International Journal of Statistics and Applied Mathematics 2024; SP-9(1): 276-281

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2024; SP-9(1): 276-281

https://www.mathsjournal.com
Received: 08-11-2023
Accepted: 13-12-2023

Subrat Kumar Mahapatra
Ph.D. Research Scholar, Department of Agricultural Statistics, Institute of Agriculture, Visva-Bharati, Sriniketan, West Bengal, India

Digvijay Singh Dhakre
Assistant Professor, Department of Agricultural Statistics, Institute of Agriculture, Visva-Bharati, Sriniketan, West Bengal, India

Debasis Bhattacharya
Professor and Head, Department of Agricultural Statistics, Institute of Agriculture, Visva-Bharati, Sriniketan, West Bengal, India

Kader Ali Sarkar
Assistant Professor, Department of Agricultural Statistics, Institute of Agriculture, Visva-Bharati, Sriniketan, West Bengal, India

Corresponding Author:
Subrat Kumar Mahapatra
Ph.D. Research Scholar, Department of Agricultural Statistics, Institute of Agriculture, Visva-Bharati, Sriniketan, West Bengal, India

Forecasting of finger millet production in Odisha by ARIMA & ANN model: A comparative study

Subrat Kumar Mahapatra, Digvijay Singh Dhakre, Debasis Bhattacharya and Kader Ali Sarkar

Abstract

The Present study has been conducted to compare the forecasting of Autoregressive Integrated Moving Average (ARIMA) model and Artificial Neural Network (ANN) model for the finger millet production in Odisha. The Production data of finger millet from 1971-72 to 2020-21 have been collected from Department of Agriculture and Farmers empowerment, Govt of Odisha. The models were fitted using the 90% of dataset and the other 10% dataset have been used for cross validation. Different models have been identified and based on the lowest value of Root Mean Square Error and Mean Absolute Percentage Error; the most efficient forecasting model have been selected. The study finds that the Neural Network Autoregressive (NNAR) model is the best fitted model due to the lowest value of Root Mean Square Error (RMSE) and mean absolute percentage Error (MAPE). The best fitted model NNAR (5,4) is used to forecast the finger millet production for the upcoming years in Odisha. From this study, it is found that the production of finger millet will follow both increasing and decreasing trends in future years. The Production expected to decrease in 2021-22 then it will increase in 2022-23 again it follows decreasing trend in 2023-24 and increasing trend in 2024-25. it will expect to decrease in 2025-26.

Keywords: Forecasting, finger millet, ARIMA, ANN, RMSE, MAPE

1. Introduction

Finger Millet is an important Millet crops of Odisha, which is an annual herbaceous plant. It is widely cultivated in the arid and semiarid areas of Asia and covers 12% area in terms of global millet cultivation. Finger Millet or Ragi is a rich source of calcium, iron, protein and fibre. It is considered as one of the most nutritious cereals. Millet are being cultivated in 54495.83 ha, with Finger Millet occupying over 86% of area in Odisha. In 2017, Govt of Odisha started the Odisha Millet Mission (OMM) to increase in production, promoting household level consumption, improving the productivity, Promoting FPOs for marketing etc. So far more than 11 lakhs farmers of the state have taken up millet cultivation through improved agronomic practices. Millet is also helpful in income generation, nutritional security and helps in climate resilience (Patra and Mahapatra, 2020) [13]. Odisha is a tribal dominated state. Small and poor tribal farmers are involved in cultivation of millets. Millets are almost grown in the rural areas and the tribal dominated areas, where as the state government and OMM aims to promote the millet cultivation in almost all the districts of Odisha. As millet is more nutritious crops than the other cereals and pulses, the environment is suitable for the crop cultivation. The Present study has been taken up to analyse and Forecasting the Finger Millet production in different Statistical and Machine Learning Models such as ARIMA and ANN model. Forecasting of Crop Production and Productivity is considered to be the most important task as information on various factors are needed for the purpose of accurate forecasting. Different conditions like weather conditions, soil conditions and different crop management techniques have significant impacts on crop production and productivity (Singh *et al*, 2014) [25]. Unfavorable, unpleasant and aberrant weather conditions lead to crop loss up to 30% (Atri *et al*, 2003) [6]. For this, there is a huge demand for suitable statistical and machine learning models for forecasting of crop production and productivity, which will help the farmers for proper decision making and planning in cropping season. (Mahapatra and Satapathy, 2019) [19].

In recent days, prediction of crop productions and productivity using ARIMA and ANN models are quite popular, different attempts has been made to compare the ARIMA and NNAR model for the price forecasting of wheat and rice (Reza and Debnath, 2020) [1].

2. Materials and Methods

2.1 Data Sources

The secondary data on Finger Millet production from 1971-72 to 2020-21 were collected from Department of Agriculture and Farmers empowerment, Govt of Odisha. In this study, Autoregressive integrated moving average (ARIMA) model and Artificial Neural Network (ANN) model is used to compare the performance the forecasting Finger Millet production in Odisha. The Data from 1971-72 to 2015-16 have been used for model calibration and data from 2016-17 to 2020-21 have been used for model validation.

2.2 Autoregressive integrated moving average or ARIMA model

For the purpose of forecasting, a statistical model known as the autoregressive integrated moving average, or ARIMA model, is used. The goal of developing this model was to address the challenge of fitting Moving Average (MA) and Autoregression (AR) to characterize the data's dynamic structure. The ARMA models incorporate the differencing sequence, which is used to stationarize the data. The ARIMA model with parameter (p,d,q) is fitted using the univariate Box-Jenkins techniques (Box and Jenkins, 2015) [8]. This model includes the moving average of order q, the autoregressive of order p, and the order of differencing, which is d. If the mean and variance of a time series stay the same, it is said to be stationary. First, the original data is plotted and its stationarity is checked. If the data is shown on the graph to be non-stationary, the first difference of data is plotted and its stationarity is confirmed (Dash and Mahapatra, 2020) [13]. We continue in this manner until the data become stationary. The maximum order of differencing (d) is often two. Utilizing the Augmented Dicky-Fuller Test (ADF test), the stationarity of the data has been verified. David Dickey and Wayne Fuller developed the Dicky-Fuller test in 1979 to ascertain whether or not a given time series of data is stationary at the unit root. According to the hypothesis, a p-value of less than 0.05 must be obtained in order to prove stationarity. We evaluated the data for the first differencing to make sure it was stationary if the p-value was greater than 0.05. First order differencing is carried out in the first step; if the data are not stationary, second order differencing is carried out until the data are stationary.

$$\Delta Z_t = Z_t - Z_{t-1}$$

The above equation mentions the first order differencing.

2.3 Artificial Neural Network (ANN)

Artificial Neural Network is the important forecasting techniques which uses the Machine learning algorithms. The NNAR (Neural Network Autoregressive) model has been taken into consideration in this study for the purposes of prediction and comparison. The lagged values in the time series data are typically used as the input for the NNAR Model's prediction procedure. NNAR (p, k) indicates that there are k nodes and p-lagged values in the hidden layer. While ARIMA models are utilized for datasets with linear relationships, NNAR models typically deal with non-linearity in data. ANN is therefore regarded as a particular kind of non-linear machine learning forecasting method. It consists of three interconnected layers: the hidden layer (one to three layers of neurons), the input layer (nodes or units), and the output layers of neurons.

2.4 Model Performance Evaluation

The performance of the selected models was evaluated using the root mean square error (RMSE), the mean absolute percentage error (MAPE). The formulae of the model evaluation measures are shown below

$$\text{RMSE} = \frac{\sum_t e_t^2}{n-2}$$

$$\text{MAPE} = \frac{1}{n}\frac{\sum_t |Y_t - \hat{Y}_t|}{Y_t} \times 100$$

The models among all the selected ARIMA and ANN models which have lowest value of RMSE and MAPE is considered to be the best-fit model.

2.5 Calculation of Percent forecast error

For the aim of cross-validating the model, about 10% of the data are used. In this study, the model building (calibration) phase uses data from 1971-72 to 2015-16, while the model validation phase uses data from 2016-17 to 2020-21.

$$\text{Percent forecast error} = \left(\frac{Y-\hat{Y}}{Y}\right) \times 100$$

Where Y is the value that was recorded between 2016-17 and 2020-21,

The predicted value from 2016-17 to 2020-21 is denoted by $\hat{Y}$.

Better prediction performance is indicated by a lower percent forecast error value.

The best fit model is chosen and utilized for forecasting after a successful validation process. Here, the Finger millet production from 2021-2022 to 2025-2026 is forecasted using the best fit model that was chosen. The ARIMA model and ANN are used for time series modeling and forecasting are analyzed by using R Software.

3. Results

3.1 Checking for Stationarity

Figure 1 displays the original plot of the Finger Millet Production data, which indicates that the data are non-stationary. ADF, or the Augmented Dicky-Fuller Test, is used to verify the stationarity of data. The ADF test yields a p-value of 0.335, which is above the 5% significance level (α=0.05), indicating non-stationarity in the data. As a result, first order differencing is carried out, and following this, the test's p-value is discovered to be 0.01. Therefore, Figure 2 leads us to the conclusion that the data are now stationary.

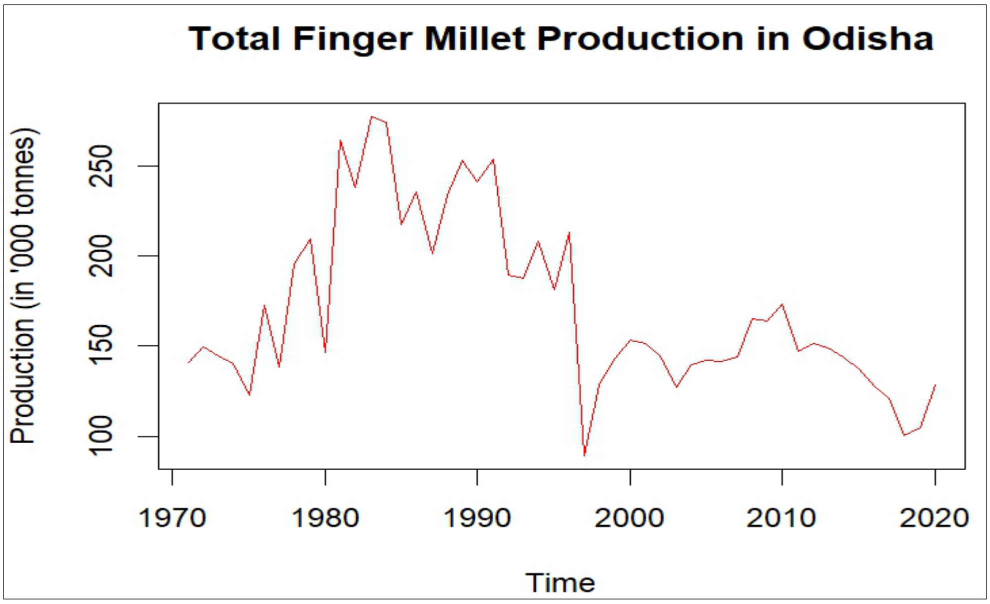

Fig 1: Plot of original value of Production vs time

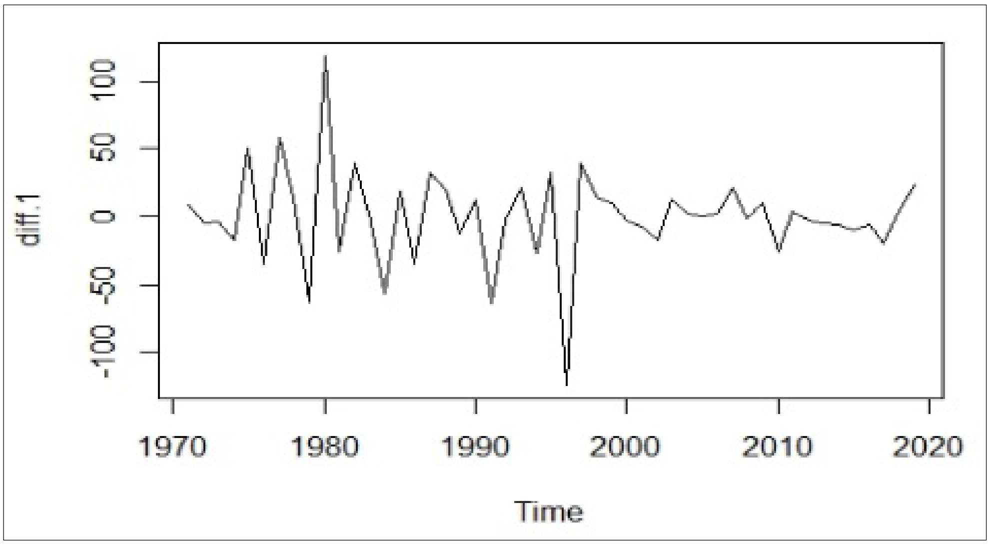

Fig 2: Plot of first differences value of Production vs time

The first difference value of Finger Millet Production plotted as an ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) in Figure 3 indicates that q=0, p=1 and q=1, p=1 are the approximate values of p and q that would be appropriate for yield. As a result, ARIMA (1,1,0) and ARIMA (1,1,1) are the ARIMA models chosen for this investigation. Similarly, the various values of p and k are examined, and NNAR (2, 5), NNAR (3, 2), NNAR (3, 4), and NNAR (5, 4) are the chosen artificial neural network model for this investigation.

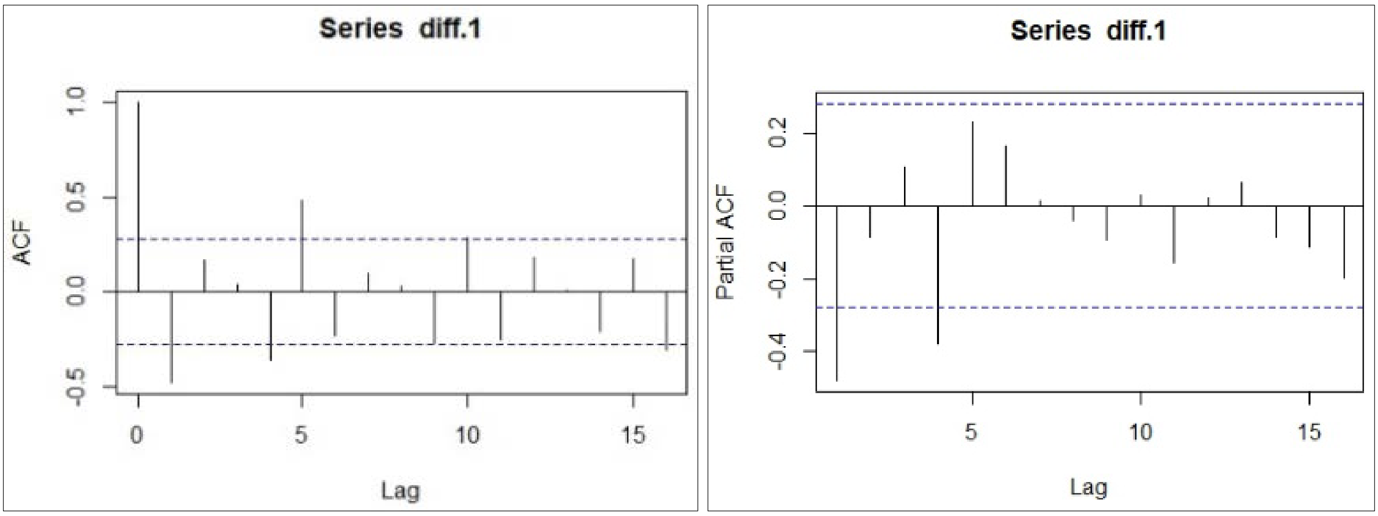

Fig 3: ACF and PACF plot after first differencing the value of production

3.2 Evaluation of the Model Performance

The performance of the models was evaluated using the root mean square error (RMSE), the mean absolute percentage error (MAPE). The value of RMSE and MAPE of all the selected models are presented in Table 1.

Table 1: Comparison of different models for training and testing

	Criteria	ARIMA (1,1,0)	ARIMA (1,1,1)	NNAR (2,5)	NNAR (3,2)	NNAR (3,4)	NNAR (5,4)
Calibration (Training)	RMSE	31.65	31.54	14.13	20.31	15.07	11.15
	MAPE	27.75	26.41	31.90	34.58	28.79	26.28
Validation (Testing)	RMSE	13.40	13.41	6.69	9.48	6.67	5.18
	MAPE	21.98	22.63	26.82	29.11	23.04	20.02
Whole dataset	RMSE						12.27
	MAPE						21.15

From the above table, it has been found that the lowest value of RMSE and MAPE is seen in NNAR (5,4) for both the training data, testing data and the whole dataset, so that NNAR (5,4) is selected as the best fitted model for Finger Millet production in Odisha.

3.3 Cross Validation of Model and Calculation of percentage forecasting error

The actual production, predicted production and their corresponding forecasting errors for each fitted model is presented in Table 2.

Table 2: The actual Production, predicted Production and their corresponding error

Year	Actual Production	ARIMA (1,1,0)			ARIMA (1,1,1)			NNAR (2,5)		
		Predicted Production	% Error	Mean % Error	Predicted Production	% Error	Mean % Error	Predicted Production	% Error	Mean % Error
2016-17	127.65	140.43	10.01		141.18	10.59		146.27	14.58	
2017-18	120.92	138.95	14.91		139.80	15.61		147.37	21.87	
2018-19	100.58	139.66	38.85	20.98	140.30	39.49		146.12	45.27	
2019-20	104.92	139.32	32.78		140.12	33.54		145.78	38.94	
2020-21	128.73	139.48	8.35		138.87	7.87	21.42	146.18	13.55	26.84

Table 2: Continued

Year	Actual Production	NNAR (3,2)			NNAR (3,4)			NNAR (5,4)		
		Predicted Production	% Error	Mean % Error	Predicted Production	% Error	Mean % Error	Predicted Production	% Error	Mean % Error
2016-17	127.65	148.08	16.00		139.67	9.41		135.58	6.21	
2017-18	120.92	149.52	23.65		145.69	20.48		140.34	16.06	
2018-19	100.58	150.22	49.35	26.32	144.14	43.30	23.11	136.21	35.42	20.07
2019-20	104.92	138.65	32.14		142.74	36.04		138.76	32.25	
2020-21	128.73	142.23	10.48		136.87	6.32		142.12	10.40	

From Table 2, it is concluded that the mean percent forecast error in NNAR (5,4) is found to be 20.07, which is lowest among all the selected model.

3.4 Forecasting for coming years

From this comparative study, it is found that the Neural Network Autoregression or the ANN model is perform better than the traditional time series model ARIMA. Out of the six selected ARIMA & ANN model, The best fitted NNAR (5, 4) model is used to forecast the finger millet production for the next 5 years.

Table 3: Forecasted Finger Millet Production by using NNAR (5, 4) Model

Year	Yield
2021-22	138.12
2022-23	163.11
2023-24	157.13
2024-25	171.12
2025-26	147.48

The Forecast of Finger Millet production from 2021-22 to 2025-26 by NNAR (5,4) model is presented in Figure 4.

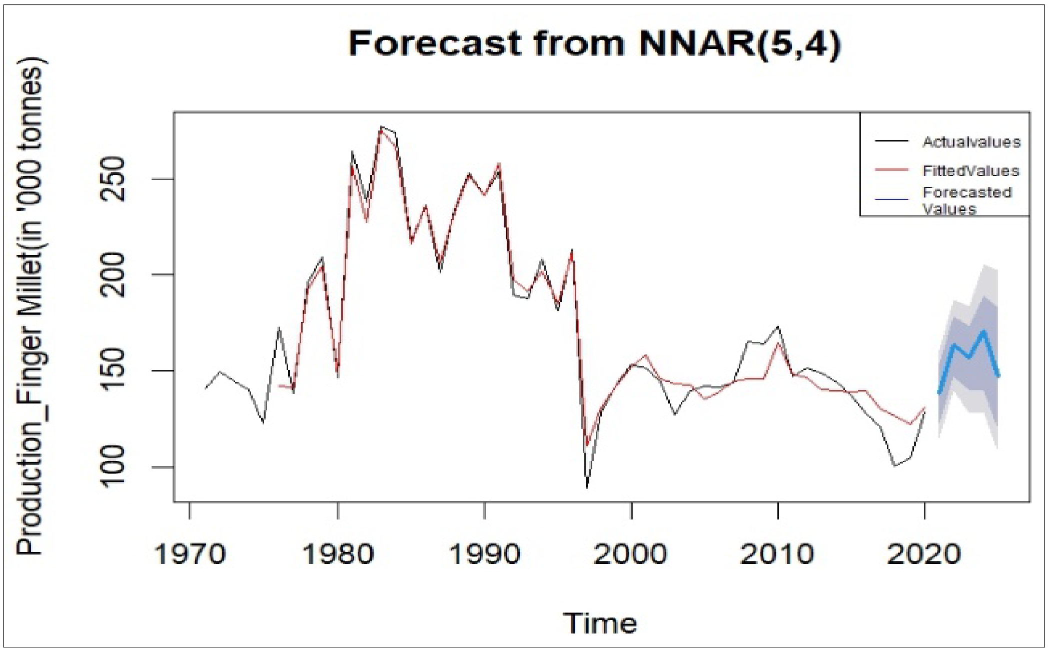

Fig 4: Forecasted Value of Finger Millet production from 2021-22 to 2025-26

4. Conclusion

In the comparative analysis, different ARIMA and ANN model have been selected based on different parameters. The models are evaluated by using their values of RMSE and MAPE. Among all the selected model, NNAR (5,4) is selected as the best fitted model due to the low value of RMSE & MAPE. From the above comparative study, it has been found that the ANN model performs better than the ARIMA Model as ANN is the non-linear forecasting model, having a potential advantage in the analysis as compare to the traditional ARIMA model. ANN model performance was found to be more effective in the production of finger millet in Odisha. This finding is supported by Setiya *et al* (Setiya *et al*, 2022) [23]. Forecasting of Finger Millet production from 2021-22 to 2025-26 is done by using the best fitted NNAR (5,4) model. From the study, it is concluded that the production of finger millet is expected to decrease in 2021-22 then it will increase in 2022-23 again it follows decreasing trend in 2023-24 and increasing trend in 2024-25. it will expect to decrease in 2025-26.

5. Author Contributions: Conceptualization-SKM, methodology-DB, KAS, software- SKM, DSD, validation and formal analysis- SKM, DSD, writing - original draft preparation and editing-SKM, DSD, DB, KAS

6. Funding: This research received no funding

7. Conflicts of Interest: The authors declare no conflict of interest.

8. References

1. Aashiqur Reza DS, Debnath T. An Approach to Make Comparison of ARIMA and NNAR Models for Forecasting Price of Commodities; c2020.
2. Abotaleb M, Ray S, Mishra P, Karakaya K, Shok C, Al Khatib A, *et al*. Modelling and forecasting of rice production in south Asian countries. Ama, Agricultural Mechanization in Asia, Africa & Latin America. 2021;51:1611-1627.
3. Alpaydin E. Introduction to Machine Learning. 2nd ed; c2010. Available from: https://books.google.nl/books?hl=nl&lr=&id=TtrxCwAAQBAJ&oi=fnd&pg=PR7&dq=introduction+to+machine+learning&ots=T5ejQG_7pZ&sig=0xC_H0agN7mPhYW7oQsWiMVwRnQ#v=onepage&q=introduction%20to%20machine%20learning&f=false.
4. Annamalai N, Johnson A. Analysis and Forecasting of Area Under Cultivation of Rice in India: Univariate Time Series Approach. SN Computer Science. 2023;4:193. https://doi.org/10.1007/s42979-022-01604-0.
5. Arvind KS, Vashisth A, Krishnan P, Das B. Wheat Yield Prediction Based on Weather Parameters Using Multiple Linear, Neural Network and Penalized Regression Models. J Agrometeorol. 2022;24:18-25.
6. Attri SD, Rathore LS. Pre-Harvest Estimation of Wheat Yield for NW India Using Climate and Weather Forecast. Mausam. 2003;54:729-738.
7. Badmus MA, Ariyo OS. Forecasting Cultivated Areas and Production of Maize in Nigerian using ARIMA Model. Asian Journal of Agricultural Sciences. 2011;3(3):171-176.
8. Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control. John Wiley & Sons; c2015.
9. Casdagli M, Eubank S, Farmer JD, Gibson J. State Space Reconstruction in the Presence of Noise. Phys. Nonlinear Phenom. 1991;51:52-98.
10. Das B, Murgaonkar D, Navyashree S, Kumar P. Novel Combination Artificial Neural Network Models Could Not Outperform Individual Models for Weather-Based Cashew Yield Prediction. Int. J Biometeorol. 2022;66:1627-1638.
11. Das B, Nair B, Arunachalam V, Reddy KV, Venkatesh P, Chakraborty D, *et al*. Comparative Evaluation of Linear and Nonlinear Weather-Based Models for Coconut Yield Prediction in the West Coast of India. Int. J Biometeorol. 2020;64:1111-1123.
12. Das B, Nair B, Reddy VK, Venkatesh P. Evaluation of Multiple Linear, Neural Network and Penalised Regression Models for Prediction of Rice Yield Based on Weather Parameters for West Coast of India. Int. J. Biometeorol. 2018;62:1809-1822.

13. Dash A, Mahapatra SK. Using ARIMA model for yield forecasting of important pulse crops of Odisha. Amazonian Journal of Plant Research. 2020;4(3):646-659.
14. Filippi P, Jones EJ, Wimalathunge NS, Somarathna PDSN, Pozza LE, Ugbaje SU, *et al*. An approach to forecast grain crop yield using multilayered, multi-farm data sets and machine learning. Precis. Agric; c2019. https://doi.org/10.1007/s11119-018-09628-4.
15. Hastie T, Tibshirani R, Friedman JH, Friedman JH. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer: New York, NY, USA. 2009;2:1-758.
16. Hemavathi M, Prabakaran K. ARIMA model for forecasting of area, production and productivity of rice and its growth status in Thanjavur District of Tamil Nadu, India. International Journal of Current Microbiology and Applied Sciences. 2018;7:149-156.
17. Karadas, Celik, Eyduran, Hopoglu. Forecasting Production of Some Oil Seed Crops in Turkey Using Exponential Smoothing Methods. J Anim Plant Sci; c2017, 27(5).
18. Khaki S, Wang L. Crop yield prediction using deep neural networks. Front. Plant Sci. 2019;10:621.
19. Mahapatra SK, Satapathy A. An application of box-jenkins methodology for forecasting of green gram productivity in Odisha. J Pharmacogn Phytochem. 2019;8(3):4383-4387.
20. Patra C, Mahapatra SK. Forecasting of Ragi Production in Koraput Districts of Odisha. International Journal of Current Microbiology and Applied Science. 2020;9(07):1923-1929.
21. Safa M, Samarasinghe S, Nejat M. Prediction of Wheat Production Using Artificial Neural Networks and Investigating Indirect Factors Affecting It: Case Study in Canterbury Province, New Zealand. J Agric. Sci. Technol. 2015;17:791-803.
22. Satpathi A, Setiya P, Das B, Nain AS, Jha PK, Singh S, *et al*. Comparative Analysis of Statistical and Machine Learning Techniques for Rice Yield Forecasting for Chhattisgarh, India. Sustainability. 2023;15:2786.
23. Setiya P, Satpathi A, Nain AS, Das B. Comparison of Weather-Based Wheat Yield Forecasting Models for Different Districts of Uttarakhand Using Statistical and Machine Learning Techniques. J Agrometeorol. 2022;24:255-261.
24. Sharma PK, Dwivedi S, Ali L, Arora RK. Forecasting maize production in India using ARIMA model. Agro-Economist. 2018;5(1):1-6.
25. Singh AK, Vashisth A, Sehgal VK, Goyal A, Pathak H, Parihar SS, *et al*. Development of Multi Stage District Level Wheat Yield Forecast Models. J Agric. Phys. 2014;14:189-193.
26. Soto-Ferrari M, Chams-Anturi O, Escorcia-Caballero JP. A time-series forecasting performance comparison for neural networks with state space and ARIMA models.
27. Sridhara S, Ramesh N, Gopakkali P, Das B, Venkatappa SD, Sanjivaiah SH, *et al*. Weather-Based Neural Network, Stepwise Linear and Sparse Regression Approach for Rabi Sorghum Yield Forecasting of Karnataka, India. Agronomy. 2020;10:1645.
28. Tripathi R, Nayak AK, Raja R, Shahid M, Kumar A, Mohanty S, *et al*. Forecasting rice productivity and production of Odisha, India, using autoregressive integrated moving average models. Adv Agriculture; c2014. p. 1-9. https://doi.org/10.1155/2014/621313.
29. Wang SC. Artificial Neural Network. In Interdisciplinary Computing in Java Programming; Springer US: Boston, MA, USA; c2003. p. 81-100.
30. Xu X, Gao P, Zhu X, Guo W, Ding J, Li C, *et al*. Design of an integrated climatic assessment indicator (ICAI) for wheat production: a case study in Jiangsu Province, China. Ecol. Ind. 2019;101:943-953. https://doi.org/10.1016/j.ecolind.2019.01.059.