JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

TIME SERIES FORECASTING FOR THAI TOURISM DATASET

[1] Rahul Kumar, [2] Bijay Singh [3] Preeti Sharma
[1] Department of Computer Science and Engineering, NIET Greater Noida ,India
2 Department of Computer Science and Engineering,ARKA Jain University,Jharkhand
[2]Department of Printing Technology and Engineering , SITM ,Rewari
email: ***Rahulgate59@gmail.com***; bijaym.tech13@gmail.com preetisharmag1995@gmail.com

Abstract : Tourists' get attracted towards Thailand because of its diverse culture and geography. Apart from heritage and culture, the tourists from all over the world come here for various other purposes like medical, business, education and sports. The tourism industry of Thailand is economically important and is growing rapidly. The tourism industry in Thailand helps in the growth of other sectors like agriculture, small scale industries, self-employment, etc. This makes forecasting of tourists' arrivals in Thailand a prime focus of the government. Forecasting is the process of making predictions of the future based on past and present data and analysis of trends. Tourism forecasting plays an important role in providing awareness and support for future development of the Thailand tourism industry. In this paper, an attempt has been made to forecast tourists' arrival using ARIMA, Holtwinter and Naïve Bayes Methods.

Keywords—Forecasting; Holt-Winter's exponential smoothing model; ARIMA; Auto regression; Moving average; Naïve Bayes

I. INTRODUCTION

Modeling and forecasting tourists demand has received substantial attention among policy makers, hospitality management, researchers and other interest groups globally. Tourist's footfalls immensely contribute towards the growth of economy's Gross Domestic Product (GDP). It is one of the leading sources of foreign exchange earnings as well as generating employments opportunities. Despite the continued existence of number of antisocial activities and economic setbacks such as terrorism, Naxal activities etc., India is still a wonderful and attractive destination for international tourists. Forecasting tourist arrival being a significant activity for its beneficiaries and stake holders', several forecasting models have been applied to estimate and forecast the tourism demand globally. Large numbers of research papers have applied widespread time series models for forecasting tourism demand globally. Highly structured and an extensive survey of literature of earlier studies is provided by Crouch [1], Li et al. [2], Witt et al. [3].

On the other hand, Song et al. review the literature for post 2000 studies. This paper highlights few recent studies which applied time series model for forecasting tourism demand. The studies of Smeral et al. [4] applied ARIMA, SARIMA and naïve methods for forecasting tourism demand. The result reveals that advanced models like ARIMA or SARIMA model could not even outperform the simple Naïve model. Applying ARCH and GARCH model Chan et al. [6] tries to estimate and forecast volatility in tourism demand and its affect to various shocks. On the other hand, Turner et al. have applied the structural equation modeling. Cho [9] concluded that artificial Neural Network

(ANN) model outperform the exponential smoothing and ARIMA mode in modeling and forecasting the tourism demand for Hong Kong.
The main objective of the study is to forecast tourists' arrival in India using statistical time series modeling techniques- Holt Winters method and ARIMA modeling. Further comparative analysis of the both the methods is done on the basis of certain performance metrics.

II. DATASET AND METHODS

A. Dataset Description

The annual tourists' arrival in Thailand for the period 2010-2016 is collected from Thailand Tourism Statistics, Ministry of Tourism, and Government of Thailand. Dataset contains the data of tourist coming from 53 regions. The prediction period is from 2017-2020. ARIMA, Holt Winter's Exponential Smoothing and ETS models are statistical tools used for forecasting of the number of tourists' arrival in Thailand from 2017-2020.The dataset format is given in Table-1.

Table-1. Dataset format of Thai tourism

region	nationality	year	month	tourists
Africa	AfrOthers	2010	1	6553
Africa	AfrOthers	2010	2	5618
Africa	AfrOthers	2010	3	6689
Africa	AfrOthers	2010	4	5210
Africa	AfrOthers	2010	5	4537
Africa	AfrOthers	2010	6	4683
Africa	AfrOthers	2010	7	6323
Africa	AfrOthers	2010	8	5717
Africa	AfrOthers	2010	9	5951
Africa	AfrOthers	2010	10	6327
Africa	AfrOthers	2010	11	6356
Africa	AfrOthers	2010	12	6866
Africa	AfrOthers	2011	1	6547
Africa	AfrOthers	2011	2	5143
Africa	AfrOthers	2011	3	6374
Africa	AfrOthers	2011	4	5821
Africa	AfrOthers	2011	5	5513
Africa	AfrOthers	2011	6	5838
Africa	AfrOthers	2011	7	7005

B. Stationary Time Series

Stationary time series is a random process whose joint probability distribution does not change with time. There are several methods that can be used to convert non-stationary series to stationary series. However, the most widely used variance stabilization method is Box-Cox transformation. The Augmented Dickey Fuller (ADF) test, Phillip -Perron (P-P) test and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test can be used to check whether the series is stationary or not

Test Name	Value	Lag order	P value
Dickey- Fuller Test	-3.7821	4	0.02369
KPSS Test	1.4149	3	0.01

C. Seasonal and Non-seasonal unit root testing

This section examines whether the y_t series contains unit roots at the seasonal and non-seasonal (annual) frequencies. The existence of one or more unit roots in a non-seasonal time series implies that the series is non stationary. This characteristic creates a number of problems for determining an appropriate forecasting model. It is now customary to transform the series into a stationary process by the application of various filters that difference the data at first or higher order levels.

Given a seasonal time series, the presence of unit roots at the seasonal frequency implies that the seasonal fluctuations change over time. This is clearly relevant to the issue of climate change, where a major concern is whether seasonal patterns are changing.

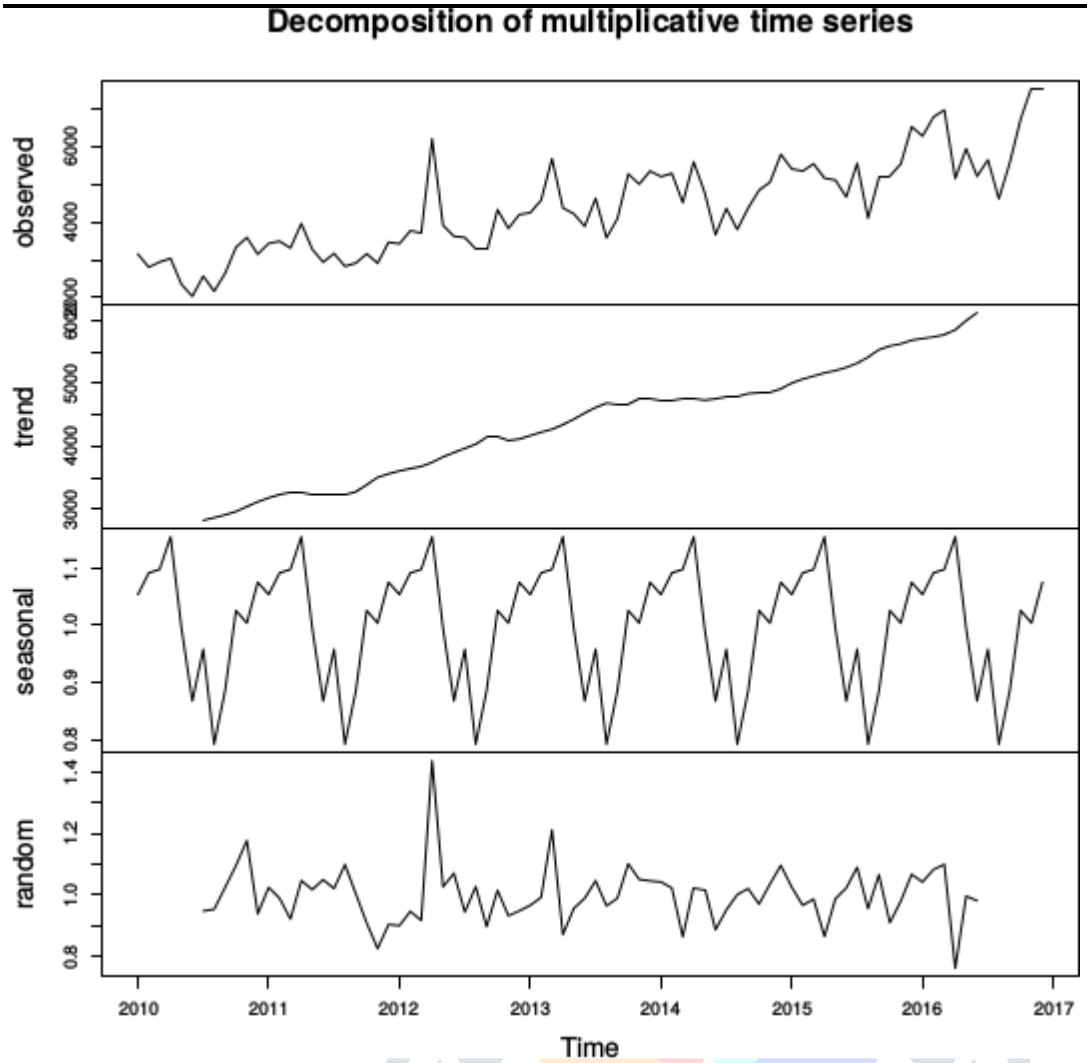

Figure- 2. Seasonality decomposition plot.

D. ACF and PACF

Autocorrelation and partial autocorrelation function are a type of graphs that contain correlations of different time lags. ACF and PACF can be used to determine the behaviors of the series, whether stationary or not and to identify the number of components in an ARMA model. The number of significant spikes in the ACF indicates the number of MA terms in the model, while the number of significant spikes in PACF indicates the number of AR terms in the model.

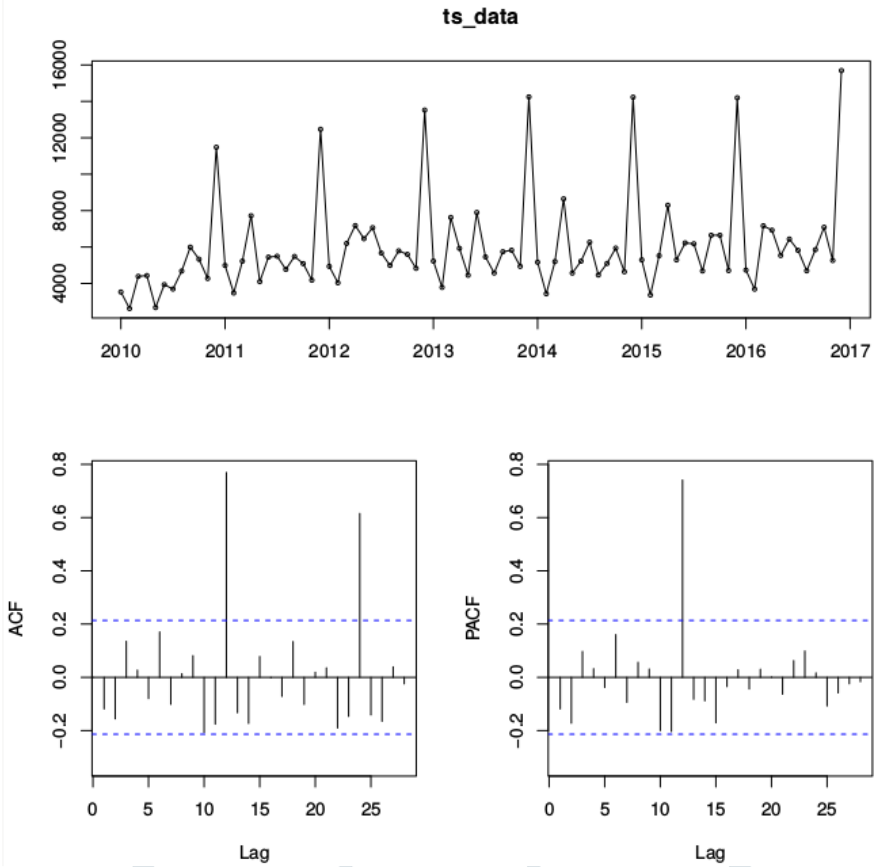

Figure-1. ACF and PACF plot.

E. Holt winters exponential smoothing (HWES)

Exponential smoothing modeling's are simple, fast and inexpensive. They are used frequently throughout the world. Exponential smoothing methods are a class of methods that produce forecasts with simple formulae, taking into account trend and seasonal effects of the data. These procedures are widely used as forecasting techniques in inventory management and sales forecasting. Ord et al. [5] had put exponential smoothing procedures on sound theoretical ground by identifying and examining the underlying statistical models.

The HWES method estimates three smoothing parameters, associated with level, trend and seasonal factors. The seasonal variation can be of either an additive or multiplicative form. The multiplicative version is used more widely and on average works better than the additive [5]. If a data series contains some values equal to zero, the multiplicative method may not be used. In such cases additive Holtwinters forecasting model is used. A problem which affects all exponential smoothing methods is the selection of smoothing parameters and initial values, so that forecast is better in accord with time series data. The parameters of smoothing (and initial) in HWES are estimated by minimizing the mean square error (MSE). The Holt Winters' Exponential Smoothing Model is given in equations (1)

Level : $E_t = \alpha\,(Y_t - S_{t-p}) + (1-\alpha)(E_{t-1} - T_{t-1})$

Trend : $T_t = \beta\,(E_t - E_{t-1}) + (1-\beta)\,T_{t-1}$ (1) Seasonality : $S_1 = \gamma\,(Y_t - E_t) + (1-\gamma)\,S_{t-p}$

Forecast : $F_{t+1} = (E_t + mT_t) + S_{t-p+m}$

Where α, β, γ are smoothing constants and are chosen so that MSE is minimized.

F. **Auto regressive integrated moving average models (ARIMA)**

Univariate ARIMA models use only the information contained in the series itself. Thus, models are constructed as linear functions of past values of the series and/or previous random shocks (or errors). Forecasts are generated under the assumption that the past history could be translated into predictions for the future. The ARIMA model uses the fact that arrival of tourist's is a stochastic time series. This modeling regresses the dependent variable Yt on p-lags of the dependent variable (Autoregressive) and q lags of the error term (Moving Average). Sometimes instead of dependent variable Y_t, $L_d Y_t$ can be used as the dependent variable. Here L is the one step lag operator, i.e., $LY_t = Y_{t-1}$.

$(1 - \Sigma\,\alpha_k L_k)\,(1 - L)_d X_t = (1 + \Sigma_q \beta_k L_k)$ (2)

Where ε_t is white noise error? It is identically and independently distributed with mean zero and common variance σ_2 across all observations. In ARIMA model following steps are followed:

Step 1: Model identification: According to Box and Jenkins [3] two graphical procedures are used to access the correlation between the observations within a single time series data. These devices are called an estimated autocorrelation functions and the estimated partial autocorrelation function. These two procedures measure statistical relationships within the time series data. Next step for identification is summarization of statistical correlation within the time series data. One has to choose the appropriate ARIMA model from the whole family of ARIMA as suggested by Box and Jenkins. The autocorrelation function (ACF) and partial autocorrelation functions (PACF) of a series together are the most powerful tool, usually applied to reveal the correct values of the parameters. The ACF gives the autocorrelations calculated at lags 1, 2 and so on, while PACF gives the corresponding partial autocorrelations, controlling the autocorrelations at intervening lags. Every ARIMA model have their unique ACF and PACF associated with it. One has to select the model who's theoretical ACF and PACF resembles the anticipated ACF and PACF of the time series data [1].

Step 2: Parameter estimation: Maximum Likelihood Estimation Method (MLE) or Modified Least Squares Method (MLS), whichever suitable for the time series data is used to estimate the coefficients of the model. The final results include the parameter estimates, standard errors, estimates of residual variance, standard error of the estimate, natural log likelihood, Akaike's Information Criterion (AIC). Model selection is based on the minimization of AIC. To identify the optimal ARIMA model, different combinations of AR and MA are tested. The one for which AIC have minimum values are considered to be optimal model. AIC is given by:
AIC = -2logL + 2m (3)
m=p+q and L is likelihood function (3).

Step 3: Diagnostic checking: Diagnostic checks help to determine if the anticipated model is adequate. In this step, an examination of the residuals from the fitted model is done and if it fails the diagnostic tests, it is rejected and one have to repeat the cycle until appropriate models is achieved.

Step 4: Forecast: These models are regression models that use lagged values of the dependent variables and/or random distributing term as explanatory models. These models rely heavily on the auto correlation pattern in the data. This model regresses the dependent variable on p lags of the dependent variable (Auto Regressive) and q lags of the error term (Moving Average).

Performance evaluation: To evaluate the performance of the various models the Root Mean Square Error (RMSE) and the Mean Absolute Percentage Error (MAPE) are used, which are as follows:

$$RMSE = \frac{1}{n}\sum_{i=1}^{n} (Y_t - F_t)^2 \quad (4)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left| \frac{Y_t - F_t}{Y_t} \right| \times 100 \quad (5)$$

Where Y_t is the observed value and F_t is the forecast value and n is the number of time period used as forecasting.

G. ETS model

Point forecasts are obtained from the models by iterating the equations for\ (t=T+1,\dots, T+h\) and setting all \(\varepsilon_t = 0\) for \(t>T\). These forecasts are identical to the forecasts from Holt's linear method, and also to those from model ETS(A,A,N)[6]. Thus, the point forecasts obtained from the method and from the two models that underlie the method are identical (assuming that the same parameter values are used).
ETS point forecasts are equal to the medians of the forecast distributions. For models with only additive components, the forecast distributions are normal, so the medians and means are equal [6]. For ETS models with multiplicative errors, or with multiplicative seasonality, the point forecasts will not be equal to the means of the forecast distributions.

III. Result and Discussions

The main aim of our project is to predict number of tourists' arrival in Thailand and compare the three forecasting models. Table 2, shows the error measure of three different method. Figure- 3,4 and 5 shows the time series forecasting plot using ARIMA, ETS and HoltWinters method respectively.

A. Forecasting using ARIMA model

ARIMA uses the fact that foreign tourists' arrival is a stochastic time series. In this project we have implemented ARIMA model using R-programming. The forecasted plot using ARIMA is shown in Figure-3.

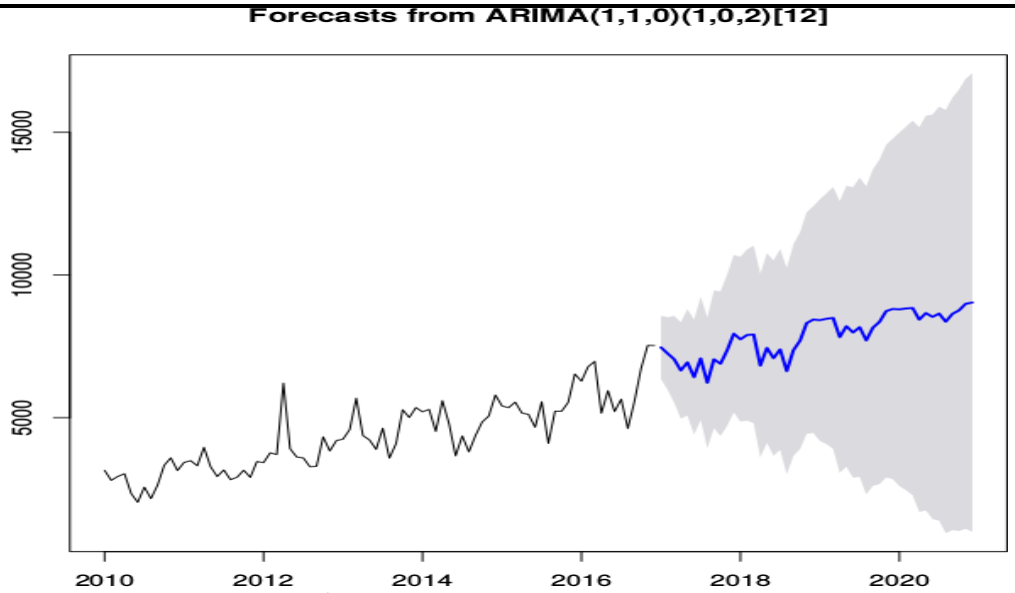

Figure-3. Forecasting using ARIMA.

B. Forecasting using ETS

ETS(Error, Trend, Seasonal) model focuses on trend and seasonal components. The flexibility of the ETS model lies in its ability to trend and seasonal components of different traits. The forecasted plot using ETS is shown in Figure-4.

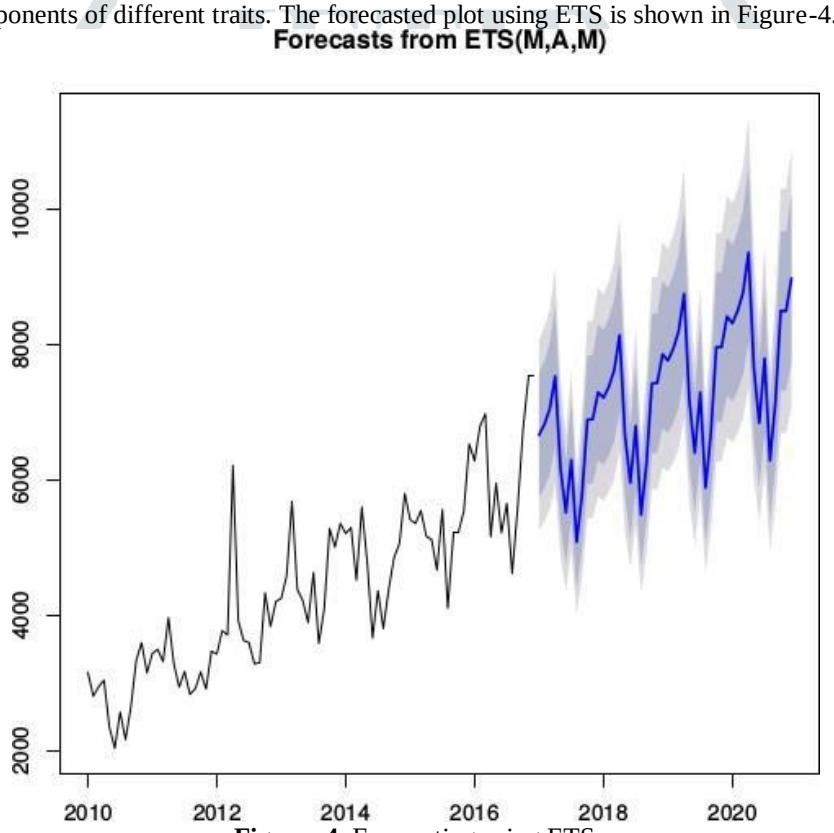

Figure- 4. Forecasting using ETS

C. Forecasting using Holt-winter's exponential smoothing (HWES)

HWES model is also appropriate when trend and seasonality are present in the time series. It decomposes the series down into three components that are base, trend and seasonal components. We have implemented HWES using R-programming. The forecasted plot using HWES is shown in Figure- 5.

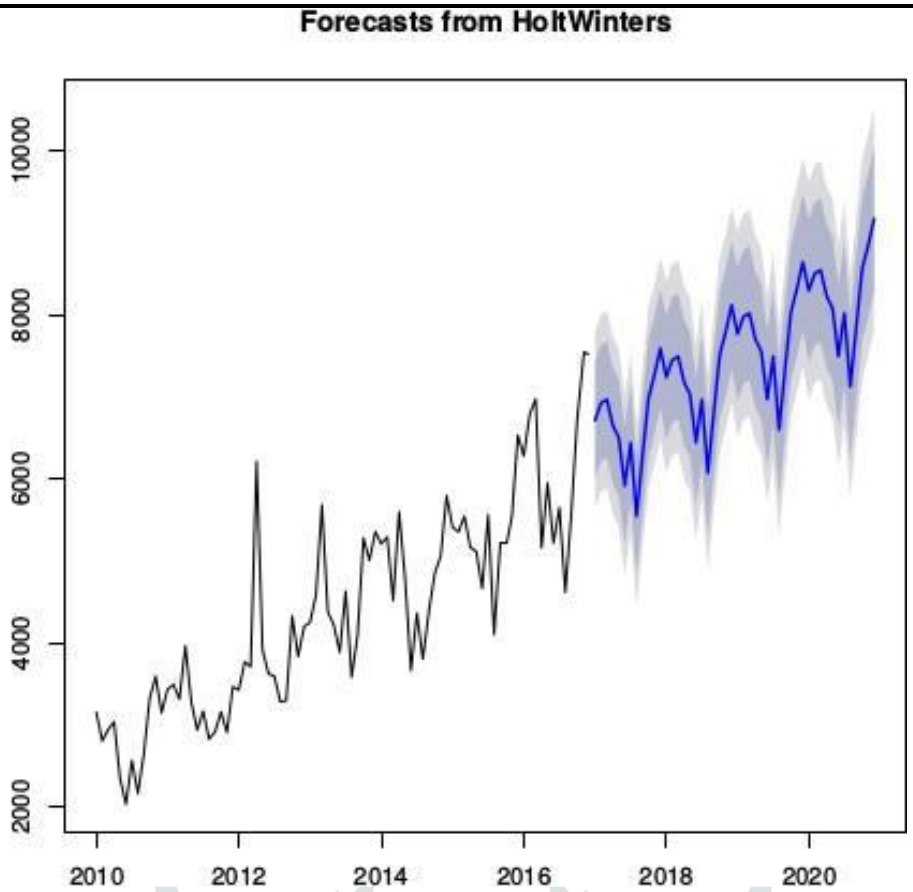

Figure- 5. Forecasting using HoltWinters.

D. Performance Measure

The accuracy of any model can be measured using ME, RMSE, MAE, MPE, MAPE, MASE and ACF1. To estimate which model is best fit for the given data we have tabulated the above given accuracy factor in the below TABLE-2.

TABLE-2. Error measure of different method.

ERROR FACTOR	ARIMA	ETS	HWES
ME	39.22914	-11.7759	61.64748
RMSE	543.6277	465.8642	549.4518
MAE	403.5192	332.5409	395.3534
MPE	-0.19810	-1.42877	0.324539
MAPE	9.162874	7.52197	8.352323
MASE	0.573305	0.472461	0.561703
ACF1	-0.03501	0.111452	0.110006

IV. CONCLUSION

The aim our project is to forecast tourists' arrival in Thailand and to compare ETS, HWES and ARIMA model result based on MAPE and RMSE. ETS and HWES model both are quite efficient for forecasting foreign tourists' arrival in Thailand. On the basis of results obtained ETS model is better than Holt Winter's Exponential Smoothing. Therefore, ETS model is found to be the best fit model for foreign tourists' arrival in Thailand.

REFERENCES

[1] Lim, Christine, and Michael McAleer. "Time series forecasts of international travel demand for Australia." *Tourism Management* 23.4 (2002): 389-396.

[2] Song, Haiyan, et al. "Forecasting tourist arrivals using time-varying parameter structural time series models." *International Journal of Forecasting* 27.3 (2011): 855-869.

[3] Li, Gang, et al. "Tourism demand forecasting: A time varying parameter error correction model." *Journal of Travel Research*45.2 (2006): 175-185.

[4] Song, Haiyan, and Gang Li. "Tourism demand modelling and forecasting—A review of recent research." *Tourism management* 29.2 (2008): 203-220.

[5] Koehler, Anne B., Ralph D. Snyder, and J. Keith Ord. "Forecasting models and prediction intervals for the multiplicative Holt–Winters method." *International Journal of Forecasting* 17.2 (2001): 269-286.

[6] Kourentzes, Nikolaos, Fotios Petropoulos, and Juan R. Trapero. "Improving forecasting by estimating time series structural components across multiple frequencies." *International Journal of Forecasting* 30.2 (2014): 291-302