Proceedings of Science and Mathematics
Faculty of Science, Universiti Teknologi Malaysia
Vol. 2, 2021, page 193-202

Forecasting of Loans Approved by Purpose for Banking System in Malaysia

[1]Naqibah Aminuddin Jafry and [2]Nur Arina Bazilah Kamisan

[1,2]Department of Mathematical Sciences
Faculty of Science, Universiti Teknologi Malaysia,
81310 Johor Bahru, Johor, Malaysia.

e-mail: [1]naqibah@graduate.utm.my, [2]nurarinabazilah@utm.my

Abstract Forecasting time series data is a common occurrence that needs to be addressed by researchers. In this study, a hybrid forecasting model was developed and several existing forecasting techniques such as Seasonal Auto Regressive Integrated Moving Average (SARIMA) and Exponential Smoothing (ES) method were utilized to make future prediction of the Loans Approved by Purpose for Banking System in Malaysia for Purchase of Transport Vehicles. The accuracy of the forecasted values is measured using Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). From the findings, hybrid model was found to be the best forecasting model for the loan approved by purpose for banking system in Malaysia for transport vehicles, followed by ES model, and SARIMA.

Keywords Forecasting; loans; banking; SARIMA; Holt-Winters; hybrid; time series.

1 Introduction

Forecasting is an approach that relies on historical data to generate reliable prediction which will be used to assess the course of future trends. It has long been used by decision makers to predict the term structure of government bond yields [1], the number of tourist arrivals [2], and the list goes on. Banking sectors plays a vital part in developing the growth of the country's economy, because banks are amongst the entities that contributes significantly to the country's economic development as well as providing credit facility or loan [3]. ARIMA was used to forecast the number of personal loan consumers at Bhutan Development Bank in Bhutan as they believe the future prediction will benefit and assist the responsible party to anticipate and function accordingly [3]. As the bank will gain from the interest on the loans they credit, the gain or loss of a borrower rests on a vast number of loans, whether the consumer pays back the loan or defaulting [4].

ES is amongst the most renowned conventional forecasting approach to predict a time series univariate data. ES is a forecasting technique that calculates the observed time series unequally; where latest observations have more weight compared to the older and distant observations [5]. Simple Exponential Smoothing (SES) method is the simplest technique of Exponential Smoothing method, since it assumes that series has no trend and seasonal pattern [6]. Double Exponential Smoothing (DES) method is suitable to forecast data with trend. If the data exhibits both trend and seasonality, Triple Exponential Smoothing or Holt-Winters method is deemed to be more reliable for future predictions. Forecast accuracy were measured using the primary production of electricity data in Slovakia and they computed the smoothing constants with an increment of 0.01 every time to determine the most optimal parameters of the smoothing constant [5]. SES technique will only yield favourable results for short-term forecasting and if the data has no seasonal pattern and trend. Forecasting performance of Two Mediterranean Islands where Actual Static (AS) (econometric) model, DES, Holt-Winters and Autoregressive Moving Averages (ARMA) models were considered [6]. They concluded that the most optimal forecasting technique for nearly all cases of data with seasonality is Holt-Winters method. Decomposition, Holt-Winters and Seasonal ARIMA (SARIMA) were chosen to be compared in the study to forecast monthly exchange rate in Nigeria [7]. The result from the study revealed that Holt-Winters yields the best forecasting results as it depicts the lowest RMSE values.

Box-Jenkins or ARIMA is a conventional method to find the best fit of a time series model using past values of time series. There are three stages to model this approach, which are model identification, parameters estimation and selection, and adequacy checking. Then, the best fitted ARIMA model will be selected for forecasting the time series data. ARIMA was used to forecast the credit loan from households in South Korea and the fitted ARMA (1,5) model demonstrates excellent performance when the result yields the lowest MAPE, RMSE, and MAE [8]. ARIMA were applied by researchers in estimating sugarcane production in India [9]. The authors highlighted that ARIMA is very famous since it can generate reliable predictions based on summary of historical univariate data and manages to represent both stationary and non-stationary time series. ARIMA model was selected to forecast stock prices in India from different sectors [10]. The authors concluded that the accuracy of ARIMA model is above 85% in forecasting stock prices in India for seven different sectors.

SARIMA is an extension of ARIMA model. This forecasting technique is appropriate when the data exhibit trend and seasonality. By comparing four error metrics, a study in a Malaysian university were conducted to establish the optimum model for projecting energy load demand for 3 months [11]. Holt-Exponential Winter's Smoothing (HWES) and SARIMA are two forecasting methodologies that were utilized in the study. The results revealed that the dataset has seasonal and trend components and SARIMA $(0,0,1)(1,0,0)_{12}$ models was found to be the best model and were chosen to forecast power load demand since it had the lowest error measurement when compared to HWES.

An extremely favourable performance of forecasting results will be attained if a hybrid approach is applied such that ARIMA technique is paired with a neural network [12]. In this case, the hybrid of ANN and ARIMA were developed to predict the future index value and trend of Indian stock market and the hybrid results in better forecast as compared to each of the individual model itself [13]. Financial time series are not easy to forecast since most of the data are dynamic, non-linear and complicated, hence it is not suitable to be forecasted using traditional model such as ARIMA [14]. The combination of linear ARIMA model with non-linear models such as Support Vector Machines (SVM), ANN and random forest (RF) models were developed to forecast the stock index returns. The performance of three hybrid models, ARIMA-SVM, ARIMA-ANN and ARIMA-RF were compared with the individual models which are ARIMA, SVM, ANN and RF models and the authors discovered that the hybrid ARIMA-SVM model is the best model to forecast the stock index returns as it yields the highest accuracy and better returns than the other models.

This study aims to find the best forecasting model for loans approved by purpose for banking system in Malaysia between three models which are SARIMA, Exponential Smoothing and a hybrid model. The forecasting performance of the models will be measured using MAPE, RMSE, and MAE.

2 Methodology

2.1 Exponential Smoothing

2.1.1 Double Exponential Smoothing

This study will use DES for the purpose of hybrid forecasting model. The formula of DES can be seen as below:

$$L_t = \alpha y_t + (1-\alpha)(L_{t-1} + T_{t-1}) \tag{1}$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \tag{2}$$

where α and β are smoothing constants between 0 and 1. While, L_t represent the level series, T_t is the level value at time t. The k-step-ahead of forecast at time t is calculated as:

$$F_{t+k} = L_t + kT_t \tag{3}$$

where,

y_t: the actual values which include seasonality,
L_t: level component of the series,
α: the smoothing constant for level where $0 < \alpha < 1$,
β: the smoothing constant for the trend estimation where $0 < \beta < 1$,
k: number of steps ahead to be forecast,
F_{t+k}: forecast for k-step-ahead.

2.1.2 Additive Holt-Winters Exponential Smoothing

This study will use the Holt-Winters method since the data exhibits trend and seasonal pattern. Three smoothing constants which are α, β, and γ are going to be used. There are two types of Holt-Winters method which include Multiplication Holt-Winters method and Additive Holt-Winters method. This study will be focusing on Additive Holt-Winters method and the general formula of Additive Holt-Winters model can be expressed as:

$$L_t = \alpha(y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1}) \tag{4}$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \tag{5}$$

$$S_t = \gamma(y_t - L_t) + (1-\gamma)S_{t-s} \tag{6}$$

where L_t represent the level series, b_t is the trend estimate and S_t is the seasonality factor. The m-step-ahead of forecast at time t is calculated as:

$$F_{t+m} = L_t + mb_t + S_{t+m-s} \tag{7}$$

where,

y_t: the actual values which include seasonality,
L_t: level component of the series,
b_t: the estimation of the trend component,
S_t: the estimation of seasonality component,

s: the length of seasonality,
α: the smoothing constant for level where $0 < \alpha < 1$,
β: the smoothing constant for the trend estimation where $0 < \beta < 1$,
γ: the smoothing constant for seasonality estimation where $0 < \gamma < 1$,
m: number of steps ahead to be forecast,
F_{t+m}: forecast for m-step-ahead.

2.2 Seasonal Auto Regressive Integrated Moving Average

ARIMA model are also capable of modelling a wide range of seasonal data. A seasonal ARIMA is formed by including additional seasonal terms in the Seasonal ARIMA (SARIMA) models $(p, d, q) \times (P,D,Q)_m$ can be formulated as follows:

$$\emptyset_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D y_t = \theta_q(B)\emptyset_Q(B^S)u_t \tag{8}$$

where,

B: non-seasonal backward operators,
B^S: seasonal backward operators,
$\emptyset_p$: non-seasonal AR component coefficients with order p,
Φ_P: seasonal AR component coefficients with order P,
θ_q: non-seasonal MA component coefficients with order q,
$\emptyset_Q$: seasonal MA component coefficients with order Q,
d: non-seasonal differencing order,
D: seasonal differencing order,
y_t: the time series,
u_t: the white noise residual.

2.3 Hybrid Approach

Hybrid model in this study will be the combination of Additive Holt-Winters and DES. Additive Holt-Winter model was selected to perform the hybrid model because it yields the lowest forecast error. The data will be fitted using residual of the Additive Holt-Winters method. Residuals can be calculated by measuring the difference between actual value and forecasted value. DES will be used to forecast the in-sample residuals of Additive Holt-Winters that were obtained previously. After that, the forecasted data of residual by DES and forecasted data of Additive Holt Winters will be combined which then formed a hybrid forecast. Hybrid method be formulated as below:

$$F_{t+k} = f_k + e_k \tag{9}$$

where,

f_k: out-sample forecast from Additive Holt-Winters,
e_k: out-sample residual forecast from DES,
F_{t+k}: hybrid forecast for k-step-ahead.

3 Results and Discussion

3.1 Additive Holt-Winters Exponential Smoothing

The value of the optimal parameters which aims to minimize the sum of square error for the smoothing constant, α =0.2018 (smoothing constant for level series), β = 0 (smoothing constant for trend), γ =0.0547 (smoothing constant for seasonality factor) were obtained automatically from Solver function in Microsoft Excel.

3.2 SARIMA

3.2.1 Model Identification

SARIMA model is suitable to fit with the stationary series. We can observe the stationarity of a model by considering the ACF graph and we can also conduct a formal stationary test which is the Augmented Dickey-Fuller (ADF) unit root test. If a series is non-stationary, differencing needs to be done before we can proceed to the next stage. From the ACF plotted in Figure 1, we can see that the lags die slowly, or the ACF shows a slowly decreasing pattern, thus we can say that the series is non-stationary. We can also see an obvious seasonal component from the ACF in Figure 1. Thus, SARIMA model needs to be used to forecast the time series data.

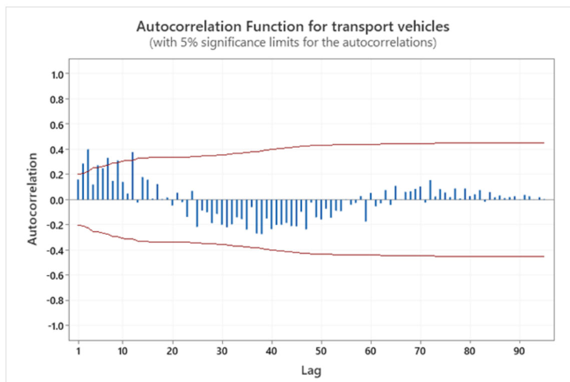

Figure 1: ACF Plot for Loan Data

Another formal way to check for the stationarity of a time series data is conducting ADF unit root test to evaluate the series. Figure 2 below that were obtained from RStudio shows that the data is non-stationary since the p-value = 0.4964 > α= 0.05, so we accept H_0 and conclude that series is non-stationary. Data differencing is needed until a stationary data is achieved.

```
> adf.test(TV)

        Augmented Dickey-Fuller Test

data:  TV
Dickey-Fuller = -2.1936, Lag order = 4, p-value = 0.4964
alternative hypothesis: stationary
```

Figure 2: ADF Test Results of original Loan Data

After regular differencing and seasonal differencing is done, the stationarity of the data was checked again using ADF test. We obtain the result as shown in Figure 3. Since p-value =0.01 < α=0.05, null hypothesis is rejected. Therefore, the model is stationary at level of significance, α =0.05.

```
> adf.test(seasonaldiffTV)

        Augmented Dickey-Fuller Test

data:  seasonaldiffTV
Dickey-Fuller = -7.027, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(seasonaldiffTV) : p-value smaller than printed p-value
```

Figure 3: ADF Test Results of differenced Loan Data

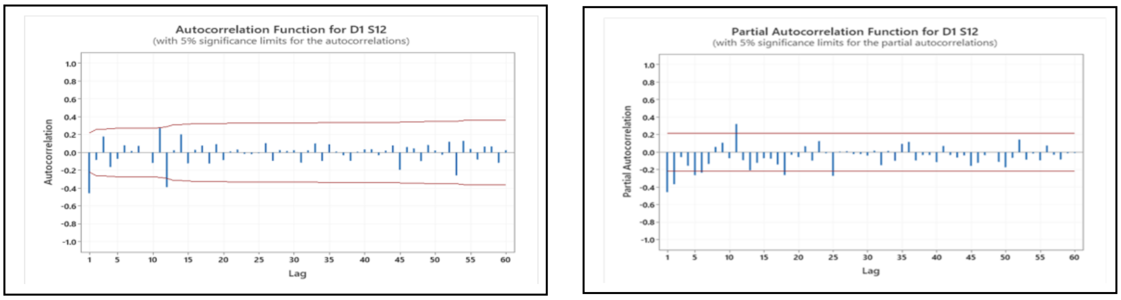

Figure 4: ACF and PACF plot of differenced loan data

Stationarity of a data can also be observed through visual inspection of ACF and PACF plot. From the ACF and PACF plot in Figure 4, we can also observe that most of the spikes in ACF and PACF plot are within the standard error band, thus we can assume that the series is stationary after seasonal differenced and regular differenced process were completed. The plots in Figure 4 further supports the ADF test results that were obtained previously.

3.2.2 Parameter Estimation

Now we can proceed to identify the number of orders in SARIMA $(p, d, q)(P, D, Q)_{12}$. Regular differencing and seasonal differencing were conducted one time each previously, thus we can fill in the *d* and *D* with 1. We can see that from PACF plot, there are no significant spikes greater than lag 2. For *P*, the spike is significant at lag 12 and lag 24, thus the possible *P* is 2. Observing the ACF plot, the most significant spikes can be seen at lag 1. To determine the value of Q, we are looking at the ACF plot and we can see that the spike is significant at lag 12 only, thus the Q is equal to 1. The best SARIMA model was chosen by looking at the model with the lowest Akaike Information Criterion (AIC). The possible SARIMA model can be seen in the Table 1.

Table 1: Possible SARIMA model

Model	MSE	AIC
SARIMA (0,1,1) (0,1,1) $_{12}$	**0.0093819**	**10.0554**
SARIMA (0,1,1) (2,1,0) $_{12}$	0.0104791	11.9594
SARIMA (0,1,0) (2,1,0) $_{12}$	**0.0180271**	**9.4881**

As there are many possible models that can be chosen from the estimated parameters before, we are going to look at AIC and MSE values to select the best SARIMA model. Model with lowest AIC and MSE value will be selected as the best forecasting model. According to the performance of the in-sample data from Table 1, SARIMA (0,1,0) $(2,1,0)_{12}$ has the lowest AIC value, while SARIMA (0,1,1) $(0,1,1)_{12}$ has the second lowest AIC value.

3.2.3 Model Selection and Diagnostic Checking

SARIMA $(0,1,0)(2,1,0)_{12}$ is the best model since it yields the lowest AIC value, while the SARIMA $(0,1,1)(0,1,1)_{12}$ has the second lowest AIC value. The adequacy of the SARIMA models will be measured by checking the Ljung-Box test. The adequate model with lowest AIC value will be considered as the best fitted model. After that, we are going to check for the parameter terms, whether they are statistically significant or not:

H_0: No association between parameter terms and the loans approved by purpose for purchase of transport vehicles.
H_1: There exist association between parameter terms and the loans approved by purpose for purchase of transport vehicles.

If the *p*-value is smaller than significance level at $\alpha = 0.05$, null hypothesis will be rejected, and a conclusion can be drawn where there exist association between the parameter terms and the loans approved by purpose for purchase of transport vehicles. All parameter terms in both models SARIMA $(0,1,0)(2,1,0)_{12}$ and SARIMA $(0,1,1)(0,1,1)_{12}$ are statistically significant because all of the *p*-values are smaller than $\alpha = 0.05$, and it should be kept.

After that, Ljung-Box test will be conducted to test whether there exists a serial of autocorrelation of residuals. Adequacy of the model can also be determined from the Ljung-Box test. If the *p*-value of Ljung-Box test statistics is larger than critical value at the significance level $\alpha = 0.05$, we can conclude the residuals of the model is not autocorrelated.

Table 2: Ljung-Box Test for SARIMA $(0,1,0)(2,1,0)_{12}$

Lag	**12**	**24**	**36**	**48**
Chi-Square	44.04	74.45	108.86	126.03
DF	9	21	33	45
***p*-value**	**0.000**	**0.000**	**0.000**	**0.000**

Although SARIMA (0,1,0) $(2,1,0)_{12}$ has the lowest AIC value, however the model is not an adequate model since the *p*-values are not greater than 0.05. Hence, we need to continue estimating another possible parameter, and we checked the Ljung Box test with the SARIMA (0,1,1) $(0,1,1)_{12}$ model as it yields the second lowest AIC value.

Table 3: Ljung-Box Test for SARIMA $(0,1,1)(0,1,1)_{12}$

Lag	**12**	**24**	**36**	**48**
Chi-Square	8.91	21.83	34.12	51.77
DF	9	21	33	45
***p*-value**	**0.446**	**0.409**	**0.414**	**0.227**

Now, we checked the Ljung Box test with the SARIMA (0,1,1) $(0,1,1)_{12}$ model and found that the model is an adequate model since all *p*-values are greater than 0.05, which further indicates that the model fitted the data very well and forecasting can be completed using this model. We can come to a decision that SARIMA $(0,1,1)(0,1,1)_{12}$ was chosen to forecast the loans for purchase of transport vehicles. After all, we can come to a conclusion that SARIMA $(0,1,1)(0,1,1)_{12}$ has satisfied all assumptions and this model is an adequate forecasting model.

3.3 Hybrid Methodology

Based on the forecast accuracy and the results that were obtained previously, we are going to fit the data using residual of the Additive Holt-Winters method. The residuals can be calculated by measuring the difference between actual data and the forecasted data of Additive Holt-Winter method. The time series plot of the residuals of Additive Holt-Winters method can be observed in Figure 5.

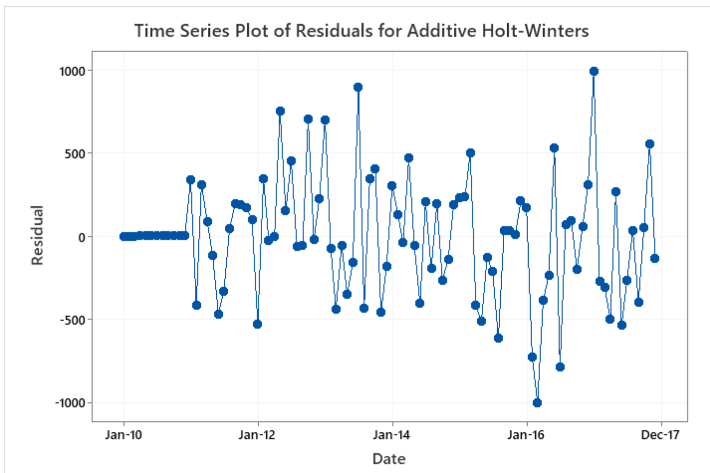

Figure 5: Time Series Plot of Residuals for Additive Holt-Winters

DES will be used to forecast the in-sample residuals of Additive Holt-Winters that were obtained previously. The out-sample data that was forecasted using DES and Additive Holt Winters will be combined to form a hybrid forecast. The hybrid model is a combination of Additive Holt Winters method and its residuals which were then forecasted using DES were produced and the forecast accuracy will be measured by comparing the hybrid model with the actual data. The value of smoothing constant α and β were set between 0 and 1. The optimal smoothing parameter which aims to minimize the SSE of α is 0 and β is 0.10.

3.4 Comparison between forecasting models

The forecasting accuracy and performance of each model will be tested by using three most common tools which are MAE, MAPE and RMSE. The out-sample loan data comprises of 12 months from January 2018 until December 2018 is utilized to measure the forecasting performance and forecasting accuracy of all three models. The measurement of forecast error can be seen in the following Table 4.

Table 4: Error measurement for forecasting model

Forecasting model	MAPE	RMSE	MAE

Additive Holt-Winters	**12.8619**	820.9751	577.7550
SARIMA $(0,1,1)(0,1,1)_{12}$	14.6299	1021.4307	683.8435
Hybrid	13.0181	**795.6100**	**574.8531**

Table 4 summarized the measurement error for each forecasting models. Additive Holt-Winters yields the lowest MAPE value as compared to the other two models. However, hybrid forecasting model managed to attain the lowest RMSE and MAE value as compared to Additive Holt-Winters and SARIMA. We can fairly conclude that hybrid model is the best model to forecast the monthly loans approved by purpose for banking system in Malaysia for purchase of transport vehicles data. Additive Holt-Winters is the second best forecasting model followed by SARIMA $(0,1,1)(0,1,1)_{12}$ and SARIMA yields the highest error when being tested with all three measurement error tools. The accuracy and performance of all three forecasting models can also be seen through line plot as shown in Figure 6.

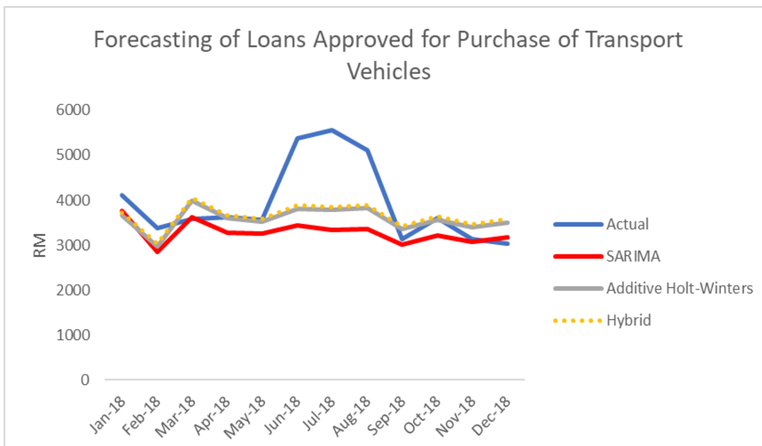

Figure 6: Comparison monthly loan data between actual values and forecasted values

4 Conclusion

A hybrid model which is the combination of DES and Additive Holt-Winters method were developed in this study. This study managed to measure the accuracy of three forecasting models which are Additive Holt-Winters, SARIMA, and hybrid model. Hybrid methodology is a better forecasting model as compared to SARIMA and Additive Holt-Winters model since it yields the lowest RMSE and MAE value, and the second lowest MAPE value. Additive Holt-Winters was deemed to be the second-best forecasting model as it generates the lowest MAPE value, but the second lowest for RMSE and MAE value after hybrid model. In this study, SARIMA was found to be the least accurate forecasting model since it generates the highest error for all three-measurement accuracy. To sum up, hybrid model can be considered as a good forecasting model. We also examine the pattern of the forecasted data using the best forecasting model which is hybrid methodology and found that the loan data decrease gradually for the first few months in 2018, and it rose steadily after.

Hybrid model was developed mainly to cater the problem of a univariate forecasting model and to improve the forecasting accuracy. However, selecting the right model is very critical or else the hybrid model might generate worser forecast than the single model itself. In future, this study can be improved by using a more advanced and robust model, then, combining them with a conventional model, such as Neural Network with ES, or ARIMA. Researchers could also combine two advanced and robust model which will lead to a better forecasting model that can capture the pattern and trend of all kinds of data so that a better prediction can be attained.

References

[1] Diebold, F. X., & Li, C. (2006) ‘Forecasting the term structure of government bond yields’, *Journal of econometrics.* 130(2): 337-364.

[2] Sun, S., Wei, Y., Tsui, K.-L., & Wang, S. (2019) ‘Forecasting tourist arrivals with machine learning and internet search index’, *Tourism Management.* 70: 1-10.

[3] Choden, S., & Unhapipat, S. (2019) ‘Statistical model for personal loan prediction in bhutan’, *Journal of Advanced Research in Dynamical and Control Systems.* 11(9): 416-422.

[4] Sheikh, M. A., Goel, A. K., & Kumar, T. (2020). *An Approach for Prediction of Loan Approval using Machine Learning Algorithm.* Paper presented at the 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC).

[5] Ostertagova, E., & Ostertag, O. (2012) ‘Forecasting using simple exponential smoothing method’, *Acta Electrotechnica et Informatica.* 12(3): 62.

[6] Yorucu, V. (2003) ‘The analysis of forecasting performance by using time series data for two Mediterranean islands’, *Review of Social, Economic & Business Studies.* 2: 175-196.

[7] Adenomon, M., Obazee, O., & Mayaki, J. (2019) ‘COMPARATIVE STUDY OF UNIVARIATE TIME SERIES MODELS ON FORECASTING MONTHLY EXCHANGE RATE IN NIGERIA’.

[8] Jeong, D.-B. (2017) ‘Forecasting for a Credit Loan from Households in South Korea’, *The International Journal of Industrial Distribution & Business.* 8(4): 15-21.

[9] Kumar, M., & Anand, M. (2014) ‘An application of time series ARIMA forecasting model for predicting sugarcane production in India’, *Studies in Business and Economics.* 9(1): 81-94.

[10] Mondal, P., Shit, L., & Goswami, S. (2014) ‘Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices’, *International Journal of Computer Science, Engineering and Applications.* 4(2): 13.

[11] Mansor, R., Zaini, B. J., & May May, C. C. (2021) ‘Medium-term electricity load demand forecasting using holt-winter exponential smoothing and SARIMA in university campus’, *SSRG International Journal of Engineering Trends and Technology.* 69(3): 165-171.

[12] Kotillová, A. (2011) ‘Very short-term load forecasting using exponential smoothing and ARIMA models’.

[13] Merh, N., Saxena, V. P., & Pardasani, K. R. (2010) ‘A comparison between hybrid approaches of ANN and ARIMA for Indian stock trend forecasting’, *Business Intelligence Journal.* 3(2): 23-43.

[14] Kumar, M., & Thenmozhi, M. (2014) ‘Forecasting stock index returns using ARIMA-SVM, ARIMA-ANN, and ARIMA-random forest hybrid models’, *International Journal of Banking, Accounting and Finance.* 5(3): 284-308.