Table 2 show that the Prophet approach is superior to SARIMA, ARIMA and grey. This is because a method with a PI value approaching 2 considered more efficient. Considering that the Prophet PI value of 1,1260 is the highest among the other methods, it can be stated that the Prophet predictions are in close association with the actual data set than the other methods as can also be seen in Figure 3.

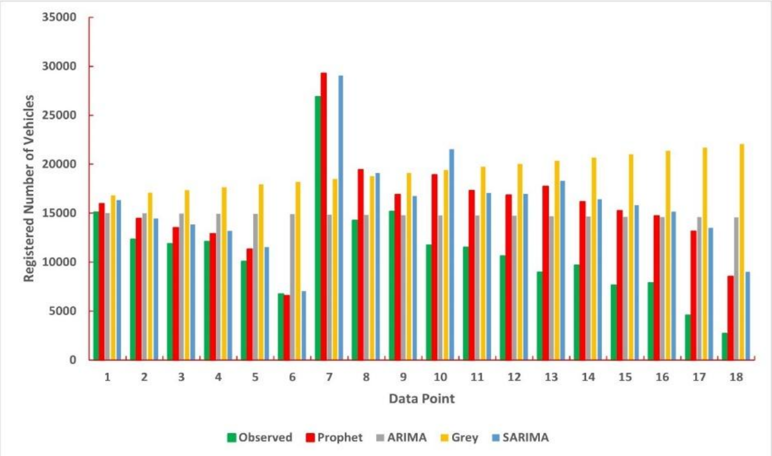

Figure 3: Line graph of test predictions and actuals

Figure 4 shows a graphical illustration of the VAF results.

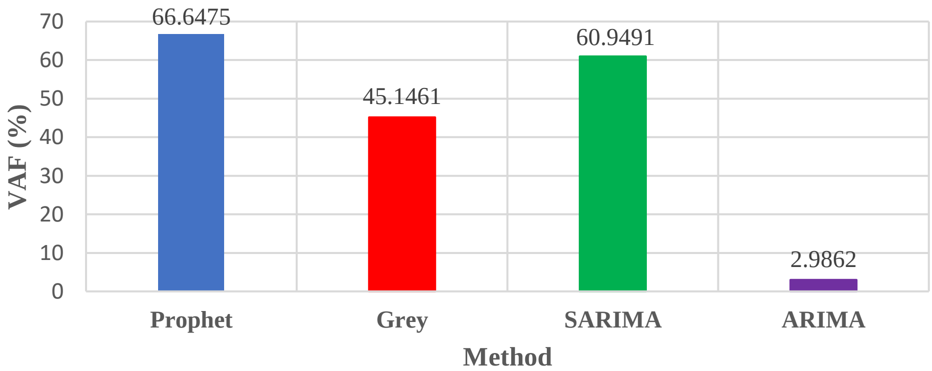

Figure 4: TRMV Models Validation using VAF

Figure 5 shows a graphical illustration of the model validation results based on PI.

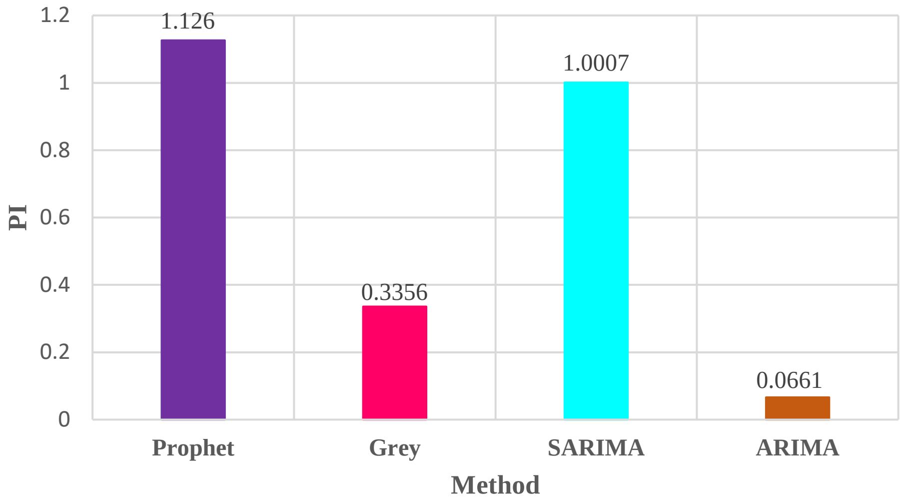

Figure: 5 TRMV models validation using PI

Boah-Mensah (2013) and Agunbiade and Peter (2013) advocated that the solution to the pedestrian and bicyclists' fatalities problem would be resolved if DVLA policymakers and service providers would have exhaustive information on the annual additional number of vehicles deemed for registration in the country. In that regard, management and policymakers could depend on the proposed Prophet model for making estimations. This would aid them in making well-informed organisational decisions and proper arrangements for good road infrastructure to be put in place to eradicate the challenges pedestrians and bicyclists suffer.

This study will not only be beneficial to researchers who need to model and estimate the number of motor vehicles a country needs to register, but will also benefit policymakers, road transport managers as well as the government to put in place good road infrastructures. This will help to eradicate the vulnerable road user crashes of pedestrians and bicyclists' fatalities with vehicles. It can be concluded that the presented paper can be useful for researchers who need to model and predict annual number of motor vehicles using Prophet Method.

6. CONCLUSIONS

In this study, we have developed time series models useful to predict the number of registered motor vehicles in a country using four main methods of interest: Facebook Prophet, SARIMA, ARIMA, and grey methods. The aim of the study was to examine the models' predictive capabilities using the statistical performance metrics: NRMSE, R, VAF, and PI.

Based on the test data, the robustness of the models was assessed and results showed that the Facebook Prophet approach was the most efficient model useful to predict the number of registered motor vehicles because of its low NRMSE (0.2176) value and high R (0.8229), VAF (66.6475%), and PI (1.1260) values.

REFERENCES

1. Aguilera, H., Guardiola-Albert, C., Naranjo-Fernández, N., and Kohfahl, C. (2019). Towards Flexible Groundwater-Level Prediction for Adaptive Water Management: Using Facebook's Prophet Forecasting Approach. Hydrological Sciences Journal, 64(12), 1504-1518.
2. Agunbiade, D. A. and Peter, E. N. (2013). Modeling and Forecasting Vehicle Registration System: An ARMA Approach. Society for Mathematical Services and Standards, 2(1), 1-13.
3. Akoglu, H. (2018). User's Guide to Correlation Coefficients. Turkish Journal of Emergency Medicine. 18, 91-93. https://doi.org/10.1016/j.tjem.2018.08.001.
4. Almazrouee, A. I., Almeshal, A. M., Almutairi, A. S., Alenezi, M. R., and Alhajeri, S. N. (2020). Long-Term Forecasting of Electrical Loads in Kuwait Using Prophet and Holt-Winters Models. Applied Sciences, 10(16), 5627.
5. Balochian, S. and Baloochian, H. (2020). Improving Grey Prediction Model and its Application in Predicting the Number of Users of a Public Road Transportation System. Journal of Intelligent Systems, 30(1), 104-114.
6. Boah-Mensah, E. (2013). The Number of Cars in Ghana Increased by 23%. Accessed: March 19, 2015, www.vehicle population in Ghana.
7. Fullerton, T. M., Novela, G., Torres, D., and Walke, A. G. (2015). Metropolitan Econometric Electric Utility Forecast Accuracy. International Journal of Energy Economics and Policy, 5(3), 738-745.
8. Ghimire, B. N. (2017). Application of ARIMA Model for River Discharges Analysis. Journal of Nepal Physical Society, 4(1), 27-32.

9. Huang, J., Tang, Y., and Chen, S. (2018). Energy Demand Forecasting: Combining Cointegration Analysis and Artificial Intelligence Algorithm. Mathematical Problems in Engineering, 1-13. https://doi.org/10.1155/2018/5194810.
10. Hyndman, J. R., and Athanasopoulos, G. (2018). Forecasting: Principles and Practice.
11. Hyun, J., Lee, S. H., Son, H. M., Park, J. U., and Chung, T. M. (2020). A Synthetic Data Generation Model for Diabetic Foot Treatment. In Future Data and Security Engineering. Big Data, Security and Privacy, Smart City and Industry 4.0 Applications: 7th International Conference, FDSE 2020, Quy Nhon, Vietnam, Proceedings 7, Springer Singapore, 249-264.
12. Islam, M. R., Ali, S. M., Fathollahi-Fard, A. M., and Kabir, G. (2021). A Novel Particle Swarm Optimisation-Based Grey Model for the Prediction of Warehouse Performance. Journal of Computational Design and Engineering, 8(2), 705-727.
13. Jameel, A. K., and Evdorides, H. (2019). Review of the Road Crash Data Availability in Iraq. International Journal of Transport and Vehicle Engineering, 13(9), 521-526.
14. Kordani, A. A., Rooyintan, M., and Boroomandrad, S. M. (2019). Economical and Technical Analysis of Urban Transit System Selection Using TOPSIS Method According to Constructional and Operational Aspects. International Journal of Transport and Vehicle Engineering, 13(7), 398-402.
15. Ma, D., and Li, J. (2016). Two Novel Grey System Models and their Applications on Landslide Forecasting. Journal of Control Sciences and Engineering, 2016, 1-6.
16. Papastefanopoulos, V., Linardatos, P., and Kotsiantis, S. (2020). COVID-19: A Comparison of Time Series Methods to Forecast the Percentage of Active Cases Per Population. Applied Science, 10, 1-15. https://doi.org/10.3390/app10113880.
17. Qasim, T., Alqawasmi, M., Hawash, M., Betar, M., and Qasim, W. (2019). Modeling Jordan University of Science and Technology Parking Using Arena Program. International Journal of Transport and Vehicle Engineering, 13(6), 324-328.
18. Saini, L. M. (2008). Peak Load Forecasting Using Bayesian Regularization, Resilient and Adaptive Backpropagation Learning-Based Artificial Neural Networks. Electric Power Systems Research, 78(7), 1302–1310. https://doi.org/10.1016/j.epsr.2007.11.003.
19. Shen, J., Valagolam, D., and McCalla, S. (2020). Prophet Forecasting Model: A Machine Learning Approach to Predict the Concentration of Air Pollutants ($PM_{2.5}$, PM_{10}, O_3, NO_2, SO_2, CO) in Seoul, South Korea. Peer Journal, 8, 1-18. https://doi.org/10.7717/peerj.9961.
20. Shomar, S. S. H. (2019). Investigating the Effective Parameters in Determining the Type of Traffic Congestion Pricing Schemes in Urban Streets. International Journal of Transport and Vehicle Engineering, 13(7), 379-385.
21. Stefenon, S. F., Seman, L. O., Mariani, V. C., and Coelho, L. D. S. (2023). Aggregating Prophet and Seasonal Trend Decomposition for Time Series Forecasting of Italian Electricity Spot Prices. Energies, 16(3), 1371.
22. Tabachnick, B. G., and Fidell, L. S. (2007). Using Multivariate Statistics, Fifth Edition, Pearson Education, Incorporated.
23. Taylor, S. J., and Letham, B. (2018). Forecasting at Scale. The American Statistician, 72(1), 37-45. https://doi.org/10.1080/00031305.2017.1380080.
24. Wei, N., Peng, C., Li, X., Zeng, F., and Lu, X. (2019). Conventional Models and Artificial Intelligence-Based Models for Energy Consumption Forecasting: A Review. Journal of Petroleum Science and Engineering, 181, 1-22. https://doi.org/10.1016/j.petrol.2019.106187.
25. Weytjens, H., Lohmann, E., and Kleinsteuber, M. (2019). Cash Flow Prediction: MLP and LSTM Compared to ARIMA and Prophet. Electronic Commerce Research, 1-21. https://doi.org/10.1007/s10660-019-09362-7.
26. Yenidoğan, I., Çayır, A., Kozan, O., Dağ, T., and Arslan, ç. (2019). Bitcoin Forecasting Using ARIMA and Prophet. Conference Paper, 1-6. https://doi.org/ 10.1109/UBMK.2018.8566476.
27. Yuan, C., and Chen, D. (2016). Effectiveness of GM (1, 1) Model on Linear Growth Sequence and its Application in Global Primary Energy Consumption Prediction. Emerald, 45(9), 1472-14.
28. Zhang, X., Zhang, T., Young, A. A., and Li, X. (2014). Applications and Comparisons of Four-Time Series Models in Epidemiological Surveillance Data. Plos One, 9(2), e88075.