A Comparison of Classical Time Series Models and Machine Learning LSTM Model to Forecast Paddy Production in Sri Lanka

IBI Sandaruwani[1#] and RAB Abeygunawardana[1]

[1] Department of Statistics, Faculty of Science, University of Colombo, Sri Lanka
#imesha4sandaruwani@gmail.com

Abstract: *The most common & effective traditional methods of univariate time series forecasting are Autoregressive Integrated Moving Average (ARIMA) family models and Exponential Smoothing family models. With the recent advancement in more advanced machine learning algorithms and approaches such as Long-Short-Term-Memory modeling approaches, new algorithms are developed to analyze and forecast time series data. The objective of the study is to identify the best time series forecasting model among classical time series models and machine learning LSTM model to forecast the annual paddy production of Sri Lanka. Based on the RMSE, MAE and MAPE values, the results showed that the estimated error of ARIMA & Double Exponential Smoothing (DES) models are much higher than the estimated error of chosen LSTM model. Hence LSTM model outperforms to the traditional-based algorithms like ARIMA and smoothing models for forecasting the paddy production of Sri Lanka. The forecasts for paddy production from 2022 to 2024 were 4.92, 4.89 and 5.34 million Mt respectively. This model can be used by researchers for forecasting paddy production in Sri Lanka and it should be updated continuously with incorporation of recent data.*

Keywords: *Time series, Forecasting, Paddy Production, Sri Lanka, ARIMA, Double Exponential Smoothing, LSTM*

1. Introduction

Rice is the dietary staple and the major domestic crop cultivated in Sri Lanka since ancient times. The livelihood of more than 1.8 million Sri Lankan farmers is paddy production. Specially with the current economic crisis in Sri Lanka, prices of imported basic food products is high and the expenditure on rice sector has increased continuously mainly due to inorganic fertilizers, agro-chemicals and fuel shortage. Therefore, having a very accurate forecast on the production of main food in the country for importing requirements is much significant at this time to ensure food security.

The forecasts of paddy production in Sri Lanka for the future years will no doubt to be useful for policy makers, country planners and research workers. This study will be beneficial to the government and other people concerned for a reason that they can generate analysis from the brief forecast which will help them in decision making and planning in the future as reliable management information on production estimates is a prerequisite for policy decision making on rice imports to Sri Lanka.

Forecasting by Machine Learning models attracted much attention in recent times compared to traditional statistical medelling. Hence in this study, a comparison of forecasting accuracy between classical ARIMA & DES models with machine learning LSTM modeling was done to forecast the paddy production of Sri Lanka. For this purpose, univariate time series data of annual paddy production from 1952 to 2021 was used. This research can be useful for future researchers who will do analysis relating to the same topic because this will be material for their reference through adding knowledge on how the time series forecasting approach works and help for further understanding.

2. Methodology

The annual paddy production data of Sri Lanka from 1952 to 2021 was used for this study from the Department of Census and Statistics, Sri Lanka. The First 95% of the data is used to estimate the models and the remaining 5% data is used for model validation.

The classical time series models that have been considered in the present study are ARIMA and DES models. The forecasting performance of these traditional time series models are compared with LSTM model forecasting performance.

A. ARIMA model

In ARIMA model fitting, Box Jenkins methodology is used. After checking and making data to stationary by transformations and differncing the ARIMA model fitting was done. Logarithmic transformation is used in this study because they are interpretable and constrain the forecasts to stay positive on the original scale.

A series is called ARIMA (p,d,q) if we need to difference the series d times to make it stationary before applying ARMA (p,q) model. Here, p denotes the number of autoregressive terms and q denotes the number of moving average terms.

B. Double Exponential Smoothing model

DES is used when data have a trend and do not have a seasonal component to deliver short-term forecasts. This procedure calculates dynamic estimates for two components: level and trend. In addition to the level (α) parameter for controlling smoothing factor for the level, an additional smoothing factor is added to control the decay of the influence of the change in trend called trend parameter (Υ) in this model.

The initialization method used to determine how the smoothed values are obtained in one of two ways: with optimal weights or with specified weights. In optimal weights method, ARIMA (0,2,2) model is fitted to data in order to minimize the sum of squared errors and the trend and level components are then initialized by back casting. In specified weights method a linear regression model to time series data (y variable) versus time (x variable) is fitted. The constant from this regression is the initial estimate of the level component, the slope coefficient is the initial estimate of the trend component.

C. LSTM model

LSTMs are a special kind of recurrent neural network. It is capable of learning long-term dependencies by having memory cells and gates that controls the information flow along with the memory cells. The LSTM generates the cell states (c_t) and hidden states (h_t) for the consumption of the next time step LSTM. There are two methods in LSTM to generate these forecasts. That is an unrolling forecast scenario and a rolling forecast scenario. In the rolling forecast scenario, model will be used to make a forecast for the time step, then the actual expected value from the test set will be taken and made available to the model for the forecast on the next time step. Because this methodology involves moving along the time series one-time step at a time, it is called Walk

Forward Validation. This method is the standard method of model evaluation in LSTM. The default activation function for LSTMs is the hyperbolic tangent (tanh), which outputs values between -1 and 1. In compiling the network, we must specify a loss function and optimization algorithm. Mean squared error is better as the loss function and the efficient Adam

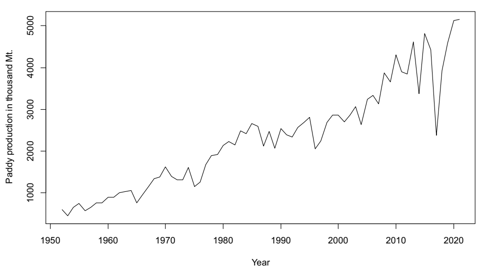

Figure 1. Time series plot of annual paddy production of Sr Lanka.

optimization algorithm is better as the optimization algorithm for small data sets. By default, the samples within an epoch are shuffled prior to being exposed to the network. This is undesirable for the LSTM because it needs the network to build up state as it learns across the sequence of observations. Hence the shuffling of samples can be disabled. The internal state at the end of the training epoch can be reset & ready for the next training iteration. To calculate the weight of the network the gradient descent method is utilized and adjust the weight of interconnection to minimize the sum-squared error (SSE) of the network.

D. Model selection criteria

In statistical modelling one of the main challenges is to select a suitable model to characterize the underlying data. The two most commonly used criteria in model selection of time series are the Akaike information criterion and Bayesian information criterion.

E. Forecast Performance Measures

There are several ways to evaluate the performance of forecasting models. In this study, Mean Absolute Error(MAE), Root Mean Squared Error(RMSE) and Mean Absolute Percentage error(MAPE) is considered as the metrics to evaluate the time series models.

3. Results

As the first step of the time series analysis, the time series plot was drawn by using the original annual paddy production data of Sri Lanka. The figure shows the time series plot of annual paddy production.

To verify the stationarity of annual paddy production series unit root tests are applied to the original annual paddy production data of Sri Lanka. As ADF test suggested that the data is not stationary, logarithmic transformed series was considered and checked for stationarity. All ADF, KPSS and PP test suggested that first order differenced logarithmic transformed series is stationary.

In order to identify the tentative ARIMA model for the data set, after identifying the order of difference it is needed to plot the Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) to get the value of p and q. The ACF and PACF plots have come up with the possible model as ARIMA(2,1,1). However, all possible models up to lag 5 are fitted and among which ARIMA (2, 1, 1) has been selected as the best model with the lowest AIC and BIC values to forecast the paddy production of Sri Lanka. Table 1 contains the estimated values of selected ARIMA (2, 1, 1) model.

Table 1. Estimated values of ARIMA (2,1,1)

	AR1	AR2	MA1
Coefficients:	-0.8778	- 0.4581	0.4990
S.E.	0.2749	0.1349	0.2726
sigma^2 = 0.02755		log likelihood = 25.83	
AIC= -43.66	AICc= -42.99	BIC= -34.96	

To observe the forecasting behavior of fitted model ARIMA (2,1,1), the forecasted value obtained by the model is evaluated. The figure 1 shows the forecasted values alog with actual values.

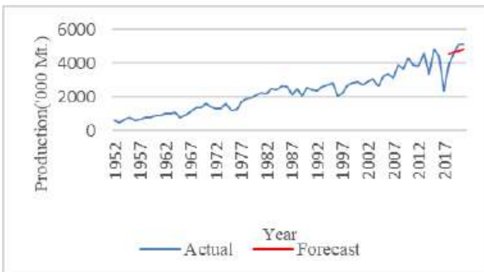

Figure 2. Actual value along with the Forecasted value using ARIMA (2,1,1).

The MAPE obtained was 31.32 which is bigger than 10 indicates good forecasting but performs a little less enough than expected.

Since the original paddy data set has a clear upward trend with no seasonality, a Double Exponential Smoothing model is fitted as next step. DES model employs a level component and a trend component at each period. To find the optimal values of the parameter of level(α) and trend(γ), various combinations of level and trend based on range between 0.1 to 0.9 with increments of 0.1 were tried. Following table shows the mean absolute deviation (MAD) & mean squared deviation (MSD) values of some fitted models which had comparatively minimum values.

Table 2. MAD & MSD values of fitted Double Exponential Smoothing models

Level smoothing constant (α)	Trend smoothing constant (γ)	MAD	MSD
0.1	**0.4**	**305**	**207859**
0.3	0.9	311	229887
0.6	0.9	316	234269
0.7	0.7	319	237528
0.8	0.5	317	236059

The model that have the lowest MAD and MSD value tend to give slightly better results than the other models. Hence, here optimal values of the parameter of α and γ is found as 0.1 & 0.4 respectively. Next, annual paddy production of Sri Lanka is forecasted for testing sample time period using selected DES model.

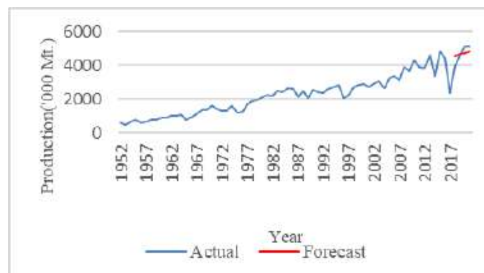

Figure 3. Actual value along with the Forecasted value using DES model.

Forecasting performance of fitted model is checked through the MAPE. Since the MAPE is only 7.8% which is less than 10%, the fitted model performs extremely well compared to best fitted ARIMA model. Therefore, it can be used to forecast annual paddy production of Sri Lanka.

A machine learning LSTM model is developed to forecast annual paddy production of Sri Lanka. In here also, first 95% of data was used for machine learning model training purpose and the rest 5% of observations are used for testing the model performance.

A non-stationary series will introduce more error in predictions and force errors to compound faster. Stationary data is easier to model and will very likely result in more skillful forecasts. While stationarity is not an explicit assumption of LSTM, it does help immensely in controlling error. Hence before building the model, the series is checked for stationarity and found that the first order differenced yearly paddy production data is stationary.

To train the LSTM neural network model the data series was framed as a supervised learning frame. The input component of neural network model was some number of prior observations. The number of lag observations to use in the input component was also selected by which provides minimum RMSE. Based on that number batch sizes were selected.

As model is being trained using batch gradient descent, the selected batch sizes were equal and more than the supervised training sample size in the trial-and-error method. The model also has a single hidden layer with some number of nodes. The rectified linear activation function was used on the hidden layer as it performs well. The number of nodes to use in the hidden layer was also selected by trial-and-error method from range 1 to 5 based on minimum RMSE. The output component was the paddy production in next year because the model is developed to make next step forecasts.

A linear activation function was used on the output layer as a continuous value is predicted.

To evaluate the fitted LSTM neural network model, walk forward validation was used. Below table shows the RMSE value of some best fitted models.

Table 3 RMSE of best fitted LSTM models

Input component	Num. of nodes	Num. of epochs	Batch size	Difference order	RMSE
3	1	2	63	1	149.7
4	2	3	62	1	164.9
5	3	2		1	111.4
8	4	3	5861	1	189.5
10	5	2	56	1	344.1

According to table 3, the LSTM model with 5 input component, 3 nodes, 61 batch size and 2 epochs has the lowest RMSE value. As model is being trained using batch gradient descent method, the selected batch sizes were equal and more of in supervised learning sample size. As data set is framed as a supervised learning problem, in the optimal model there are 61 samples that could be used to train the model. Hence this LSTM network was selected as the best LSTM model that yields best results for forecasting annual paddy production of Sri Lanka. As next step, annual paddy production of Sri Lanka is forecasted for testing time period using above fitted LSTM model.

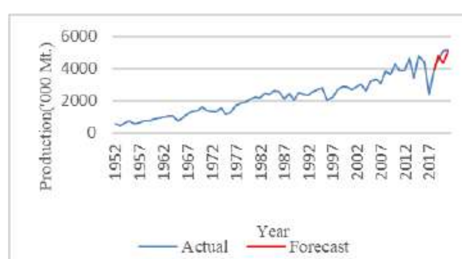

Figure 4. Actual value along with the Forecasted value using LSTM model.

As the MAPE of this model is only 2.61%, the fitted model performs very well. Therefore, it can be used to forecast the annual paddy production of Sri Lanka.

As the next step, the forecasting performance of all three models are compared using forecasting intensity measurements. All approaches ARIMA, DES and LSTM are compared regarding their forecasting behavior using three accuracy measures such as RMSE, MAE and MAPE.

Table 4. Comparing forecasting performance of ARIMA, DES and LSTM models

	ARIMA (2,1,1)	DES	LSTM
RMSE	1647.93	408.03	395.45
MAE	1522.99	352.75	138.83
MAPE	0.3132	0.0785	0.02618

When comparing all three models forecasting measurements, all RMSE, MAE & MAPE is the lowest for LSTM model. Hence based on the MAPE, RMSE and MAE, it is clear that the forecasts based on LSTM machine learning model is more accurate than ARIMA and DES models. Hence the fitted LSTM model can be selected as the best model to forecast annual paddy production of Sri Lanka as it performs superior compared to other two models.

The next three years yearly paddy production was forecasted using the fitted LSTM model. The output is shown in below table.

Table 5. Three years ahead forecasted values of annual paddy production of Sri Lanka using fitted LSTM model

Year	Forecasted value (000 Mt.)
2022	4,918.983
2023	4,886.369
2024	5,341.231

According to the LSTM model forecasts for next three years, in year 2022 Sri Lanka is expected to have 4, 919 metric tons of paddy production. In years 2023 & 2024 Sri Lanka is expected to have 4, 886 and 5, 341 metric tons of paddy production respectively.

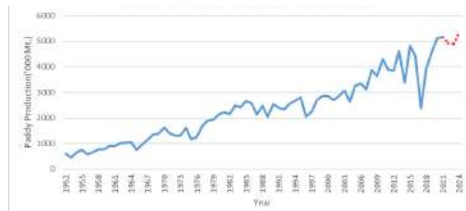

Figure 5. Three years ahead forecasted values of annual paddy production of Sri Lanka using selected LSTM model

4. Discussion and Conclusion

The basic motives of this study was to identify the best time series forecasting model to forecast the annual paddy production of Sri Lanka for the next coming three years. In order to model the annual paddy production of Sri Lanka, three approaches were used. Namely, ARIMA modelling approach, Smoothing modelling approach and Machine Learning modelling approach.

Before estimating and developing a stationary model, the stationarity of data series was checked. The original annual paddy production series was nonstationary, and variance was not constant with time. Therefore, the original annual paddy production series were converted to logarithmic transformed series and fitted a stationary model. since the transformed series was not stationary, first lag differenced transformed series was considered. As there was no seasonality indicated by the plot, in search for the suitable ARIMA model to modeling and forecasting the paddy production of Sri Lanka, ACF and PACF plots recommended possible ARIMA models, among them ARIMA (2, 1, 1) has been selected as the best ARIMA model according to the minimum AIC and BIC values to forecast annual paddy production of Sri Lanka.

A DES model was also fitted to the original annual paddy production series. The best fitting parameter values were found as 0.1 & 0.4 respectively for level and trend parameters with minimum MAD and MSD values.

Several parameter values were checked for input component, nodes, batch size and epochs and found optimal values as 5 input components, 3 nodes, 61 batch size and 2 epochs which has the lowest RMSE value. Hence this LSTM network was selected as the best fitting LSTM model that yields best results for forecasting annual paddy production of Sri Lanka.

To compare the forecasting behavior of classical statistical models and machine learning model for forecasting paddy production of Sri Lanka, three forecasting evaluation techniques namely RMSE, MAE and MAPE of test data set are obtained and compared. The results based on these three measures of error have showed that the performance of LSTM model with single hidden layer, three neurons and two epochs is better than ARIMA (2,1,1) model and DES model with 0.1 level and 0.4 trend parameters for forecasting annual paddy production of Sri Lanka. Finally, with the help of chosen LSTM model, the future annual production of paddy in Sri Lanka for next recent three years was estimated and presented the results in table 5.

In conclusion, it can be said that the LSTM model is better than the ARIMA (2,1,1) model and DES model to forecast the annual paddy production in Sri Lanka. The other finding from the forecasting analysis is that the paddy production will be decreased in the year 2022 & 2023 and will be increased in 2024. This model can be used to predict the future values of annual production of paddy in Sri Lanka. As per this study, LSTM model has been recommended to forecast the time series of annual data. Because in this study, LSTM model gives better result comparing with the classical statistical time series models for forecasting the paddy production of Sri Lanka.

These forecasting values can help mostly the policy makers to plan and make decisions on rice imports and to encourage paddy farmers on higher production to make country self-sufficient in rice. As well as this brings awareness of paddy production, indirectly the rice production towards positive side.

Reference

Karunanayaka , K., & Aponsu, G. (2019). Forecasting Paddy Production in Yala And Maha Seasons in Sri Lanka. *11th Symposium on Applied Science, Business & Industrial Research*, 293-299.

Namini, S. S., Tavakoli, T., & Namin, A. S. (2018). A Comparison of ARIMA and LSTM in Forecasting Time Series. *International Conference on Machine Learning and Applications*, 5386-6805.

Shabri, A., Samusudin, R., & Ismail, Z. (2009). Forcasting of the Rice Yields Time Series Forcasting using Artificial Neural Network and Statistical Model. *Journal of Applied Sciences 9*, 4168-4173.

Sivapathasundaram, V., & Bogahawatte, C. (2012). Forecasting of Paddy Production in Sri Lanka: A Time Series Analysis using ARIMA Model. *Tropical Agricultural Research Vol. 24 (1)*, 21-30.

Sultana, A., & Khanam, M. (2020). Forecasting Rice Production of Bangladesh Using ARIMA and Artificial Neural Network Models. *Dhaka University Journal of Science*, 143-147.

Vijay, N., & Mishra, G. (2018). Time Series Forecasting Using ARIMA and ANN Models for Production of Pearl Millet (BAJRA) Crop of Karnataka, India. *International Journal of Current Microbiology and Applied Sciences*, 880-889.