INTERNATIONAL JOUNAL OF: SCIENCE RESEARCH AND TECHNOLOGY

TIMBOU-AFRICA ACADEMIC PUBLICATIONS
INTERNATIONAL JOURNAL, AUGUST, 2022 EDITIONS
VOL. 10 NO. 9 ISSN: 2623-7861

APPLICATION OF SUPPORT VECTOR MACHINES AND HOLT WINTERS EXPONENTIAL SMOOTHING MODELLING APPROACHES IN AIRLINES PASSENGERS' TIME SERIES FORECASTING

[1]ALHAJI ISMAILA SULAIMAN; [2]MOHAMMED SALISU ALFA; & [1]ABUBAKAR MUHAMMAD AUWAL,

[1]Department of Statistics, Nasarawa State University, Keffi. Nasarawa State. [2]Department of Statistics, The Federal Polytechnic, Bida. Niger State

ABSTRACT

This paper examines the application of Artificial Intelligence (AI) and statistical methods which are support vector machine (SVM) and Holt Winters Exponential Smoothing (HW) models respectively in time series forecasting. The aim of this paper is to examine the feasibility of applying the two models to airlines passenger's data so as to determine the more efficient among the two

INTRODUCTION

Both Support Vector Machine (SVM) and Holt Winters Exponential Smoothing (HW) models are very useful in time series forecasting. Exponential smoothing models have been found to be amongst the most effective forecasting models. It has been applied in many fields of human endeavors. However, it suffers from the limitation of being able to capture only linear features in time series data. Support vector machines (SVM) on the other hand though new has made remarkable in roads in the field of time series forecasting. While SVM model deals with the nonlinear aspect of the data, HW model deals with the linear parts. Exponential smoothing technique is one of the most important quantitative techniques for forecasting the accuracy of time series which depends on experimental constant whose choice of the value is very crucial in minimizing the error in

models. These two models are very important in the field of time series forecasting, SVM deals with linear data while HW deals with nonlinear patterns. The results indicate that using the criteria evaluation performance HW model has Mean Square Error (MSE) of 996.10, Mean Average Error (MAE) of 44.64 and coefficient of correlation (R) of 0.75 while SVM model has MSE of 925.922, MAE of 25.24 and R of 0.89. These results show that using the time series for the airlines passenger, SVM model is the most efficient having the least error of performance and the highest coefficient of correlation. SVM model has therefore proved to be superior to the HW model and therefore recommended to be used in forecasting the airlines passengers' data because of its advantage over HW model to forecast the airlines passengers' data.

***Keywords**: SVM, HW, Time series, Forecasting, Airlines Passengers'*

forecasting (Supriatna et al., 2017). The model uses three equations and two smoothing constants. HW's method can be used for time series since it contains both trend and seasonal variations and has two versions, additive and multiplicative models which depend on the characteristics of the time series (Gelper et al., 2008). Kalekar & Bernard, (2004) stated that HW techniques are sensitive since the smoothed values are affected by the up-date equations which involve current and past values of the series and the second is the selection of the parameters used in the updating scheme.

SVM offers remarkable generalization performance in many areas such as pattern recognition, text classification and regression estimation and also that some researchers deal with the application of SVM in time series forecasting. According to Wang et al., (2013) in recent years, SVM has become a popular tool for pattern recognition and machine learning. SVM is used for classification problems and its goal is to optimize "generalization" (Wang, 2005). Support vector machines (SVMs), which were introduced by Vapnik and his coworkers in the early 1990's Cortes & Vapnik, (1995) and Vapnik, (1999), are proved to be effective and promising techniques for data mining (Chen & Wang, 2007). SVM performs better than conventional evolution methods with advantages of high efficiency, lower cost and easy implementation and similarly the new SVM-based approach holds further advantages with characteristics of lower cost, higher reliability and on-line implementation (Pai et al., 2010). SVM is a machine learning algorithm that started to be in use in the middle of 19th century and also statistical theory employs the criterion that minimizes the structure risk (Tian et al., 2012).

Among the applications of time series is forecasting and the time series forecasting technique is made up of exponential smoothing among others, (Ismail et al., 2009). In the view of Kalekar & Bernard, (2004) exponential smoothing is a process for frequently revising a forecast in the light of more recent experience. Exponential Smoothing ascribes exponentially decreasing weights as the observations get older. In other words, recent observations are given relatively more weight in forecasting than the older observations.

It is a known fact that time series forecasting has been and still being studied extensively over the years by different Scholars. Different kinds of forecasting models have been developed and statistical techniques have been relied upon by the researchers to predict time series data. Such techniques are referred to as time series forecasting. The main objective of time series forecasting is therefore to forecast future events which is based on known past events. Zhang, (2003) proposed that the accuracy of time series forecasting is fundamental to various decision process and consequently, the research that improves the effectiveness of forecasting models have not been stopped. In another opinion of Zhang, (2003), time series forecasting as an important area in which the past observed values of the variable are gathered and analyzed in order to develop a model to describe the relationship and further more use the model to extrapolate the time series into the future. Cordeiro & Neves, (2009), viewed forecasting future values of a time series as one of the motives in time series analysis in which a lot of forecasting methods have been developed and evaluated its performance. Forecasting methods can be qualitative where no formal mathematical model is required and quantitative in which historical data on variables of interest are available as it is the case in this study.

Methodology

These are the methods used in this paper to achieve the desired objective. These include SVM model and HW models used in time series forecasting.

Support Vector Machine (SVM) model

SVM model has shown its potentiality in the time series forecasting applications with their nonlinear modeling capability. The selection of an appropriate subset of variables from a set of measured potential input variables for inclusion as inputs to model the system under investigation is a vital step in model development. This is particularly important in data driven techniques, such as SVM model, as the performance of the final model is heavily dependent on the input variables used to develop the model. Selection of the best set of input variables is essential to be able to model the system under consideration reliably. When the available data set is high dimensional, it is necessary to select a subset of the potential input variables to reduce the number of free parameters in the model in order to obtain good generalization with finite data.

SVM are used on concept of decision planes which defines decision boundaries. This decision plane separates a set of objects that belong to different class members. This is illustrated by having two objects belonging to either class Blue or Yellow.

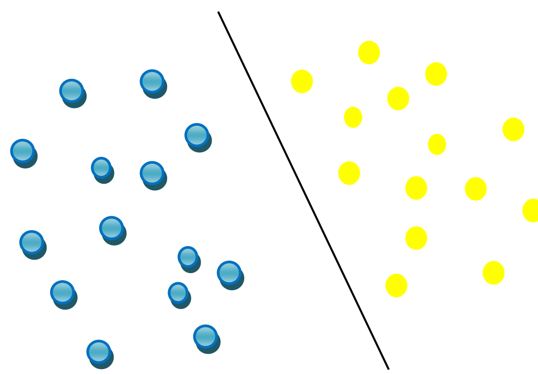

Figure 1: Input space original data

The object (Figure1) above is a typical example of a linear classifier (that is a classifier that separates the set of objects into their respective groups, blue and yellow). The line in between them defines a boundary on one side and the other on the other side of the line. However, some of these classifications are not simple and sometimes more complex structures are required in order to make an optimal separation, which is by classifying new objects (test cases) on the basis of the available train cases. This is often achieved by using curves for separation as shown below:

TIJSRAT
ISSN: 2623-7861

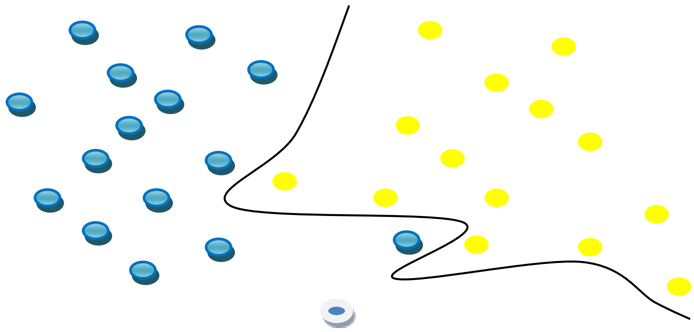

Figure 2: Mapped data in the feature space

A situation where a curve is used instead of lines as shown in figure 2 to separate the group is regarded as more complex than the line. These classifications in which curves are used for separation to differentiate between the objectives of different class members are referred to as hyperplane classifiers. SVMs are suitable to handle this.

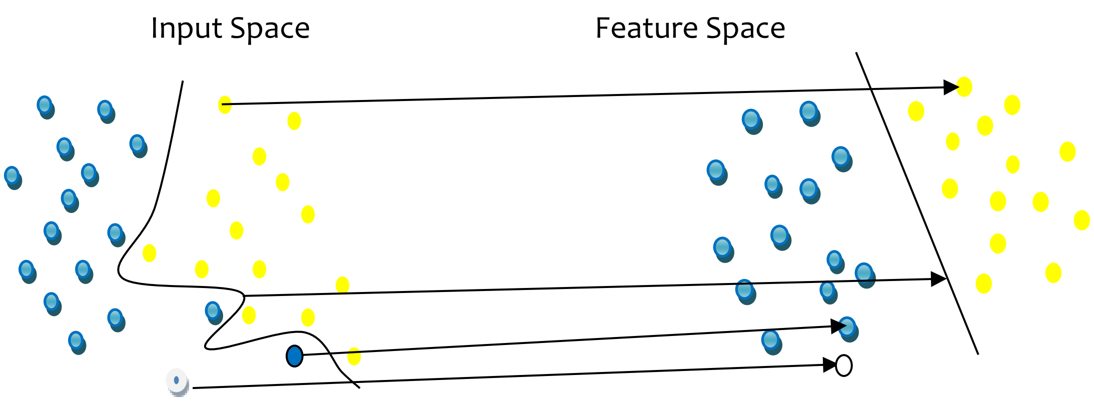

Figure 3: Support vector machines

The simplest formulation of SVM is the linear one, where the hyperplane lies on the space of the input data x. In this case the hypothesis space is a subset of all hyperplanes of the form:

$$f(x) = w.x + b \tag{1}$$

So to summarize, the hypothesis space used by SVM is a subset of the set of hyperplanes defined in some space formally written as $\{ f : \|f\|_k^2 < \infty \}$

Where k is the kernel $K(x_1, x_2) = x_1.x_2$, the functions considered are of the form $f(x) = w.x + b$ which is simply the norm of these functions is simply the norm of w

In fact SVM consider subsets of this space, namely sets of the form $\left\{ f : \| f \|_k^2 \leq A^2 \right\}$

Now solving the primal problem:

$$\text{Min } \frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i$$

subject to $y_i \left(w^T x_i + \beta \right) \geq 1 - \xi_i \;, \xi_i \geq 0$ For $i = 1, 2, \ldots, N$,

Where $x_i = (1, 2, \ldots N)$ are the N training points, y_i is the label of each point with values +1or -1 and C is the penalty cost for the sample points which are not classified correctly by the SVM, however, a large C corresponds to a higher penalty to errors.

Generally, SVM models can be classified into four different groups as follows:

a. Classification SVM

i. Classification SVM Type1:- this is known as C-SVM classification

In this type of SVM, training involves minimization of the error function usually given by:

$$\frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i$$

Subject to the constraints: $y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i \text{ and } \xi_i \geq 0, \; i = 1 \ldots N$

Where C is the capacity constant; *w* is the vector of coefficients, *b* a constant, ξ_i is a parameter for handling data (inputs). *i* is the index which label the *N* training cases. $y \in \pm 1$ is the class labels x_i is the independent variables.

The kernel ϕ is used to transform data from the input (independent) to the feature space. It should be noted that the larger the C, the more the error is penalized. Thus, C should be chosen with care to avoid over fitting.

ii. Classification SVM Type2:- this is known as nu-SVM classification

Similar to classification SVM type1, the classification SVM type2 model minimizes the error function: $\frac{1}{2} w^T w - v\rho + \frac{1}{N} \sum_{i=1}^{N} \xi_i$

Subject to the constraints $y_i \left(w^T \phi(x_i) + b \right) \geq \rho - \xi_i, \; \xi_i \geq 0, \quad i = 1 \ldots N$ and $\rho \geq 0$

b. Regression SVM

In regression, SVM which is the functional dependence of the independent variable x has to be estimated. Its assumption is like other regression problems, which states that the relationship between the independent and dependent

TIJSRAT
ISSN: 2623-7861

variables are given by a deterministic function f in addition to some additive noise:

$$y = f(x) + noise \tag{2}$$

The issue is now to find functional form for f that can correctly predict new cases for SVM. This is achieved by training the SVM model on a sample set. This process is like classification and the sequential optimization of the error function. Two types of SVM models can be applicable depending on the definition of this error function. These are regression SVM type1 and regression SVM type 2. This continues as follows:

iii. Regression SVM Type1:- this is known as epsilon-SVM regression

This type of SVM uses the error function: $\frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i + C\sum_{i=1}^{N}\xi_i^*$

This is minimized subject to:

$$\left(w^T\phi(x_i)+b\right) - y_i \le \in + \xi_i$$
$$y_i - \left(w^T\phi(x_i)+b_i\right) \le \in + \xi_i^*$$
$$\xi_i, \xi_i^* \ge 0,\ i = 1,\ldots,N$$

iv. Regression SVM Type2:- this is known as nu-SVM regression

In this type of model, the error function is given by:

$$\frac{1}{2}w^T w - C\left[v \in + \frac{1}{N}\sum_{i=1}^{N}(\xi_i + \xi_i^*)\right]$$

This minimizes subject to:

$$\left(w^T\phi(x_i)+b\right) - y_i \le \in + \xi_i$$
$$y_i - \left(w^T\phi(x_i)+b_i\right) \le \in + \xi_i^*$$
$$\xi_i\, \xi_i^* \ge 0, i = 1,\ldots, N \in \ge 0$$

Parameters that control the regression quality are the cost of error C, the width of the tube $\in$ and the mapping function ϕ. According to Sujjaviriyasup, (2013), intuitively, ε value should be small for large sample size than for small sample size and the SVM model value of ε in the ε- insensitive loss function is to be selected and it also affect the smoothness of the SVM response which affect the number of support vectors. It is useful to fix parameter ε by specifying the desired accuracy in advance.

TIJSRAT
ISSN: 2623-7861

In this paper, six model structures were developed to study the model performance of the input variables. The model structures were obtained by setting the input variables equal to the number of variables from the airline passengers' of previous periods, $x_{t-1}, x_{t-2}, \ldots, x_{t-p}$ where p= 2, 4, 6, 8, 10, 12. Table1 show the input structure of the SVM model considered in forecasting airlines passengers' time series.

Table 1 showing the input structure of the models for forecasting time series data

Model	Input structure
M2	$X_t = f(X_{t-1}, X_{t-2})$
M4	$X_t = f(X_{t-1}, X_{t-2}, X_{t-3}, X_{t-4})$
M6	$X_t = f(X_{t-1}, X_{t-2}, X_{t-3}, X_{t-4}, X_{t-5}, X_{t-6})$
M8	$X_t = f(X_{t-1}, X_{t-2}, X_{t-3}, X_{t-4}, X_{t-5}, X_{t-6}, X_{t-7}, X_{t-8})$
M10	$X_t = f(X_{t-1}, X_{t-2}, X_{t-3}, X_{t-4}, X_{t-5}, X_{t-6}, X_{t-7}, X_{t-8}, X_{t-9}, X_{t-10})$
M12	$X_t = f(X_{t-1}, X_{t-2}, X_{t-3}, X_{t-4}, X_{t-5}, X_{t-6}, X_{t-7}, X_{t-8}, X_{t-9}, X_{t-10}, X_{t-11}, X_{t-12})$

In the training and testing of SVM model, the same input structures of the data set (M2, M4, M6, M8, M10 and M12) are used.

Exponential smoothing

Exponential smoothing is the most widely used class of procedures for smoothing discrete time series in order to forecast the immediate future (Makatjane & Moroke, 2016). The work of Rahman et al., (2016) revealed that the idea of exponential smoothing is to smooth the original series the way the moving average does and to use the smoothed series in forecasting future values of the variable of interest. In exponential smoothing, however, we want to allow the more recent values of the series to have greater influence on the forecast of future values than the more distant observations. Exponential smoothing is a simple and pragmatic approach to forecasting, whereby the forecast is constructed from an exponentially weighted average of past observations. The largest weight is given to the present observation, less weight to the immediately preceding observation, even less weight to the observation before that, and so on (exponential decay of influence of past data).

Simple exponential smoothing

This is also known as simple exponential smoothing. Simple smoothing is used for short-range forecasting, usually just one month into the future. The model assumes that the data fluctuates around a reasonably stable mean (no trend or consistent pattern of growth).
The specific formula for simple exponential smoothing is:

$$S_t = \alpha * X_t + (1-\alpha) * S_{t-1} \quad (3)$$

The basic or initial equation for the simple exponential smoothing is given by:

$$F_{t+1} = \alpha\, y_t + (1-\alpha)F_t \quad (4)$$

Where, F_{t+1}= forecast for the next period, α = smoothing constant, y_t = observed value of series in period t, F_t= old forecast for period t.
The forecast F_{t+1} is based on weighting the most recent observation y_t with a weight α and weighting the most recent forecast F_t with a weight of 1- α. Equation 4 can be rewritten as:

$$F_{t+1} = F_t + \alpha\,(y_t - F_t) \quad (5)$$

This is to explain the role of the parameter constant α. The value of smoothing constant α must be between 0 and 1. To estimate α, Forecasts are computed for α equal to 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9 and the sum of squared forecast error is computed for each of them. The value of α with the smallest RMSE is considered for use in producing the future forecasts.

Holts (Double) Exponential Smoothing Method

This method is used when the data shows a trend. Exponential smoothing with a trend works much like simple smoothing except that two components must be updated each period - level and trend. The level is a smoothed estimate of the value of the data at the end of each period. The trend is a smoothed estimate of average growth at the end of each period. The specific formula for double exponential smoothing is:

$$S_t = \alpha * y_t + (1-\alpha) * (S_{t-1} + b_{t-1}) \qquad 0 < \alpha < 1 \quad (6)$$

$$b_t = \Upsilon * (S_t - S_{t-1}) + (1 - \Upsilon) * b_{t-1} \qquad 0 < \Upsilon < 1 \quad (7)$$

The model equation for the double exponential smoothing is given by:
The smoothed estimate of the level

$$L_t = \alpha\, y_t + (1 - \alpha)(L_{t-1} + b_{t-1}) \quad (8)$$

TIJSRAT
ISSN: 2623-7861

The trend estimate:

$$b_t = \beta(L_t - L_{t-i}) + (1 - \beta)b_{t-1} \quad (9)$$

Forecast *m* periods into the future:

$$F_{t+m} = L_t + mb_t \quad (10)$$

Where, L_t = Estimate of the level of the series at time t, α = smoothing constant for the data, y_t = new observation or actual value of series in period t, β = smoothing constant for trend estimate, b_t = estimate of the slope of the series at time t, m = periods to be forecast into the future.

The parameter constants α and β can be selected subjectively or by minimizing a measure of forecast error such as RMSE. The initialization process for Holt's linear exponential smoothing requires two estimates, one to get the first smoothed value for L_1 and the other to get the trend value b_1.

Triple (Winter's) Exponential Smoothing Method,

This method is used when the data shows trend and seasonality. To handle seasonality, we have to add a third parameter. We now introduce a third equation to take care of seasonality. The resulting set of equations is called the "Holt-Winters" (HW) method

$$L_s = \alpha(y_t - S_{t-s}) + (1-\alpha)(L_{s-1} + T_t) \quad (11)$$

The trend estimate

$$T_t = \phi(L_t - L_{t-1}) + (1-\beta)T_{t-1} \quad (12)$$

The seasonality estimate

$$S_t = \gamma(y_t - L_t) + (1-\gamma)S_{t-s} \quad (13)$$

Forecast m period into the future

$$\hat{y}_{t+m} = L_t + mT_t + S_{t+m-s} \quad (14)$$

y_t is the time series at time t, L_t is a linear trend component, S_t is a additive seasonal factor, T_t is smoothed additive trend, *s* is length of seasonality, Let the length of the season be L periods.

As with Holt's linear exponential smoothing, the weights α, β, and γ can be selected subjectively or by minimizing a measure of forecast error such as root mean square error (RMSE). As with all exponential smoothing methods, we need initial values for the components to start the algorithm. To start the algorithm, the

TIMBOU-AFRICA ACADEMIC PUBLICATIONS
AUGUST, 2022 EDITIONS, INTERNATIONAL JOURNAL OF:
SCIENCE RESEARCH AND TECHNOLOGY VOL. 10

initial values for L_t, the trend T_t, and the indices S_t must be set. To determine initial estimates of the seasonal indices we need to use at least one complete season's data (i.e. periods).Therefore, the season's data (i.e. periods) initialize trend and level at period s. Initialize level as:

$$L_s = \frac{1}{S}(y_1 + y_2 + \ldots y_s) \quad (15)$$

Initialize trend as:

$$T_s = \frac{1}{S}\left(\frac{y_{s+1} - y_1}{S} + \frac{y_{s+2} - y_2}{S} + \ldots + \frac{y_{s+s} - y_s}{S} \right) \quad (16)$$

Initialize seasonal indices as:

$$S_1 = \frac{y_t}{L_s},\ S_2 = \frac{y_2}{L_s} \ldots S_t = \frac{y_s}{L_s} \quad (17)$$

Multiplicative model: -This model is used when the data exhibits Multiplicative seasonality. The model is appropriate for series with a linear trend and a multiplicative seasonality effect that depends on the level of the series. Its smoothing parameters are level, trend, and seasonality. The four equations which are used to construct the winter's multiplicative methods are:

The exponentially smoothed series:

$$L_t = \alpha \frac{y_t}{S_{t-S}} + (1 - \alpha)(L_{t-1} + b_{t-1}) \quad (18)$$

The trend estimate:

$$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1} \quad (19)$$

The seasonality estimate:

$$S_t = \gamma \frac{y_t}{L_t} + (1 - \gamma)S_{t-s} \quad (20)$$

Forecast *m* period into the future:

$$\hat{y}_{t+m} = (L_t + mb_t)S_{t+m-s} \quad (21)$$

Where, L_t = level of series, α = smoothing constant for the data, y_t = new observation or actual value in period t, β = smoothing constant for trend estimate, b_t = trend estimate, γ = smoothing constant for seasonality estimate, S_t =seasonal component estimate, m = Number of periods in the forecast lead period, S = length of seasonality (number of periods in the season), F_{t+m} = forecast for m periods into the future.

As in the case with Holt's linear exponential smoothing, the parameters, α, β, and γ can be selected subjectively or by minimizing a measure of forecast error like

TIJSRAT
ISSN: 2623-7861

RMSE. As obtained with the previous exponential smoothing models, we need initial values for the components to begin the computations. To start the computation, the initial values for L_t, the trend b_t, and the indices S_t must be set. To determine initial estimates of the seasonal indices we need to use at least one complete season's data (i.e. s periods). Consequently, we initialize trend and level at period S.

As in the case with Holt's linear exponential smoothing, the parameters, a, b, and g can be selected subjectively or by minimizing a measure of forecast error like RMSE. Consequently, we initialize trend and level at period S.

Initialize level as:

$$L_s = \frac{1}{S}(y_1 + y_2 + \ldots y_s) \quad (22)$$

Initialize trend as:

$$b_s = \frac{1}{S}\left(\frac{y_{s+1} - y_1}{S} + \frac{y_{s+2} - y_2}{S} + \ldots + \frac{y_{s+s} - y_s}{S}\right) \quad (23)$$

Initialize seasonal indices as:

$$S_1 = \frac{y_t}{L_s}, S_2 = \frac{y_2}{L_s}, \ldots, S_t = \frac{y_s}{L_s} \quad (24)$$

Initialize level as:

$$L_s = \frac{1}{S}(y_1 + y_2 + \ldots y_s) \quad (25)$$

Initialize trend as:

$$b_s = \frac{1}{S}\left(\frac{y_{s+1} - y_1}{S} + \frac{y_{s+2} - y_2}{S} + \ldots + \frac{y_{s+s} - y_s}{S}\right) \quad (26)$$

Initialize seasonal indices as:

$$S_1 = \frac{y_t}{L_s}, S_2 = \frac{y_2}{L_s}, \ldots, S_t = \frac{y_s}{L_s} \quad (27)$$

Results and Discussion

This section discusses the results of the analysis and presenting various graphs and tables that gave the various results

The Results for forecasting Airline passengers' using HW model

The graph of the data plotted in figure 4 proves that the data is a multiplicative seasonality time series. For seasonal time series the suitable model of exponential smoothing model is multiplicative seasonality Holt-Winters (HW) model. The plot shows the behavior of the airline passengers' data in the training parts. We can observe a strong seasonality and growth trends in the plot. It could also be seen that the historical data extends to the future with the same pattern.

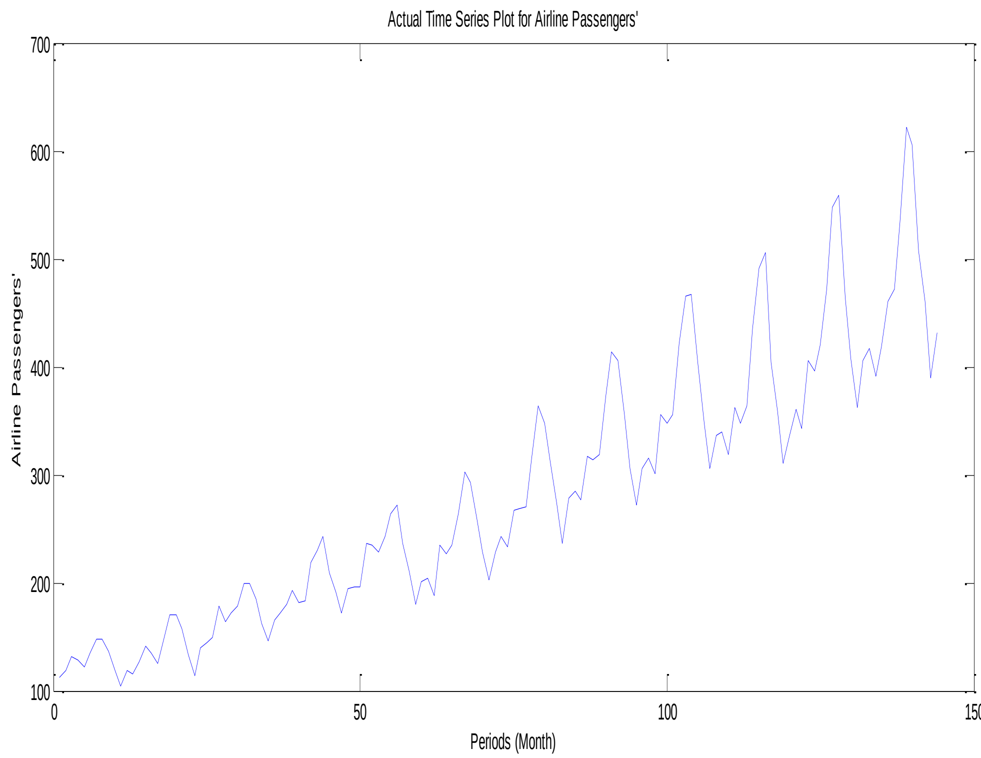

Figure 4: Time series plot of Airline passengers from 1949-1961

Secondly, Figure 5 illustrates the time series plot of the historical data and its forecast in the training part of the analysis. It can be seen that both the historical data and forecast are moving in the same direction. The movement progresses to a level and then descends again. The model predicts the historical data very well.

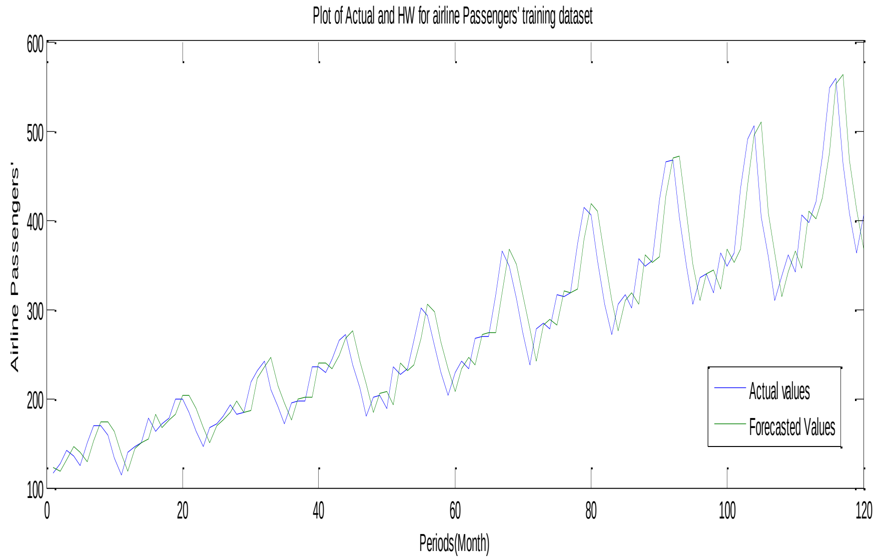

Figure 5: Time series plot of airline passengers' actual data and forecast for training data

Figure 6 shows the time series plot for both the actual data and the forecast in the testing phase which follows the same pattern.

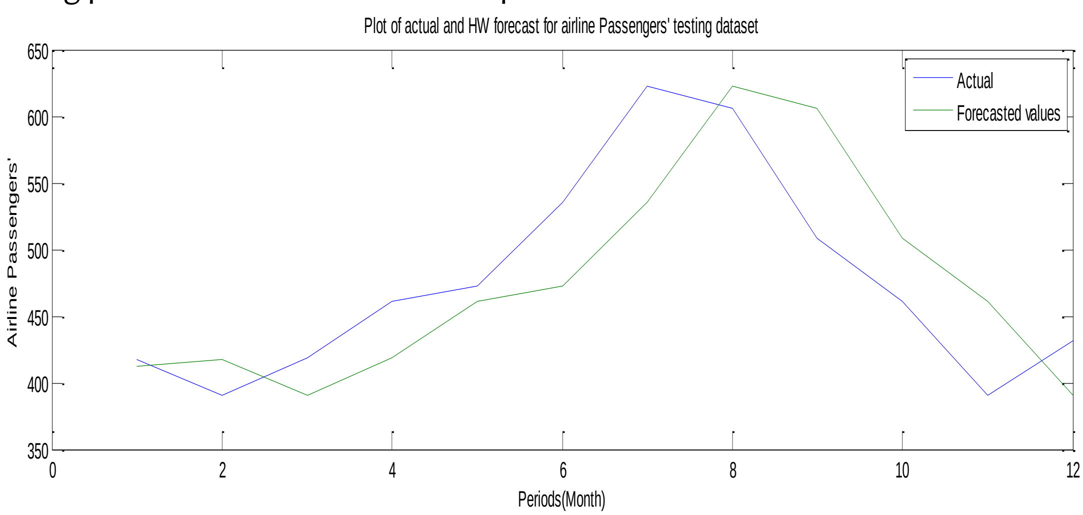

Figure 6: Time series plot of airline passengers' actual data and forecast testing data

Table 2 showing the results of performance of Airline passengers using HW

Measures	Training	Forecasting
MSE	973.3258	996.10
MAE	23.82151	44.63513

R	0.9570	0.7502

For exponential smoothing, parameter constants like α, β and γ are estimated and forecasts are computed which are equal to 0.0001 up to 0.9999 and the sum of squared forecast error is computed for each. The value of any of these parameter constants (α, β and γ) with the smallest MSE (Mean Squared Error) is chosen subjectively or by minimizing a measure of forecast for use in producing the future forecasts. In our study, the parameters used are α = 0.9999, β =0.0001 and γ = 0.1045 which gave the computed value results of analysis of forecasting airline passengers' that involves the computation of measures of performance (MSE, MAE and R) for both training and testing (forecasts) as presented in Table 2 above.

The Results for forecasting Airline passengers' using SVM model

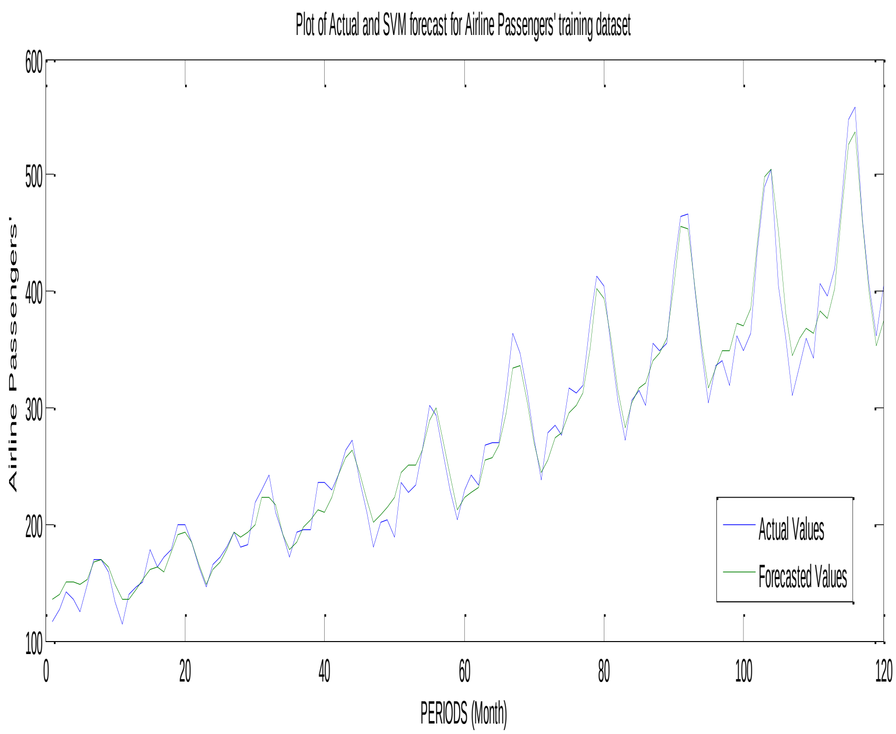

Figure 7: Time series plot of Airlines data and SVM (training)

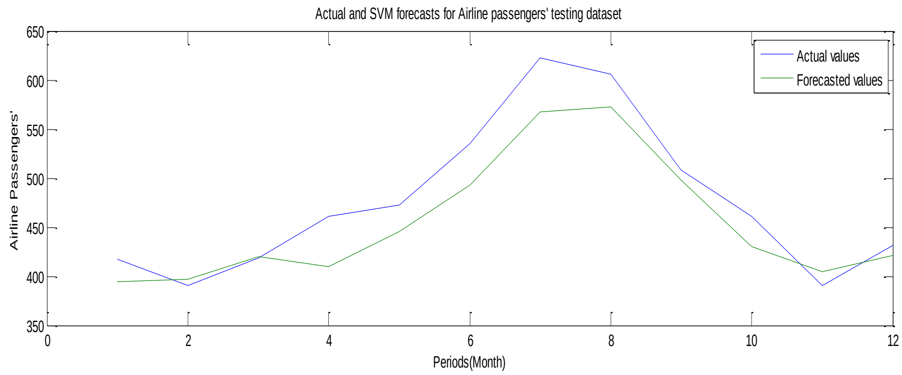

Figure 8 : Time series plot of Airlines data and SVM (testing)

Table 3 showing the results of performance of Airline passengers using SVM

Inputs	Measures	Training	Testing
2	MSE MAE R	745.7402 22.37584 0.9660	3153.7390 46.91967 0.8524
4	MSE MAE R	739.8490 22.07039 0.9663	3227.7190 47.31532 0.8257
6	MSE MAE R	714.3565 22.07702 0.9679	3342.8690 45.35504 0.8747
8	MSE MAE R	729.9602 21.74191 0.9668	3068.6220 41.27822 0.8590
10	MSE MAE R	512.9342 19.15990 0.9760	2141.3500 36.36567 0.8959
12	MSE MAE R	208.4179 11.38535 0.9900	**925.9196** **25.24209** **0.8979**

From table 3, the best model for the forecast in the testing phase is input 12 with MSE = 925.9196, MAE = 25.24209 and R =0.8979 and in the training phase, the best model is also input 12 with MSE= 208.4179, MAE = 11.38535, and R = 0.9900. These results are obtained using various inputs as contained and explained in table 2.1 under the methodology.

The Approach of HW model to Time Series modelling

For a quite number of years exponential smoothing models have been used in different aspects of time series forecasting. In the area of exponential smoothing model which is assumed to be linear, there are three types of exponential smoothing namely, simple or single, double or Holts' and triple or winters (referred to in our study as seasonal exponential smoothing, HW) exponential smoothing. This model is one of the popular forecasting models in various areas of time series research. In this study, we have applied the HW whose components can be additive or multiplicative models.

HW model which is a three parameter model is used when data exhibits seasonality. HW has three smoothing parameters namely α which is the smoothing parameter for the level, β is the smoothing parameter for the trend and γ is the smoothing parameter for the seasonal index. The range of these parameters is between 0 and 1. i.e., ($0 < \alpha < 1$, $0 < \beta < 1$ and $0 < \gamma < 1$). These parameter constants are estimated and forecasts are obtained and they are equal to 0.0001 to 0.9999, then the sum of squared forecast error is computed for each and the value of any of these parameters with the smallest MSE, MAE and the highest R is then selected as the best model in the training and testing phases. The results are summarized on the respective tables.

The Approach of SVM model to Time Series modelling

SVM models have been used in the time series forecasting and is being used to develop models using the airline passengers' data has been normalized. In selecting kernel function for this study, we used Radial Basis Function (RBF), where γ is the parameter of the kernel. For the SVM model, three parameters (C, ε and γ) were determined. The range of these parameters were set to (1, 10) with increments of 1.0 for C and (0.1, 0.5) with increments of 0.1 for ε and γ is fixed as 0.5. With the use of STATISTICA software, the computation was performed for the three data sets for SVM with the number of input 2 to 12 and the results obtained for MSE, MAE and R in which the SVM model with the minimum value for MSE, MAE and the maximum value for R is selected to be the best. These results are presented in the appropriate tables.

Forecast evaluation methods

The performance of the models in forecasting seasonal time series in training and testing are evaluated by using mean squared error (MSE), mean average error (MAE), and correlation coefficient (R). All these means of evaluation are widely used in evaluating results of time series forecasting (Dawson et al., 2007). Similarly, (Hyndman, 2014), and Shcherbakov & Brebels, (2013) and Propper & Wilson, (2003) affirmed the use of these standard statistical measures of the forecast accuracy as very useful.

The criteria for judging the best model are how relatively small these are in both the training and testing of the data. This is necessary to quantify the amount by which an estimator differs from the original (true) value. That is why the measure with smaller value is usually selected as the best. The correlation coefficient (R) measures the strength of relationship between the predicted values and the observed values and the degree to which the two values are linearly related, the strength is usually considered best as the value approaches 1.

The following forecasting accuracy measures will be considered to make comparisons among the exponential smoothing and SVM models. MSE, MAE and R are given by

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2 \qquad (28)$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}| y_t - \hat{y}_t | \qquad (29)$$

and $$R = \frac{\sum_{t=1}^{n}(y_t - \bar{y}_t)(\hat{y}_t - \bar{\hat{y}}_t)}{\left[\sum_{t=1}^{n}(y_t - \bar{y}_t)^2\right]^{1/2}\left[\sum_{t=1}^{n}(\hat{y}_t - \bar{\hat{y}}_t)^2\right]^{1/2}} \qquad (30)$$

Where $\hat{y}_t$ is the predicted value, y_t is the actual value at time *t*, and *n* is the number of Predictions.

Conclusion

TIJSRAT
ISSN: 2623-7861

Time series forecasting is a vital tool in organizational planning. There are many methods that can be used to forecast seasonal time series data. The application of time series for forecasting has become very important in recent years. Time series forecasting has become one of the most important quantitative models that has received prominence in the modern research. Therefore, using quantitative methods to make forecasts can assist in decision making. HW and SVM methodology have shown to be very effective in this direction especially in time series forecasting with some degree of accuracy nonetheless their performance may not be satisfactory.

The application of HW and SVM models in the field of forecasting time series data is studied and the accuracy of the forecasts in terms of MSE, MAE and R have been computed and compared among the two models. The result from table 3 gave an indication that the best model for the forecast in the testing phase is input 12 with MSE 925.9196; MAE 25.24209 and R 0.8979 and in the training phase, the best model is also input 12 with MSE 208.4179; MAE 11.38535 and R 0.9900. These results are obtained using various inputs as contained and explained under the methodology.

The result of comparison shows that the SVM model outperformed the HW using the evaluation performance criteria. SVM model is therefore recommended to be used in forecasting the airlines passengers' data.

References

Changxiu, Z. C. and Xu xiaoling, W. W. (2002). A novel approach based on support vector machine to forecasting the quality of friction welding. *4th World congress on intelligent control and automation*.

Chen, K., & Wang, C. (2007). *A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan*. *32*, 254–264. https://doi.org/10.1016/j.eswa.2005.11.027

Cordeiro, C., & Neves, M. (2009). Forecasting time series with Boot. EXPOS procedure. *Revstat*, *7*(2), 135–149.

Cortes, C., & Vapnik, V. (1995). *Support-Vector Networks*. *297*, 273–297.

Dawson, C. W., Abrahart, R. J., & See, L. M. (2007). HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environmental Modelling and Software*, *22*(7), 1034–1052. https://doi.org/10.1016/j.envsoft.2006.06.008

Gelper, S., Fried, R., & Croux, C. (n.d.). *Computational aspects of robust holt-winters smoothing based on m -estimation*. *2001*, 1–13.

Hyndman, L. R. J. (2014). *Forecasting : Principles & Practice*. *September*.

Ismail, Z., Ismail, Z., Yahya, A., & Shabri, A. (2009). Forecasting Gold Prices Using Multiple Linear Regression Method. *American Journal of Applied Sciences*, *6*(8), 1509–1514. https://doi.org/10.3844/ajassp.2009.1509.1514

Kalekar, P. S., & Bernard, P. (2004). *Time series Forecasting using Holt-Winters Exponential Smoothing Under the guidance of*. *04329008*, 1–13.

Makatjane, K. D., & Moroke, N. D. (2016). *COMPARATIVE STUDY OF HOLT-WINTERS TRIPLE EXPONENTIAL SMOOTHING AND SEASONAL ARIMA : FORECASTING SHORT TERM SEASONAL CAR SALES IN SOUTH AFRICA*. *6*(1), 71–82.

Pai, P., Lin, K., Lin, C., & Chang, P. (2010). Expert Systems with Applications Time series forecasting by a seasonal support vector regression model. *Expert Systems With Applications*, *37*(6), 4261–4265. https://doi.org/10.1016/j.eswa.2009.11.076

Propper, C., & Wilson, D. (2003). The use and usefulness of performance measures in the public sector. *Oxford Review of Economic Policy*, *19*(2), 250–267. https://doi.org/10.1093/oxrep/19.2.250

Rahman, H., Salma, U., & Hossain, M. (2016). *Revenue Forecasting using Holt – Winters Exponential Smoothing*. *December*.

Shcherbakov, M. V., & Brebels, A. (2013). *A Survey of Forecast Error Measures*. *24*(4), 171–176. https://doi.org/10.5829/idosi.wasj.2013.24.itmies.80032

Sujjaviriyasup, T. (2013). *Hybrid ARIMA-Support Vector Machine Model for Agricultural Production Planning*. *7*(57), 2833–2840.

Supriatna, A., Susanti, D., & Hertini, E. (2017). *Application of Holt exponential smoothing and ARIMA method for data population in West Java Application of Holt exponential smoothing and ARIMA method for data population in West Java*. https://doi.org/10.1088/1742-6596/755/1/011001

Tian, Y., Shi, Y., & Liu, X. (2012). *Recent advances on support vector machines research*. *18*(1), 5–33. https://doi.org/10.3846/20294913.2012.661205

Vapnik, V. N. (1999). *An Overview of Statistical Learning Theory*. *10*(5), 988–999.

Wang, L. (2005). *Support Vector Machines : Theory and Applications*.

Wang, S., Feng, B., & Liu, G. (2013). Application of seasonal time series model in the precipitation forecast ☆. *Mathematical and Computer Modelling*, *58*(3–4), 677–683. https://doi.org/10.1016/j.mcm.2011.10.034

Zhang, G. P. (2003). *Time series forecasting using a hybrid ARIMA and neural network model*. *50*, 159–175.