Universiti Teknologi MARA | Cawangan Kedah Kampus Sungai Petani

2021 iCMS International Conference on Computing, Mathematics and Statistics

e-PROCEEDINGS

of The 5th International Conference on Computing, Mathematics and Statistics (iCMS2021)

4-5 August 2021

Driving Research Towards Excellence

e-Proceedings
of the 5th International Conference on Computing, Mathematics and Statistics (iCMS 2021)

Driving Research Towards Excellence

Editor-in-Chief: Norin Rahayu Shamsuddin

Editorial team:
Dr. Afida Ahamad
Dr. Norliana Mohd Najib
Dr. Nor Athirah Mohd Zin
Dr. Siti Nur Alwani Salleh
Kartini Kasim
Dr. Ida Normaya Mohd Nasir
Kamarul Ariffin Mansor

e-ISBN: 978-967-2948-12-4
DOI

Library of Congress Control Number:

Publication by
Department of Mathematical Sciences
Faculty of Computer & Mathematical Sciences
UiTM Kedah

TABLE OF CONTENT

PART 1: MATHEMATICS

PART 2: STATISTICS

PART 3: COMPUTER SCIENCE & INFORMATION TECHNOLOGY

PART 4: OTHERS

STATISTICAL LEARNING OF AIR PASSENGER TRAFFIC AT THE MURTALA MUHAMMED INTERNATIONAL AIRPORT, NIGERIA

Christopher Godwin Udomboso[1,*] and Gabriel Olugbenga Ojo[2]
[1]Department of Statistics, University of Ibadan, Ibadan, Nigeria
[2]Department of Mathematics and Statistics, Federal Polytechnic, Ilaro, Ogun State, Nigeria
([1,*]udomboso@gmail.com, [2]adeoluwie@gmail.com)

Based on previous studies, aviation affair needs reliable forecasts of air passenger traffic flow. In this research, the performance of Artificial Neural Network (ANN) and Support Vector Machine (SVM) models were investigated on predicting air passenger traffic in the Murtala International Airport Nigeria. Past eleven years' monthly data (2007-2018) obtained from Statistics Department of the Nigerian Airspace Management Agency (NAMA), MMIA, Lagos was used. ANN models with backpropagation steepest descent estimation techniques were compared with the SVM models with different kernels. The comparative evaluation of these adopted models focused basically on a Root Mean Square Error (RMSE) statistical loss function. The efficiency of the ANN model was found better than that of the SVM model in predicting the domestic air passenger traffic flow, while the SVM model predicted the foreign air passenger traffic flow more efficiently than the ANN model.
***Keywords*:** Artificial Neural Network, Support Vector Machine, ReLU Activation Function

1. Introduction

Most recently in Nigeria, Aviation has been acknowledged as being a supremely important sector of the economy. However, it is one of the indices for quantifying the development of a country. The significance of this sector to the economy of Nigeria cannot be overemphasized. The effect of this which is extensively characterized in classified businesses or alternation within cycle of growth in the nation economy and down turns, brings about data series exhibiting cyclical patterns. Thence, every economy is extremely exposed to a variety of shocks of different nature like economic, climatic etc, which are liable to alter the past trends and the volatility in the data (Bougas, 2013).

Since the inception of aviation sector in Nigeria, economy growth and development has been so rapid in which the sector has recorded massive increase in air passenger traffic. It employs and supports 245,500 jobs directly or indirectly, it contributes NGN185billion to the Nigeria economy with 4.7million and 10.7million international and domestic air passengers respectively in 2013 (National Bureau of Statistics, 2014).

In order words, Okafor et al. (2019) affirmed how significant the airline passenger traffic predictions connect as a drive to the activities within the aviation industry. Despite this methodologies, fundamental issues like accuracy, minimum classification error rate are yet to be addressed. In lieu of this, unique statistical learning techniques was employed to research on the examined model in order to know better model for easy simplicity where the aforementioned would possibly be addressed. Among several other techniques, Tsui et al. (2011) affirmed the predictions of air passenger traffic in Hong Kong International Airport by comparing both the Seasonal Autoregressive Integrated Moving Average (SARIMA) and Auto-Regressive Integrated Moving Average with Exogenous Input (ARIMAX) models. Hence, deduced that the forecasted errors of SARIMA model are smaller which made it reliable as a technique in forecasting airline passenger traffic flow.

Chen et al. (2012) employed a Back-propagation Neural Network to identify some common factors that influence air passenger and air cargo demand from Japan to Taiwan. Furthermore, construction of models which possess very high forecasting accuracy in the short and medium term was then ascertained. Thence, showed that the performance of neural network models heavily depends on choosing appropriate training set.

Dingari et al. (2019) employed a Multi-Layer Perceptron Neural Network approach in forecasting the domestic air passenger traffic in India. By virtue of validity, they affirmed that the model is relevant and best fit on the data. In 2012, Yukun et al. (2012) proposed Support Vector Machines with Ensemble Empirical Model Decomposition (EEMD-SVM) and Slope-Based as a machine technique to model on air passenger traffic data in America and British, comparing various loss functions as the performance evaluation criteria. Hence, affirmed that EEMD-Slope-SVM showed more accuracy than EEMD-SVM whereby outperform ARIMA as a univariate time series model, SVM).

Fildes et al. (2011) adopted an econometric modeling technique "Autoregressive Distributed Lag (ADL), a pooled Autoregressive Distributed Lag (a pooled ADL), Time –Varying Parameter model (TVP), Vector Auto-regressive (VAR)" on testing the error measures using the air passenger traffic data of the United Kingdom and five selected countries. However, concluded that a pooled ADL outperform the alternatives. Shen et al. (2011) established despite all the model convolutions that no method was found to constantly overwhelm other in terms of performance evaluation. Therefore, the aim is to examine the efficiency of Support Vector Machine and Artificial Neural Network techniques on air passenger traffic benchmark data.

2. Materials and Methods

2.1 Data

To benchmark the adopted methodology, monthly air passenger traffic data for both foreign and domestic flights over the period of January 2007 to December 2018 were used, which was outrightly sourced from the Nigerian Airspace Management Agency (NAMA).

2.2 Artificial Neural Networks Approach

Artificial neural networks (ANN), often traced back to McCulloch and Pitts (1943), are inspired by the learning processes of the human cognitive system and the neurological functions of the brain with a capability for self-learning and automatic abstracting and with a possible benefit of reducing modeling times. One application of ANNs is an alternative modeling strategy to traditional methods of data and time series analysis.

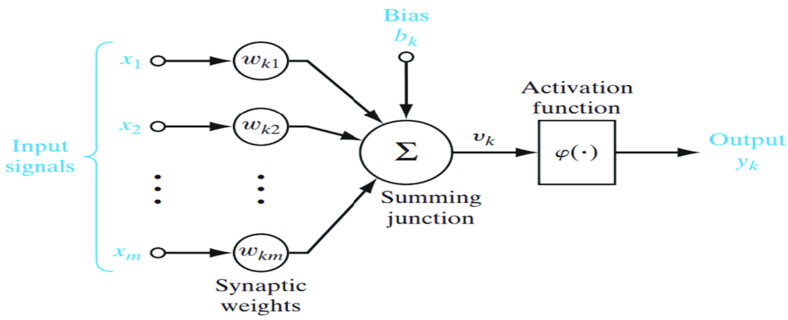

Figure 1: Nonlinear model of a neuron (Haykin, 1998).

2.2.1 Model Specification

At the hidden layers, each neuron computes (w_{ik}), a weighted sum of m input signals, x_i, for $i = 0,1,2, \dots, m$ and then applies a nonlinear activation to produce an output signal, v_k.

$$v_k = \sum_{i=0}^{m} w_{ik} . x_i + b_k = wx + b \quad (1)$$

$$y_k = \varphi(v_k) \quad (2)$$

In this study, the adoption of the rectified linear unit activation function is used to transform the output limited into an acceptable range, given as:

$$\varphi_x = \max(0, x) \tag{3}$$

Thus, the output layer y_k in Figure 1 will be obtained by the function in equation

$$y_k^m = \phi^m(w^m \phi^{m-1}(w^{m-1}\phi^{m-2}(\dots\dots w^2\phi^1(w^1 x + b^1) + b^2 + b^{m-1}) + b^m \tag{4}$$

where; y_k^m denotes the output of the overall networks, ϕ^m represents the ReLU transfer function of the m^{th}layer of the network, b^m is the bias of the neuron in the output layer. All these are in variation by adjusting the weights with the help of the feed-forward multilayer perceptron (MLP) with back propagation (BP) learning algorithm as a unique method.

2.2.2 Model Estimations

2.2.2.1 Back-Propagation Steepest Descent Techniques

The back-propagation algorithm uses the steepest-descent learning method for the weight adjustment and threshold coefficient. In other words, back propagation algorithm is used as the training method of the designed artificial neural network. Recall from (1): $v_k = \sum_{i=0}^{m} w_{ik} . x_i + b_k = wx + b$ where, $x_i{}'$s are the inputs, w_{ki}'s are the respective weights applied.

Considering a multiplayer feed forward with three-layer, where its operation is described as:

$$y_k^m = \phi^{m+1}(w^{m+1}\phi^m + b^{m+1}) \tag{5}$$

where, y^m and y^{m+1} are the outputs of the $m^{th}$and $(m+1)^{th}$ layers of the networks, respectively. b^{m+1} is the bias vector of $(m+1)^{th}$ layers of the networks, m=0,1, ---, M-1 where M is the number of layers of the neural network, the neuron of the first layer obtain input:

$$y^0 = x \tag{6}$$

Equation (6) gives the initial condition in (5) whereby the overall network which is the last layer in an output neuron can be specified and defined as:

$$y = y^m \tag{7}$$

Basically, training the network with associations between a defined set of input-output pairs $\{(x_1, t_1), (x_2, t_2), \dots\dots, (x_q, t_q)\}$, where x_q is an input to the network, and t_q is the corresponding target output is the focus intention. Therefore, by Mathematical model expression, the backpropagation algorithm uses mean square error as the performance index, which is minimized by adjusting the network parameters, as shown in below:

$$F(x) = \mathrm{E}[e^T e] = \mathrm{E}[(t - y)'(t - y)] \tag{8}$$

where x denotes the vector matrix of network weights and biases.

Applying the approximate steepest descent rule the performance index $F(x)$ becomes:

$$\hat{F}(x) = \big(t(k) - y(k)\big)^T \big(t(k) - y(k)\big) = e^T(k)e(k) \tag{9}$$

where the expectation of the squared error in (9) has been replaced by the squared error at iteration step k. The Steepest Descent Algorithm for the approximation mean square error is:

$$w_{i,j}^{m}(k+1) = w_{i,j}^{m}(k) - \alpha \frac{\partial \hat{F}}{\partial w_{i,j}^{m}} \tag{10}$$

$$b_{i}^{m}(k+1) = \ b_{i}^{m}(k) - \ \alpha \frac{\partial \hat{F}}{\partial b_{i}^{m}} \tag{11}$$

where α denotes the learning rate.

Using Chain Rule, the partial derivatives with respect to w and b in (10) and (11) can be computed as:

$$\frac{\partial \hat{F}}{\partial w_{i,j}^{m}} = \ \frac{\partial \hat{F}}{\partial v_{i}^{m}} \times \frac{\partial v_{i}^{m}}{\partial w_{i,j}^{m}} \tag{12}$$

$$\frac{\partial \hat{F}}{\partial b_{i}^{m}} = \ \frac{\partial \hat{F}}{\partial v_{i}^{m}} \times \frac{\partial v_{i}^{m}}{\partial b_{i}^{m}} \tag{13}$$

Since, the net input to layer m is an explicit function of the weights and bias then:

$$v_{i}^{m} = \sum_{j=1}^{S^{m-1}} w_{i,j}^{m}\, y_{j}^{m-1} + \ b_{i}^{m} \tag{14}$$

Therefore,

$$\frac{\partial v_{i}^{m}}{\partial w_{i,j}^{m}} = y_{j}^{m-1} \tag{15}$$

$$\frac{\partial v_{i}^{m}}{\partial b_{i}^{m}} = 1 \tag{16}$$

We now define S_{i}^{m} as the sensitivity of $\hat{F}$ to changes in the i^{th} element of the next input at layer m.

$$S_{i}^{m} \equiv \frac{\partial \hat{F}}{\partial v_{i}^{m}} \tag{17}$$

Using the defined sensitivity, the derivatives in (12) and (13) can be expressed as:

$$\frac{\partial \hat{F}}{\partial w_{i,j}^{m}} = S_{i}^{m} y_{j}^{m-1} \tag{18}$$

$$\frac{\partial \hat{F}}{\partial b_{i}^{m}} = S_{i}^{m} \tag{19}$$

Now, the simplification of the approximate steepest descent algorithm becomes:

$$w_{i,j}^{m}(k+1) = w_{i,j}^{m}(k) - \alpha S_{i}^{m} y_{j}^{m-1} \tag{20}$$

$$b_{i}^{m}(k+1) = \ b_{i}^{m}(k) - \ \alpha S_{i}^{m} \tag{21}$$

Rewritten (20) and (21) in matrix form then becomes:

$$w^{m}(k+1) = w_{i,j}^{m}(k) - \alpha S_{i}^{m} y_{j}^{m-1} \tag{22}$$

$$b^{m}(k+1) = b_{i}^{m}(k) - \ \alpha S_{i}^{m} \tag{23}$$

where;

$$S_i^m \equiv \frac{\partial \hat{F}}{\partial v^m} = \begin{bmatrix} \frac{\partial \hat{F}}{\partial v_1^m} \\ \frac{\partial \hat{F}}{\partial v_2^m} \\ \vdots \\ \frac{\partial \hat{F}}{\partial v_{S^m}^m} \end{bmatrix} \tag{24}$$

To derive the recurrence relationship between the sensitivity at layer m computed from $m+1$; the Jacobian matrix is used:

$$\frac{\partial v^{m+1}}{\partial v^m} \equiv \begin{bmatrix} \frac{\partial v_1^{m+1}}{\partial v_1^m} & \frac{\partial v_1^{m+1}}{\partial v_2^m} \cdots & \frac{\partial v_1^{m+1}}{\partial v_{S^m}^m} \\ \frac{\partial v_2^{m+1}}{\partial v_1^m} & \frac{\partial v_2^{m+1}}{\partial v_2^m} \cdots & \frac{\partial v_1^{m+1}}{\partial v_{S^m}^m} \\ \vdots & \vdots & \vdots \\ \frac{\partial v_{S^m}^{m+1}}{\partial v_1^m} & \frac{\partial v_{S^{m+1}}^{m+1}}{\partial v_2^m} \cdots & \frac{\partial v_{S^{m+1}}^{m+1}}{\partial v_{S^m}^m} \end{bmatrix} \tag{25}$$

Considering the i, j element in the matrix:

$$\frac{\partial v_i^{m+1}}{\partial v_j^m} = \frac{\partial\left(\sum_{i=1}^{S^m} w_{i,l}^{m+1} y_i^m + b_i^{m+1}\right)}{\partial v_j^m} = w_{i,j}^{m+1} \frac{\partial y_j^m}{\partial v_j^m} \tag{26}$$

Rewritten the Jacobian matrix in (25) thus becomes:

$$\frac{\partial v_i^{m+1}}{\partial v_j^m} = w_{i,j}^{m+1} \frac{\partial f^m\left(v_j^m\right)}{\partial v_j^m} = w_{i,j}^{m+1} f^m\left(v_j^m\right) \tag{27}$$

where

$$f^m\left(v_j^m\right) = \frac{\partial f^m\left(v_j^m\right)}{\partial v_j^m} \tag{28}$$

By substitution, (25) again becomes:

$$\frac{\partial v^{m+1}}{\partial v^m} = W^{m+1} F^m(V^m) \tag{29}$$

where

$$F^m(V^m) = \begin{bmatrix} f^m(v_1^m) & 0 & 0 \\ 0 & f^m(v_1^m) & 0 \\ 0 & 0 & f^m\left(v_{S^m}^m\right) \end{bmatrix} \tag{30}$$

Applying the "Chain Rule", the recurrence relation for the sensitivity is now written as:

$$S^m = \frac{\partial \hat{F}}{\partial v^m} = \left(\frac{\partial v^{m+1}}{\partial v^m}\right)^T \frac{\partial \hat{F}}{\partial v^{m+1}} \tag{31}$$

$$S^m = F^m(V^m)(W^{m+1})^T \frac{\partial \hat{F}}{\partial v^{m+1}} \tag{32}$$

Therefore

$$S^m = F^m(V^m)(W^{m+1})^T S^{m+1} \tag{33}$$

Hence, from the last layer to the first layer, the sensitivity is propagated backward through the network

$$S^M \rightarrow S^{M-1} \rightarrow \cdots \rightarrow S^2 \rightarrow S^1 \tag{34}$$

The recurrence relation of (33) at the starting point S^M is now shown as:

$$S_i^m = \frac{\partial \hat{F}}{\partial V_i^M} = \frac{\partial (t-y)^T(t-y)}{\partial V_i^M} = \frac{\partial \sum_{j=1}^{S^M}\left(t_j - y_j\right)^2}{\partial V_i^M} \tag{35}$$

$$S_i^m = -2(t_i - y_i)\frac{\partial y_i^M}{\partial V_i^M} \tag{36}$$

Now since,

$$\frac{\partial y_i}{\partial V_i^M} = \frac{\partial y_i^M}{\partial V_i^M} = \frac{\partial f^M(V_i^M)}{\partial V_i^M} = f^M\left(V_i^M\right) \tag{37}$$

Finally, express (37) in form of matrix becomes:

$$S^M = -2F^M(V^M)(t-y) \tag{38}$$

However, the iteration is implemented continuously for several layers of the neural network as obtainable in the multi-layer feed forward neural network.

2.3 Support Vector Machines Approach

2.3.1 Model Specification

In accordance to Oğcu et al. (2012), the support vector machine (SVM) is technically based on classification with an output variable restricted to take only binary values. Let us consider a supervised binary classification problem. If the training data are represented by $\{x_i, y_i\}, i = 1,2, \dots, n$, and $y_i \in \{-1,1\}$, where N is the number of training samples, $y_i = 1$ for class $\varphi_1$and $y_i = -1$ for class φ_2.

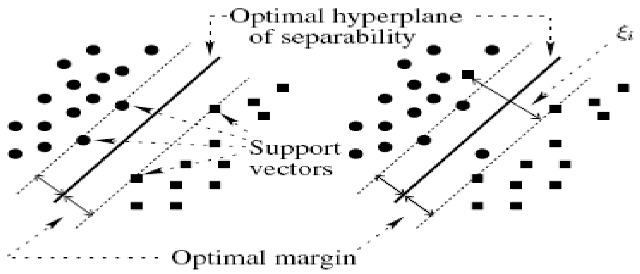

Figure 2: **Left**: linear separable case. **Right**: non-linear separable case. where ξ is the error measurement of the hyperplane (Mercier and Lennon, 2003).

2.3.2 Model Estimation

In mathematical terms, the maximal margin hyperplane for non-separable data is selected by minimizing the cost function: Considering

$$y_n[w^T \phi(x_n) + b] \leq 0 \tag{39}$$

where, w denote a weight vector, and b denote a bias value. Thus, the objective function in (40) can be used to maximize separating margin, and minimize error.

$$\min \frac{1}{2}\|w\|^2 + \eta \sum_n \xi_n \tag{40}$$

Here η is denoted by a regularized constant greater than 0 to perform balancing between the training error and model flatness.

$$\text{Subject to the constraints: } y_n[w^T \phi(x_n) + b] \geq 1 - \xi_n \text{ , } At\ \xi_n \geq 0 \text{ for all } n \tag{41}$$

where the variables ξ_n are identified as slack variables. The goal of the optimization task is to make the margin enormous and lower the number of misclassification points with $\xi > 0$. The parameter η is a positive constant that bewiled the relative influence of the two important terms.

By introducing Lagrangian multipliers, $\mu_n, \beta_n \geq 0$ ($n = \{1, \dots, N\}$) for the constraints, leads to the Primal Lagrangian:

$$\text{Minimize } L_P(w, b, \{\xi_n\}, \{\mu_n\}, \{\beta_n\}) = \left[\frac{1}{2}\|w\|^2 + \eta \sum_n \xi_n\right] + \sum_n [\beta_n(1 - \xi_n - y_n[w^T \phi(x_n) + b])] + \sum_n \mu_n(-\xi_n) \tag{42}$$

subject to $\mu_n, \beta_n \geq 0; n = 1, \dots, N$. We then look for a stationary point of L_P from (42) by taking its partial derivatives w.r.t to w, b and ξ_n and setting it gradients to zero:

$$\frac{\partial L}{\partial w} = w - \sum_n \beta_n\, y_n \phi(x_n) = 0 \tag{43}$$

$$\frac{\partial L}{\partial b} = \sum_n \beta_n\, y_n = 0 \tag{44}$$

$$\frac{\partial L}{\partial \xi_n} = \eta - \beta_n - \mu_n \tag{45}$$

Using $\eta - \beta_n - \mu_n = 0$ and $\beta_n \geq 0 \Longrightarrow \mu_n \leq \eta$, substitute (43), (44), and (45) in (42), gives the Dual Lagrangian below:

$$L_D(w, b, \{\xi_n\}, \{\mu_n\}, \{\beta_n\}) = \left[\frac{1}{2}\left(\sum_m \beta_m\, y_m \phi(x_m)\right)^T \left(\sum_n \beta_n\, y_n \phi(x_n)\right) + \eta \sum_n \xi_n\right] + \sum_n [\beta_m\{1 - \xi_n - y_n[(\sum_m \beta_m y_m \phi(x_n))^T \phi(x_n) + b]\}] + \sum_n \mu_n\,(-\xi_n) \tag{46}$$

$$L_D = \sum_n \beta_n + \frac{1}{2} \sum_m \beta_m\, \beta_n y_m y_n\ \phi^T(x_n)\phi(x_n) + \sum_n \beta_n(\eta - \xi_n - \mu_n) \tag{47}$$

$$L_D = \sum_n \beta_n - \frac{1}{2} \sum_m \beta_m\, \beta_n y_m y_n\ \phi^T(x_n)\phi(x_n) \tag{48}$$

Therefore,

$$g(\{\beta_n\}, \{\mu_n\}) = \sum_n \beta_n - \frac{1}{2} \sum_m \beta_m\, \beta_n y_m y_n\ \phi^T(x_n)\phi(x_n) \tag{49}$$

$$\underset{\{\beta_n\}}{max} g(\{\beta_n\}, \{\mu_n\}) = \sum_n \beta_n - \frac{1}{2} \sum_m \beta_m\, \beta_n y_m y_n\ \phi^T(x_n)\phi(x_n), \text{ such that } \beta_n, \mu_n \geq 0 \text{ for all } n \tag{50}$$

Diosan et al, (2012), deduced that nonlinear kernels are needed to map the input dataset into a multidimensional space where the new mapping approach brought about a linearly separable case. The real fact about the kernel function is however meant to function in the input space rather than the potentially high dimension feature space. Furthermore, the inner product need not to be evaluated in the feature space, by definition, the mapping is achieved by replacement of the inner product $(x_m, x_n) \to \phi(x_m)\phi(x_n)$ called the kernel, $\phi^T(x_m)\phi(x_n) = k(x_m, x_n)$, where $k(x_m, x_n)$ is

declared as a kernel function. Symmetrically, $k(x_m, x_n) = k(x_n, x_m)$; By positive semi-definite $\sum_m \sum_n v_m v_n k(x_m, x_n) \geq 0$.
Then,

$$\phi^T(x_n)\phi(x_n) = k(x_m, x_n) \tag{51}$$

Substitute (51) into (50) becomes:

$$\max_{\{\beta_i\}} g(\{\beta_n\}, \{\mu_n\}) = \sum_n \beta_n - \frac{1}{2}\sum_m \beta_m \beta_n y_m y_n\, k(x_m, x_n), \text{such that } \beta_n, \mu_n \geq 0 \text{ for all } n \tag{52}$$

The kernel functions adopted in this study are:

Linear $k(x_n, x) = {x_n}^T x + \mathrm{c}$, Polynomial $k(x_n, x) = (\alpha {x_n}^T x + c)^d$, Radial Basis $k(x_n, x) = exp\left(-\frac{\|x_n - x\|^2}{2\sigma^2}\right)$, and Sigmoid kernel functions $k(x_n, x) = \tanh\,(\alpha {x_n}^T x + c)$ whereby input in the basic mathematical function given as:

$$Y = f(x) = [\sum_{n=1}^{N} \beta_n\, k\,(x_n . x)] - b \tag{53}$$

In (52), k is the kernel function, N is the number of training data points, x_n represents vectors used in the training process, x *is* an independent vector, β_n and b are the parameters derived by the objective function maximization. However, each kernel function listed above has a particular parameter that must be optimized to obtain the best result performance.

2.4 Model Performance Evaluation and Forecasting

The performance of each model for both the training and testing data were evaluated according to the Root Mean Square Error (RMSE) which are widely used in time series forecasting like; Chen et al. (2012) and Coshall (2006). In this study, the RMSE was adopted to check the closeness of the predicted values to the actual values of the monthly air passenger traffic data.

$$\text{RMSE} = \sqrt{\sum_{j=1}^{n} \frac{(T_j - Y_j)^2}{n}} \tag{54}$$

where, T_j denotes the actual value and Y_j as the predicted value.

3. Data Analysis and Results

The domestic and foreign airline traffic of MMIA time plots can be evidenced in the Figures 3 and 4 respectively, which however showed seasonal variation in the data and as well clarified the existence of upward movement from time to time. Hence, deduced an influx of passengers moving in and out of the country on timely basis.

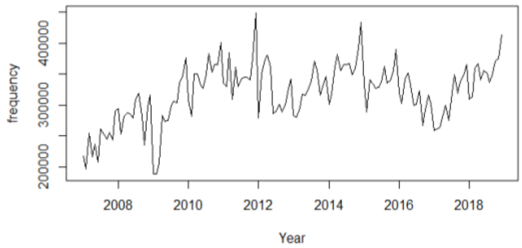

Figure 3: Time plot of domestic airline passenger traffic between 2007-2018.

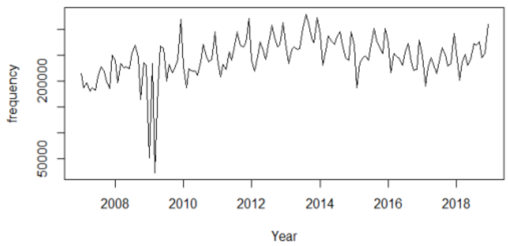

Figure 4: Time plot of foreign airline passenger traffic between 2007-2018.

Table 1: Performance Evaluation of ANN Model Using Both Domestic and Foreign Airline Traffic.

No. of Neurons	Domestic Airline		Foreign Airline	
	RMSE	Best Iterations	RMSE	Best Iterations
10	33071.8	5900	13359.09	5200
20	33003.88	5300	13357.38	3200
30	33058.59	4000	13387.26	2800
40	33020.2	3700	13348.67	2900
50	33016.37	3700	13381.88	2300
60	33012.22	3700	13382.27	2200
70	33034.08	2800	13397.82	2200
80	33034.05	2800	13420.63	1900
90	33031.52	2700	13431.18	1800

In Table 1 and Figure 5, it is well emphasized that the ANN model of the domestic airline traffic was best achieved at 5300 iterations with RMSE of 33003.88, achieved at neuron 20 compared to its counterpart models with higher RMSE respectively. This implies that neuron 20 may perform better in estimating the predictive ability of the MMIA domestic passenger traffic. While the fitted ANN model with 1 hidden layer and 40 neurons using the foreign airline traffic series showed in Table.1 and Figure 6 deduced that optimal solution was achieved at iteration 2900 with lower RMSE of 13348.67.

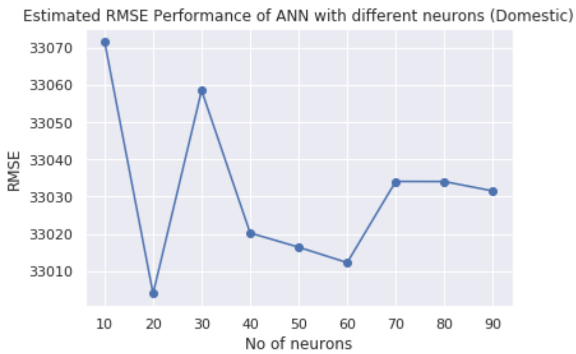

Figure 5: Graph showing RMSE performance of ANN with different number of neurons using the domestic airline traffic.

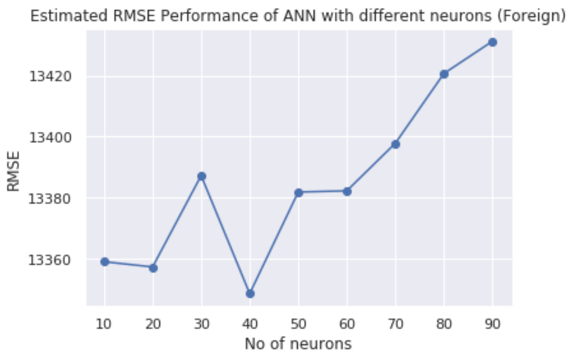

Figure 6: Graph showing RMSE performance of ANN with different number of neurons using the foreign airline traffic.

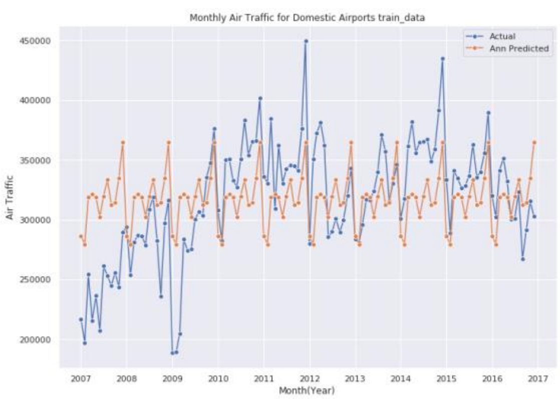

Figure 7: Plot of ANN Actual and Predicted Domestic Air Traffic (Train data).

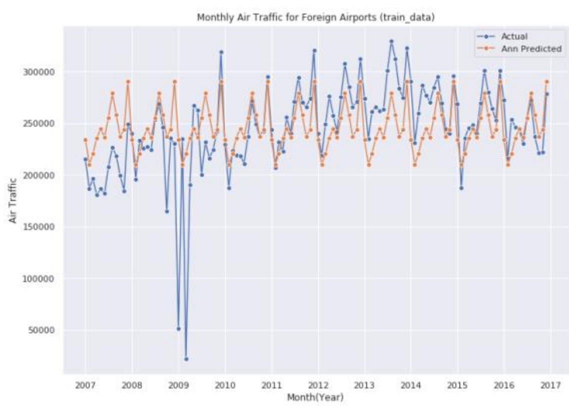

Figure 8: Plot of ANN Actual and Predicted Foreign Air Traffic (Train data).

Graphical representation of the actual and predicted plot of the trained data in Figure 7 for the ANN model indicated upward and downward changes in the trend values of the series. While Figure 8, showed the prediction accuracy of the 1 hidden layer with 40 neurons using the trained data as a proxy.

Table 2: Performance Evaluation of SVM Models Using Some Selected Kernels for Foreign and Domestic Airline Traffic.

	Domestic Airline			Foreign Airline		
Kernels	RMSE	C value	Degree Value	RMSE	C value	Degree Value
Linear	34958.413	13000	-	13345.957137	25000	-
Rbf	34959.270	15000	-	13345.930100	40000	-
Polynomial	34974.590	-	110	13345.925823	-	130
sigmoid	34958.973	16000	-	13345.956966	30000	-

It can be observed from Table 2 that the SVM model with Linear Kernel having tuning parameter C=13000 has the minimum RMSE among its counterparts of rbf, polynomial and sigmoid kernels. Thereby implied that the SVM model of the domestic airline traffic with linear kernel is more efficient in predicting local airline passenger traffic. Consequently, taking the foreign airline traffic into consideration, analysis showed that polynomial SVM kernel model with degree 130 has the minimum RMSE and has predicted out rightly well for the data among its selected counterpart models in terms of its efficiency. This can also be evidenced from Figures 9 and 10 below respectively.

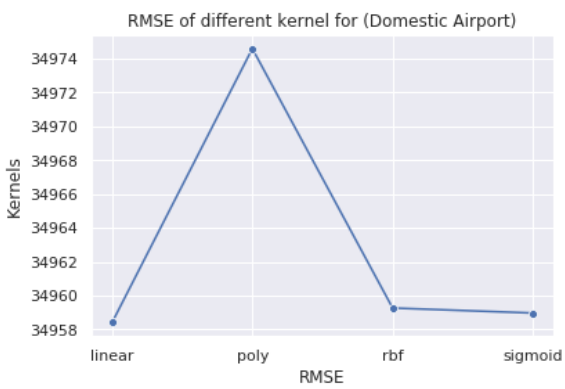

Figure 9: Graph showing the RMSE performance of different kernel in the Domestic Air Traffic Series.

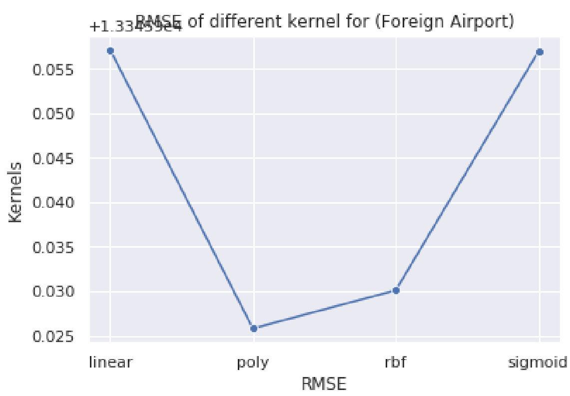

Figure 10: Graph showing the RMSE performance of different kernel in the Foreign Air Traffic Series.

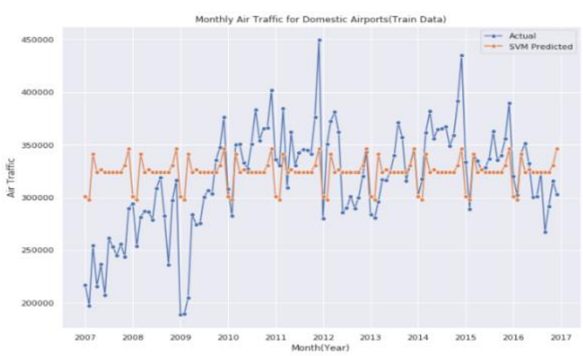

Figure 11: Plot of SVM Actual and Predicted Domestic Air Traffic (Train data).

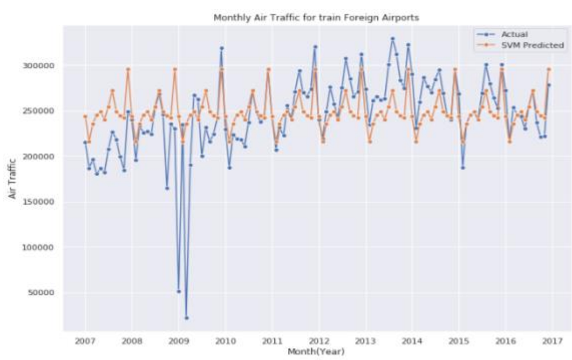

Figure 12: Plot of SVM Actual and Predicted Foreign Air Traffic (Train data).

The actual and predicted plot of both domestic and foreign air traffic trained data can be depicted in Figures 11 and 12 above. Based on the predictions, it was deduced that SVM model with linear kernel function in Figure 7 were not influenced by external forces. While SVM model with polynomial kernel function in Figure 12 showed that there exist variations between the actual and predicted values respectively. The evidenced were gathered subsequently from Table 3 below.

Table 3: Out of Sample Predictions on Selected Best Models.

Month	Domestic airline traffic out of sample predictions			Foreign airline traffic out of sample predictions		
	Actual	ANN Predicted	SVM Predicted	Actual	ANN Predicted	SVM Predicted
Jan.,2017	259120	286693.72	301164.1002	247385	233996.08	243347.9
Feb.,2017	261354	279885.97	298046.1667	190553	209700.83	215644.8993
Mar.,2017	264254	318834.28	341327.8997	225392	219846.38	235126.8995
April,2017	284987	321288.34	324046.1667	245408	235510.92	245460.9001
May,2017	299440	318665.25	326820.1002	232009	244172.88	248905.8818
June,2017	274879	302527.72	323990.9002	214790	236290.8	240232.8999
July,2017	320221	319678.88	324046.1667	243338	255067.4	254553.1001
Aug.,2017	349454	333296.47	324046.1667	264100	278702.53	271945.1002
Sept.,2017	319032	312730.53	324046.1667	251320	257625.75	248926.1
Oct.,2017	337832	313949.10	324046.1667	229491	236964.67	244734.9005
Nov.,2017	347682	334608.40	330584.0997	233907	243408.95	242365.9
Dec.,2017	365797	363591.53	346389.9002	291406	290607.03	295626.1004
Jan.,2018	308849	286693.72	301164.1002	254440	233996.08	243347.9
Feb.,2018	313391	279885.97	298046.1667	200564	209700.83	215644.8993
Mar.,2018	360356	318834.28	341327.8997	235552	219846.38	235126.8995
April,2018	366740	321288.34	324046.1667	251389	235510.92	245460.9001
May,2018	340339	318665.25	326820.1002	231146	244172.88	248905.8818
June,2018	356650	302527.72	323990.9002	241850	236290.8	240232.8999
July,2018	351414	319678.88	324046.1667	272789	255067.4	254553.1001
Aug.,2018	336657	333296.47	324046.1667	268619	278702.53	271945.1002
Sept.,2018	351409	312730.53	324046.1667	275864	257625.75	248926.1
Oct.,2018	370784	313949.10	324046.1667	245383	236964.67	244734.9005
Nov.,2018	375224	334608.40	330584.0997	253102	243408.95	242365.9
Dec.,2018	413334	363591.53	346389.9002	309685	290607.03	295626.1004

3.1 Prediction Performance

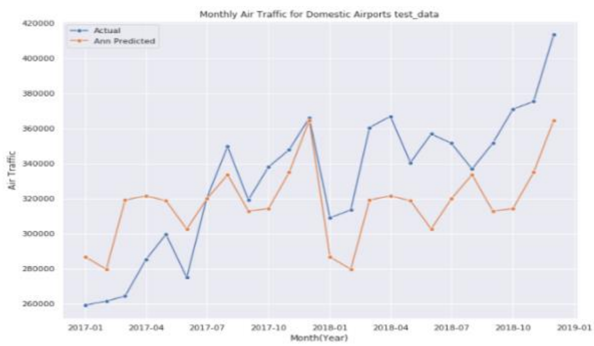

Figure 13: Out of sample Plot of Actual and Predicted Domestic Airline Passenger Traffic using ANN model.

Figure 14: Out of sample Plot of Actual and Predicted Domestic Airline Passenger Traffic using SVM model.

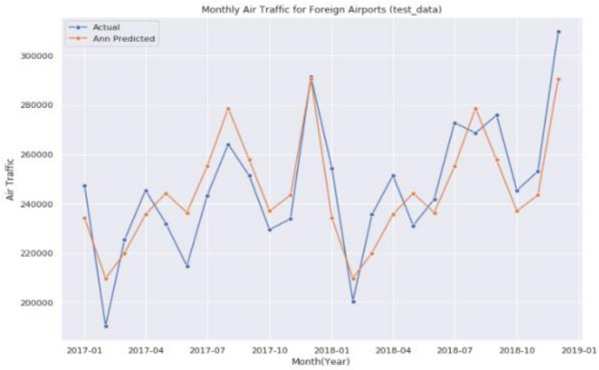

Figure 15: Out of sample Plot of Actual and Predicted Foreign Airline Passenger Traffic using ANN model.

Figure 16: Out of sample Plot of Actual and Predicted Foreign Airline Passenger Traffic using SVM model.

The out-of-sample prediction for the best identified models can be depicted in Table 3. The Artificial Neural Network and Support Vector Machine model predictions for the studied series showed little variations between the actual and predicted values of the domestic and foreign air traffic using two years of sample data spanning from year 2017 and 2018 respectively. Pictorial representation can be seen in Figure 13, 14, 15 and 16 above respectively.

3.2 Inference on Best Models Comparison

In order to make a comparison between the forecasted values of ANN and SVM methods, error values based on RMSE were used. Results based on test data which include the two years' period of 2017-2018 for both domestic and foreign; showed that ANN and SVM are capable of tracking the actual trends.

Table 4: Optimal metric for Domestic and Foreign Air Traffic of ANN and SVM.

Model	Domestic Air Traffic		Foreign Air Traffic	
	Test RMSE	Train RMSE	Test RMSE	Train RMSE
ANN	33003.878906	43868.25	13348.665039	37559.26
SVM	34958.412592	46028.98	13345.925823	38473.20

Comparison test between the actual and predicted values of the ANN and SVM methods adopted for the domestic airline passengers' traffic in Figure 17 deduced that ANN model proved to be the best based on the minimum Test data RMSE in Table 4. However, the ANN and SVM methods adopted for the foreign airline passengers' traffic in Figure 18 since showed that SVM model proved to be the best based on the minimum Test data RMSE in Table 4 as well. In order words, ANN outperforms SVM in the domestic air sector, while SVM outperforms ANN in the foreign air sector. Thence, the flexibility of the two models involved really depend on the airline sector to be predicted.

4. Conclusion

This paper was based on forecasting the air passenger traffic flow using the Root Mean Square Error (RMSE) as a performance evaluation technique on the examined models. Hence, empirical results indicate that SVM model has a better accuracy to outperforms the ANN model in the foreign air sector, while ANN model consequently, outperforms the SVM model in the domestic air sector. Thence, the exhibition of these two models were quite satisfactory as against Shen et al. (2011), where the quality testing of regression on the training set showed an extreme good prediction accuracy (Zhaoyue et al., 2021) which would certainly help the MMIA to improve and revise the affairs of the industry. For further studies, I indeed recommend several other types of Machine Learning, and Artificial Intelligence Based Forecasting Methods. However, Factors affecting the nature of the time series data in both airlines such as exogenous variables could also be captured using appropriate time series models. Also, better improvement in evaluation could be achieved if daily data are as well used for the purpose.

References

Bougas, C. (2013). *Forecasting air passenger traffic flows in Canada: an evaluation of time series models and combination methods: Master's thesis*. Université Laval, Canada.

Chen, S. C., Kuo, S. Y., Chang, K. W., and Wang, Y. T. (2012). Improving the forecasting accuracy of air passenger and air cargo demand: the application of back-propagation neural networks. *Transportation Planning and Technology*, 35(3):373-392.

Coshall, J. (2006). Time series analyses of UK outbound travel by air. *Journal of Travel Research*, 44(3):335-347.

Dingari, M., Doodipala, M. R., and Sumalatha, V. (2019). Air traffic forecasting using Artificial Neural Network. *International Journal of Scientific & Technology Research.* 8(10):556-559.

Dioşan, L., Rogozan, A., and Pecuchet, J. P. (2012). Improving classification performance of Support Vector Machine by genetically optimizing kernel shape and hyper-parameters. *Applied Intelligence*, 36:280-294.

Fildes, R., Wei, Y., and Ismail, S. (2011). Evaluating the forecasting performance of econometric models of air passenger traffic flows using multiple error measures. *International Journal of Forecasting*, 27(3):902-922.

Haykin, S. (1998). *Neural networks: A comprehensive foundation*. Prentice Hall, United States.

McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115-133.

Mercier, G. and Lennon, M. (2003). Support vector machines for hyperspectral image classification with spectral-based kernels. *International Geoscience and Remote Sensing Symposium (IGARSS)*, Vol. 1, pp 288-290. 10.1109/IGARSS.2003.1293752.

National Bureau of Statistics. (2014). www.nigerianstat.gov.ng.

Oğcu, G., Demirel, O. F., and Zaim, S. (2012). Forecasting electricity consumption with neural networks and support vector regression. *Procedia-Social and Behavioral Science*, 58:1576-1585.

Okafor, E. G., Buraimoh, O. C., Uheugho, K. O., and Soladoye, M. A. (2019). Propensity of fly in Nigeria: A forecast of domestic air passenger traffic flow in some Nigeria airports. *Nigeria Journal of Technology*, 38(2):456-464.

Shen, S., Li, G., and Song, H. (2011). Combination forecasts of international tourism demand. *Annals of Tourism Research*, 38(1):72-89.

Tsui, W., Balli, H. O., and Gower, H. (2011). Forecasting airport passenger traffic: the case of Hong Kong International Airport. *Aviation Education and Research Proceedings*, 2011:54-62.

Yukun, B., Xiong, T. and Hu, Z. (2012). Forecasting air passenger traffic by support vector machines with ensemble empirical mode decomposition and slope-based method. *Discrete Dynamics in Nature and Society*, 2012:1-12.

Zhaoyue, Z., Zhang, A., Sun, C., Xiang, S., Guan, J., and Huang, X. (2021). Research on air traffic flow forecast based on ELM non-iterative algorithm. *Mobile Networks and Applications*, 26:425-439.

2021 ICMS

INTERNATIONAL CONFERENCE ON COMPUTING, MATHEMATICS AND STATISTICS

e ISBN 978-967-2948-12-4

9 7 8 9 6 7 2 9 4 8 1 2 4