



Research Article

Development of Linear, Nonlinear and Hybrid Models for Forecasting Sugarcane Yield

Qaisar Mehmood^{1*}, Ali Raza², Asif Ali Abro³, Nargis Shaheen⁴ and Muhammad Riaz⁵

¹Department of Statistics, Government Graduate College Bahawalnagar, Pakistan; ²Department of Agriculture, Crop Reporting Service, Bahawalpur, Pakistan; ³Head of Special Studies & Performance Audit, O/o Director General Audit, Local Government, Sindh, Pakistan; ⁴Department of Statistics, Government Graduate College for Women, Bahawalnagar, Pakistan; ⁵Department of Statistics, Rahim Yar Khan Campus, Islamia University of Bahawalpur, Bahawalpur, Pakistan.

Abstract | Sugarcane is important cash crop massive contributing the agricultural economy of the Pakistan, it is necessary for future to forecast the yield of sugarcane crop. The purpose of the study has to propose the optimum forecast models of the time series, artificial neural network and their hybrid models for forecasting the yield of sugarcane. Yearly data for the yield of sugarcane crop from 1947 to 2020 for economic survey of Pakistan was used for forecasting. We compare ARIMA, ETS, TBATS, Artificial Neural Network (ANN), ARIMA-ETS, ARIMA-TBATS, and ARIMA-ANN hybrid models by calculating RMSE and MAE for each model. It was observed that the ARIMA (2, 1, 0) model was optimum because it shows the minimum values for RMSE (2345.059) and MAE (1879.447) for sugarcane yield. Forecast average yield of sugarcane crop will be increase after ten years from 63827kg to 65660.37kg per hectare from 2020 to 2030. This increase amount of yield may increase the amount of sugar to meet the country requirements. More over these forecast estimates for sugarcane yield will be important for the Government in formulating their policies to fulfill the food necessities of the nation, trade, support prices, and planning about the cultivation sector.

Received | January 09, 2024; **Accepted** | May 15, 2024; **Published** | July 10, 2024

***Correspondence** | Qaisar Mehmood, Department of Statistics, Government Graduate College Bahawalnagar, Department of Economics and Statistics, University of Management and Technology Lahore, Pakistan; **Email:** qaisarm11@gmail.com

Citation | Mehmood, Q., A. Raza, A.A. Abro, N. Shaheen and M. Riaz. 2024. Development of linear, nonlinear and hybrid models for forecasting sugarcane yield. *Sarhad Journal of Agriculture*, 40(3): 754-759.

DOI | <https://dx.doi.org/10.17582/journal.sja/2024/40.3.754.759>

Keywords | Sugarcane yield, Forecasting, ARIMA, ETS, TBATS and ANN



Copyright: 2024 by the authors. Licensee ResearchersLinks Ltd, England, UK.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Introduction

Sugarcane is an important major crop in Pakistan, production of sugarcane crop is increased during the last thirty year. After Brazil, India and China, Pakistan has fourth position between the sugarcane producing countries (FAO, 2010). Pakistan has 25.83 kg per capita consumption per year, as biggest

consumer of sugar in South Asia and Pakistan meets the 99% domestic and export level requirements of sugar from Sugarcane crop (Azam and Khan, 2010). Pakistan is also 9th sugar exporting country in the world (USDA, 2019). Sugarcane is a major crop which has an important for the agricultural economy of the Pakistan, it is much necessary to forecast the yield of sugarcane crop for the future. The objective of this

study was to suggest the best forecast models among the time series, artificial neural network and their hybrid models for forecasting the yield of sugarcane crop in Pakistan. For forecasting the sugarcane crop different models have been applied; however, in time series models Box-Jenkins's, (1976) ARIMA model was widely used.

In time series modeling and forecasting the production of sugarcane crops in Pakistan by using an appropriate measure as an ARIMA model was helpful and appropriate for policy making (Muhammad *et al.*, 1992). According to Allen (1994), forecasting agricultural production and prices was proposed to be useful for the farmers, governments, and agribusiness industries. While, Masood and Javed (2004) used a linear regression model to forecast the sugarcane area of Punjab, KPK, Sindh, and Pakistan, for overall Pakistan yield forecasting they run a regression model and estimate the sugarcane production by multiplying the forecast value of yield and area. They conclude that yield and area from the forecast model give the efficient prediction for future yield and area of sugarcane. Other researchers Yaseen *et al.* (2005) and Sajid *et al.* (2015) used the ARIMA model by checking the suitable diagnostic tests for forecasting the yield of sugarcane and cotton crops in Pakistan and predict that the forecasting values were very close to the actual values. In conducting a study of preharvest of sugarcane yield forecasting by Krishna and Suresh (2010) using climatic variables in India, they were developing a forecast model by using weather variables as a predictor and concluded that the forecast model was able to explain 87% variation in the sugarcane yield before two months harvest. We also have made a pilot study for forecasting the production of sugarcane crops taking FBS data from 1947 to 2017 using the ARIMA model, forecast results are very close to the actual values Mehmood *et al.* (2019). Nabeel (2022) forecasts the sugarcane production by using ARIMA model for Punjab, KPK, Baluchistan and Sindh taking the data from 1982 to 2016 from Pakistan Bureau of Statistics. Supriya *et al.* (2023) apply the Holt linear trend, ARIMA and ARIMAX models for forecasting the sugarcane production up to year 2030 by taking data from 1961-2020 from India, Pakistan, Bangladesh, Nepal, Sri Lanka and China. Artificial Neural Network (ANN) is a widely accepted tool as a prediction method which is used in harvest predictions under climate change Gopal and Bhargavi (2019a, b, c), Adisa *et*

al. (2019), Perera and Rathnayake (2019). Jiang *et al.* (2004) construct the artificial neural network model for forecasting the crop yield developed an artificial neural network model for estimating crop yields using remotely sensed information. Sreekanth *et al.* (2009) forecast the groundwater in Hyderabad India using the artificial neural network model; forecast values are very close to observe values. Aryal and Yao-wu (2003) develop the ANN model for forecasting the production level of the chinees construction industry which has a significant role in the GDP of China, constructed artificial neural network model has RMSE 49 percent lower than the ARIMA model. According to them, ANN has significant potential to capture the nonlinear relationship, as ANN gives the best forecast for the Chines construction industry than the ARIMA model. Laxmi and Kumar (2011) forecast the yield of rice, wheat, and Sugarcane crop using an artificial neural network-based forecast model by taking crops yield as the output variable and temperature, rainfall, and morning humidity as input variables. Kumar *et al.* (2015) developed ANN forecast models for forecasting the sugarcane yield by taking time series data from 1950 to 2011 of sugarcane yield in India. Paddy yield in Sri Lanka was predicted by an artificial neural network model by considering climate factors Vinushi *et al.* (2020). Bingjun *et al.* (2021) used a grey back propagation neural network forecasting model to forecast the grain yield for the Henan Province, China.

Omer *et al.* (2017), applied the SARIMA-ANN model to forecast the electricity load taking Turkish data. Ozozen *et al.* (2010) also use the SARIMA-ANN to forecast electricity price in Turkey. Riaz *et al.* (2023) compare the different models and found SARIMA-ANN model was the best model for forecasting the malaria cases in district Rahim Yar Khan, Pakistan. Mehmood *et al.* (2023) for forecasting the area and production of wheat crop in Pakistan, seven different models ARIMA, ETS, TBATS, ANN, ARIMA-ETS, ARIMA-TBATS and ARIMA-ANN models were applied and ARIMA-ANN model was found the best forecasting model for both area and production of wheat crop.

Materials and Methods

We chose the yearly data from 1947 to 2020 for the sugarcane yield (kg) per hectare for Pakistan. The data was taken from the official website of the Ministry of

Finance, Government of Pakistan ([Pakistan Economy Survey, 1947-2020](https://www.finance.gov.pk/)). We propose the best forecasting model for forecasting the yield of sugarcane crop applying time series ARIMA, ETS, TBATS models, artificial neural network model and ARIMA-ETS, ARIMA-TBATS, ARIMA-ANN hybrid models. Here we discuss below briefly all these models.

Autoregressive integrated moving average (ARIMA) model

Box and Jenkin (1976) have developed the ARIMA process time series forecasting methodology known as Box-Jenkin’s methodology. It is also known as the ARIMA model based on fitting mixed auto regressive, integrated, and moving average models on a set of time series data.

The ARIMA (p, d, q) model is

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + u_t + \theta_1 \mu_{t-1} + \theta_2 \mu_{t-2} + \dots + \theta_q \mu_{t-q}$$

Where; Y_t is the yield of sugarcane at time t . $Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots, Y_{t-p}$ are the lag values at time $t-1, t-2, \dots, t-p$, respectively. $u_t, u_{t-1}, \dots, u_{t-q}$ is the error term and its lag values, while the ϕ_1, \dots, ϕ_p are the coefficients of the autoregressive model and $\theta_1, \dots, \theta_q$ are the coefficients of the moving average model.

ETS model

ETS is abrevated for the error trend and seasonality, or exponential smoothing (ETS) model. Exponential smoothing is a time series forecasting methodology for data that can be applied to data consisting of both systematic trends and a seasonal component. ETS is a forecasting method which is an alternate to ARIMA model. This is simple method that handles both trend and seasonality.

TBATS model

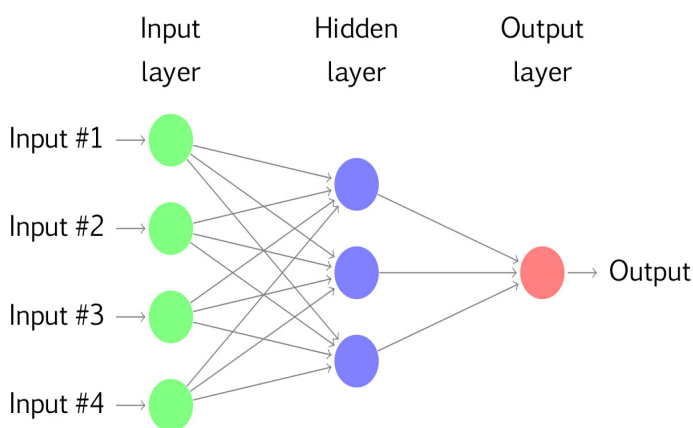
TBATS model has the ability to counter the complex seasonalities (e.g., non-integer seasonality, non-nested seasonality and large-period seasonality) with no seasonality constraints, making it possible to create detailed, long-term forecasts. The TBATS stands for Trigonometric seasonality, Box-Cox transformation, ARMA errors, Trend, and Seasonal components. The seasonal component shows the periodical variation due to in the series.

Artificial neural network (ANN) model

The artificial neural network (ANN) is intended to

work like the biological nervous systems such as the brain process the visual data. It performs like the human brain, trying to recognize the regularities and shapes of the data Rosenblatt (1958). ANN is developed to measure the non-linearity in data these tools have been recently implemented in econometrics for forecasting macroeconomic variables. The invention of ANN introduces the hidden layers among the input layers and the output layers. The relationships between independent variables and dependent variables are easily captured in hidden layers. ANN also creates the activation function (logistic function) which can measure the nonlinear function.

An example of ANN architecture is shown in the figure as.



The general expression of neural network is :

$$y_{t+1} = [\sum \alpha_j q_j = 0 (\sum \beta_{ij} y_{t-i} p_i = 0)] \dots (1)$$

Where, f and g are the activation function with “ p ” number of input nodes, q the number of hidden nodes, β_{ij} is the weight assign to the i th input node to the j th hidden node, α_j is the weight assign to the j th hidden node to the output node and y_{t-i} is the i th input (lag) of the model.

Development of hybrid time series models

In this section we develop the hybrid time series forecasting model by combining the ARIMA model with the ETS, TBATS and ANN model. These are ARIMA-ETS model, ARIMA-TBATS model and ARIMA- ANN model. In these days development of hybrid models by mixing the linear and nonlinear models has become popular. These models are suitable when the time series data have linear and nonlinear trend. Which capture both linear and nonlinear

pattern in time series.
Let time series is

$$\hat{u}_t = +$$

Where; y_t is time series values, L_t is the linear and N_t is nonlinear component. From this methodology we construct hybrid models like ARIMA-ETS, ARIMA-ANN, ARIMA-TBATS.

Results and Discussion

In this section we compare the results of ARIMA, ETS, ANN, TBATS models and hybrid time series models ARIMA-ETS, ARIMA-ANN and ARIMA-TBATS models were developed for sugarcane yield to propose a suitable model. We compare the root mean square error and mean absolute error for the seven different models for the time series data of sugarcane yield to propose the appropriate model for forecasting the sugarcane yield.

ARIMA, ETS, ANN, TBATS, hybrid ARIMA-ETS, hybrid ARIMA-ANN, and hybrid ARIMA-TBATS models were performed on the sugarcane yield data to find the optimum forecasting model.

From their diagnostic testing ARIMA model was found to be the best forecast model among the remaining forecasting models, as from the Table 1 ARIMA model has minimum root mean square error (2345.059) and mean absolute error (1879.447) as compared to other competitive models.

Table 1: Detailed summary of sugarcane yield fitted models along with accuracy.

Fitted models	RMSE	MAE
ARIMA	2345.059	1879.447
ETS	2469.525	2001.466
ANN	2583.648	2049.204
TBATS	2555.966	2066.961
ARIMA-ETS	2364.697	1917.714
ARIMA-ANN	2414.37	1925.777
ARIMA-TBATS	2393.916	1939.731

Time plot of observed sugarcane yield and fitted values of various time series models were displayed in Figure 1, while in Figure 2 the ten-year forecasted yield along with 80% and 95% confidence bands

under the best selected ARIMA model.

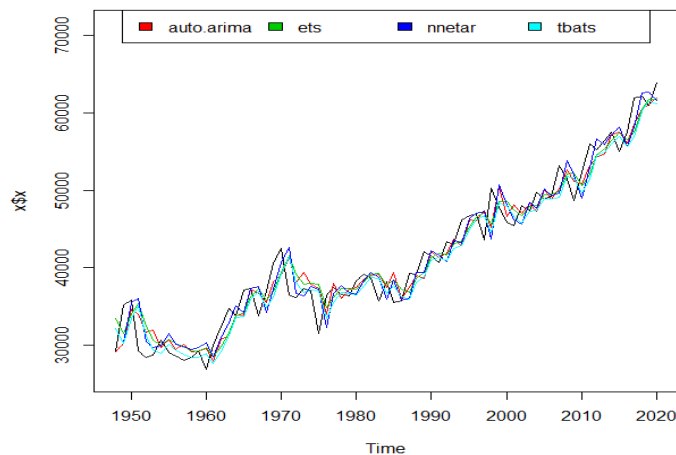


Figure 1: Actual vs fitted sugarcane yield.

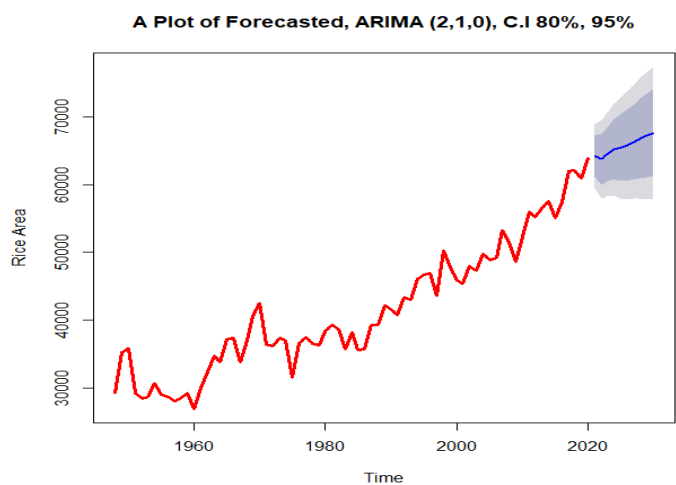


Figure 2: ARIMA forecast plot.

Table 2: Detail summary of forecasted sugarcane yield under the selected ARIMA model.

Time	Point forecast	Lo 80	Hi 80	Lo 95	Hi 95
2021	64193.35	61102.16	67284.55	59465.78	68920.93
2022	63711.72	59908.66	67514.78	57895.44	69528.00
2023	64443.01	60435.7	68450.33	58314.35	70571.68
2024	65159.95	60736.52	69583.39	58394.9	71925.01
2025	65409.93	60526.11	70293.74	57940.77	72879.08
2026	65797.78	60600.62	70994.95	57849.4	73746.16
2027	66327.92	60833.17	71822.66	57924.42	74731.41
2028	66764.19	60951.66	72576.72	57874.69	75653.69
2029	67171.81	61068.5	73275.11	57837.61	76506
2030	67623.99	61252.91	73995.07	57880.26	77367.72
Mean	65660.37	60741.6	70579.13	58137.76	73182.97

In Table 2 ten years future forecast of sugarcane yield with 80% and 95% lower and upper confidence limits for sugarcane future yield. The next ten-year forecast of sugarcane yield of Pakistan from the ARIMA

model was averaged into 65660.37 and the expected percent change was computed and showed that the yield of the sugarcane crop is expected to increase.

Conclusions and Recommendations

The purpose of this study was to propose an optimum forecasting model for the yield of sugarcane crops in Pakistan. We compare ARIMA, ETS, TBATS, Artificial Neural Network (ANN), ARIMA-ETS, ARIMA-TBATS, and ARIMA-ANN hybrid models by calculating RMSE and MAE for each model. It was observed that the ARIMA model was best because it show the lowest values for RMSE (2345.059) and MAE (1879.447) for sugarcane yield. Average forecast yield of sugarcane crop is increased after ten years from 63827kg per hectare to 65660.37kg per hectare. This increase amount of production may meet the requirements of the country. More over these forecast estimates for sugarcane crop will be important for the Government in formulating their policies to fulfill the food necessities of the nation, trade, support prices, and planning about the cultivation sector.

Acknowledgements

We are thankful to the Department of Statistics and Economics, University of Management and Technology, Lahore for providing us the research environment.

Novelty Statement

Forecasting of sugarcane yield by developing linear, nonlinear and artificial intelligence model is a unique and latest study in the field time series analysis and agricultural forecasting.

Author's Contribution

Qaisar Mehmood: Conceptualization, formal analysis, methodology, software, writing original draft, project administration.

Ali Raza & Asif Ali Abro: Writing review and editing.

Nargis Shaheen: Resources, writing review and editing

Muhammad Riaz: Funding acquisition, investigation, software.

Conflicts of interest

The authors have declared no conflict of interest.

References

- Adisa, O., J. Botai and A. Adeola. 2019. Application of artificial neural network for predicting maize production in South Africa. *Sustainability*, 11(4): 1145–1227. <https://doi.org/10.3390/su11041145>
- Allen, P.G., 1994. Economic forecasting in agriculture. *Int. J. Forecast.*, 10(1): 81-135. [https://doi.org/10.1016/0169-2070\(94\)90052-3](https://doi.org/10.1016/0169-2070(94)90052-3)
- Aryal, R.D. and W. Yao-wu 2003. Neural network forecasting of the production level of Chinese construction industry. *J. Comp. Int. Manage.*, 6(2): 45-64.
- Azam, M. and M. Khan. 2010. Significance of the sugarcane crops with special and reference to NWFP. *Sarhad J. Agric.*, 26: 289-295.
- Bingjun, L., Z. Yifan, Z. Shuhua and L. Wenyan. 2021. Prediction of grain yield in Henan province based on grey BP neural network model. *Discrete Dynamics in Nature and Society*: <https://doi.org/10.1155/2021/9919332>
- Box, G.E.P. and G.M. Jenkins. 1976. *Time series analysis: Forecasting and Control*. Revised Ed. Holden Day.
- Dickey, A. and W.A. Fuller. 1979. Distribution of the estimators for autoregressive time series with a unit-root. *J. Am. Stat. Assoc.*, 74: 427-431. <https://doi.org/10.1080/01621459.1979.10482531>
- Economic Surveys of Pakistan, Ministry of Finance Pakistan. https://www.finance.gov.pk/survey_1920.html
- FAO, 2010. *Crop production*. Food and Agriculture Organization of the United Nations.
- Gopal, P.S.M. and R. Bhargavi. 2019a. Novel approach for efficient crop yield prediction. *Comp. Electron. Agric.* 165: Article ID 104968. <https://doi.org/10.1016/j.compag.2019.104968>.
- Gopal, P.S.M. and R. Bhargavi. 2019b. Optimum feature subset for optimizing crop yield prediction using filter and wrapper approaches. *Appl. Eng. Agric.*, 35(1): 9–14. <https://doi.org/10.13031/aea.12938>
- Gopal, P.S.M. and R. Bhargavi. 2019c. Performance evaluation of best feature subsets for crop

- yield prediction using machine learning algorithms. *Appl. Artif. Intell.*, 33(7): 621–642. <https://doi.org/10.1080/08839514.2019.1592343>
- Jiang, D., X. Yang, N. Clinton and N. Wang. 2004. An artificial neural network model for estimating crop yields using remotely sensed information. *Int. J. Remote Sens.*, 25(9): 1723–1732. <https://doi.org/10.1080/0143116031000150068>
- Krishna, S.R. and K.K. Suresh. 2010. A study on pre-harvest forecast of sugarcane yield using climatic variables. *Stat. Appl.*, 7 and 8(1 and 2): 1–8.
- Kumar, S., V. Kumar and R.K. Sharma. 2015. Sugarcane yield forecasting using artificial neural network models. *Int. J. Artif. Intell. Appl.*, 6(5): 51–68. <https://doi.org/10.5121/ijia.2015.6504>
- Laxmi, R.R. and A. Kumar. 2011. Weather based forecasting model for crops yield using neural network approach. *Stat. Appl.*, 9(1-2): 55–69.
- Masood, A.M. and M.A. Javed. 2004. Forecast models for sugarcane in Pakistan. *Pak. J. Agric. Sci.*, 41(1-2): 80–85.
- Mehmood, Q., M.H. Sial, M. Riaz and N. Shaheen. 2019. Forecasting the production of sugarcane in Pakistan for the year 2018–2030, using box-jenkin's methodology. *J. Anim. Plant Sci.*, 29(5): 1396–1401.
- Mehmood, Q., M.H. Sial, S. Sharif and M. Riaz. 2023. Development of conventional, artificial neural network and hybrid models for forecasting the wheat area and production of Pakistan. *Int. J. Agric. Stat. Sci.*, 19(1): 151–160. <https://doi.org/10.59467/IJASS.2023.19.151>
- Muhammad, F., M.S. Javed and M. Bashir. 1992. Forecasting sugarcane production in Pakistan using ARIMA models. *Pak. J. Agric. Sci.*, 9(1): 31–36.
- Nabeel, H., 2022. Predicting forecast of sugarcane production in Pakistan. *Sugar Tech.*
- Omer, O.B., G. Biricik and Z.C. Tayşi. 2017. Artificial neural network and SARIMA based models for power load forecasting in Turkish electricity market. *PLoS One*, 12(4): e0175915. <https://doi.org/10.1371/journal.pone.0175915>
- Ozozen, A., G. Kayakutlu, M. Ketterer and O. Kayalica. 2010. A combined seasonal ARIMA and ANN model for improved results in electricity spot price forecasting: Case study in Turkey. In 2016 Portland Int. Conf. Manage. Eng. Technol., pp. 2681–2690. <https://doi.org/10.1109/PICMET.2016.7806831>
- Perera, A. and U. Rathnayake. 2019. Rainfall and atmospheric temperature against the other climatic factors: A case study from colombo, Sri Lanka. *Math. Probl. Eng.*, 2019: 1–15. <https://doi.org/10.1155/2019/5692753>
- Riaz, M., M.H. Sial, S. Sharif and Q. Mehmood. 2023. Epidemiological forecasting models using ARIMA, SARIMA, and holt-winter multiplicative approach for Pakistan. *J. Environ. Publ. Health*, <https://doi.org/10.1155/2023/8907610>
- Rosenblatt, F., 1958. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.*, 65(6): 386. <https://doi.org/10.1037/h0042519>
- Sajid, A., N. Badar and H. Fatima. 2015. Forecasting production and yield of sugarcane and cotton crops of Pakistan for 2013–2030. *Sarhad J. Agric.*, 31(1): 1–10.
- Sreekanth, P.D., H. Geethanjali, P.D. Sreedevi, S. Ahmed, K.N. Kumar and K.P.D. Jayanthi. 2009. Forecasting groundwater level using artificial neural networks. *Curr. Sci.*, 96(7): 933.
- Supriya, S., A. Bhooshan, K. Binita, Y. Shikha, S. Alok and N. Mahima. 2023. Comparative analysis of sugarcane production for South East Asia Region. *Sugar Tech.* <https://doi.org/10.1007/s12355-023-01346-0>
- USDA, 2019. United States Department of Agriculture. Retrieved from <https://usdasearch.usda.gov/search?utf8=%E2%9C%93&affiliate=usda&query=list+of+sugar+exporting+countries&commit=Search>
- Vinushi, A., W. Lasini, P. Anushka, J. Jeevani and R. Upaka. 2020. Artificial neural network to estimate the paddy yield prediction using climatic data. *Math. Prob. Eng.*, <https://doi.org/10.1155/2020/8627824>
- Yaseen, M., M. Zakriya, I.U.D. Shahzad, M.I. Khan and M.A. Javad. 2005. Modeling and forecasting the sugarcane yield of Pakistan. *Int. J. Agric. Biol.*, 7: 180–183.