International Journal of Statistics and Applied Mathematics 2025; 10(5): 53-58

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2025; 10(5): 53-58

www.mathsjournal.com
Received: 11-03-2025
Accepted: 14-04-2025

Sanketh Raj H
M.Sc. Student, Department of Statistics and Computer Applications, ANGRAU-Agricultural College, Bapatla, Andhra Pradesh, India

B Ramana Murthy
Assistant Professor, Department of Statistics and Computer Applications, ANGRAU-S.V. Agricultural College, Tirupati, Andhra Pradesh, India

KN Sreenivasulu
Assistant Professor, Department of Statistics and Computer Applications, ANGRAU-Agricultural College, Pulivendula, Andhra Pradesh, India

V Sita Ram Babu
Assistant Professor, Department of Agricultural Economics, ANGRAU-Agricultural College, Bapatla, Andhra Pradesh, India

Corresponding Author:
Sanketh Raj H
M.Sc. Student, Department of Statistics and Computer Applications, ANGRAU-Agricultural College, Bapatla, Andhra Pradesh, India

Hybrid statistical model for forecasting production of maize crop in Karnataka State, India

Sanketh Raj H, B Ramana Murthy, KN Sreenivasulu and V Sita Ram Babu

DOI: https://www.doi.org/10.22271/maths.2025.v10.i5a.2035

Abstract
The present study was carried out to apply the Hybrid statistical model specifically Auto Regressive Integrated Moving Average (ARIMA) combined with Time-Delay Neural Network (TDNN) on Maize production in Karnataka state for the period of 1962 to 2022. The ARIMA (3, 1, 2) model were found to be best forecasted model for Maize production by criteria, which having highest R^2, low Akaiki Information criteria (AIC) and low Root Mean Square Error (RMSE). Consequently, a nonlinear TDNN model was applied to fit the residuals. For Maize production forecasting, the TDNN (4-15-1) model with 4 input delays, 15 hidden neurons, and 1 output neuron demonstrated low RMSE and low Mean Absolute Percentage Error (MAPE) ensuring high accuracy. The ARIMA (3,1,2) model with TDNN (4-15-1) model on ARIMA residuals was found to be the best forecasting model for Maize production of Karnataka state, India. Finally, the hybrid model ARIMA with TDNN was improved the forecasting accuracy. The forecasted results for Maize production for Karnataka state were shown an increasing trend.

Keywords: Maize, Production, R^2, AIC, RMSE, MAPE, ARIMA, TDNN models and Trend

Introduction
Agriculture the backbone of Indian economy accounting for 14 percent of the nation's Gross Domestic Product (GDP) and about 11 percent of the country's exports. Nearly 65 percent of country's population still depends on agriculture for employment and livelihood**.** Cereal crops form the cornerstone of global food security, serving as staple foods for more than half of the world's population. These crops such as Rice, Wheat, Maize, Barley, Sorghum, and Millets, are rich sources of carbohydrates and play a vital role in human nutrition and animal feed. In India, cereals occupy a major share of the gross cropped area and contribute significantly to both caloric intake and rural livelihoods. Their adaptability to diverse agro-climatic conditions and suitability for rainfed farming make them integral to the country's agricultural landscape. Among these, maize has emerged as a dynamic and versatile crop with increasing relevance across sectors.

Maize (*Zea mays L.*), popularly known as the "Queen of cereals" is the third most important cereal crop in India after Rice and Wheat, owing to its high yield potential, nutritional value, and diverse industrial applications. Globally, Maize ranks first in terms of production and is a vital component of food, feed and fuel sectors. In India, during 2023-2024, Maize was cultivated around 112.41 lakh hectares, with a production of 376.65 lakh tonnes, with a productivity of 3351 kg/ha (Indiastat, 2023-24). The country has witnessed a steady rise in maize cultivation due to growing demand in the poultry feed, starch and biofuel industries.

Karnataka is the leading producer of Maize in India, contributing 15 per cent to national production. In 2023-24, the state recorded Maize cultivation 19.72 lakh hectares, yielding 56.292 lakh tonnes and a productivity of 2855 kg/ha (Indiastat, 2023-24). The districts of Davanagere, Haveri, Belagavi, Chitra Durga and Ballari are prominent Maize-growing regions, primarily under rainfed conditions, although irrigated Maize is gaining importance during the Rabi season.

Maize serves multiple roles in Karnataka's agrarian economy. It is a source of income for farmers, raw material for industries and a crucial input in livestock feed. It is rich in carbohydrates, moderately high in protein, and a good source of vitamins like B-complex. Despite these advantages, maize cultivation faces several challenges, including pest outbreaks such as Fall Armyworm, erratic rainfall patterns, fluctuating market prices, and rising input costs.

Over the past two decades (2000-2020), production of Maize in Karnataka have shown significant growth, reflecting both policy interventions and market-driven shifts. However, the year-to-year variation also shows nonlinear trends influenced by climatic factors and external shocks. This complexity underscores the need for accurate forecasting models that go beyond traditional linear approaches. Therefore, Maize has been selected as one of the key crops for this study, which explores advanced hybrid forecasting models like ARIMA-TDNN to predict its production. These models aim to provide reliable insights for stakeholders involved in agricultural planning, policy formulation, and market regulation.

Statistical forecasting is used to provide assistance in decision making and planning the future more effectively and efficiently. Forecasting is a primary aspect of developing economy so that proper planning can be undertaken for sustainable growth of the country. Considering the above mentioned facts, a study was conducted to model and forecast the Maize production in Karnataka state, India.

The major drawback of the Auto Regressive Integrated Moving Average (ARIMA) model is that it does not capture nonlinear structure. It assumes a linear correlation pattern among the time series data, and therefore cannot adequately capture nonlinear patterns often present in agricultural datasets. In reality, time series data such as those for maize production in Karnataka state often include both linear and nonlinear components due to variations in weather, market conditions, and technological adoption.

As a result, the performance of ARIMA alone becomes limited when dealing with such mixed data structures. To overcome this challenge, researchers have developed hybrid forecasting models that combine the linear modelling capabilities of ARIMA with the nonlinear learning strengths of neural networks. In this study, a hybrid ARIMA-TDNN model is used to forecast Maize production in Karnataka state. While ARIMA is used to model the linear structure of the series, TDNN is employed to capture the nonlinear residuals that ARIMA fails to explain. TDNN, being a dynamic version of neural networks, utilizes past time-lagged inputs to learn from temporal dependencies in the data.

Studies such as Zhang (2003) [10], Kumar and Prajneshu (2015) [5] and Ray *et al.* (2016) [9] have demonstrated that such hybrid models can significantly enhance the accuracy and reliability of forecasts. Thus, the integration of ARIMA and TDNN in this research provides a more comprehensive forecasting framework, particularly suitable for agricultural time series data. The resulting model offers improved predictive insights for stake holders involved in planning, policy making, and resource allocation related to Maize cultivation in Karnataka state.

Materials and Methods

Data Description

The study has been illustrated with the time series data. The secondary data on the Maize production in Karnataka state for 60 years (from 1962-63 to 2021-22) was obtained from the economic and political weekly research foundation (EPWRF India Time Series) and Indiastat.com.

Statistical Methodologies

In this study, Auto Regressive Integrated Moving Average (ARIMA), Time Delay Neural Network (TDNN), and a proposed hybrid methodology combining ARIMA-TDNN are employed to forecast the Production of Maize in Karnataka state, India. The analysis is carried out using R software. To examine the stationarity of the time series data, the Augmented Dickey-Fuller (ADF) test is performed. The Brock-Dechert-Scheinkman (BDS) test is used to detect nonlinear dependencies in the residuals, ensuring an appropriate model selection process. Additionally, the Box-Pierce/Ljung-Box test is applied to check for autocorrelation in the residuals, validating the adequacy of the fitted models. The forecasting performance of the models is assessed using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), which provide measures of predictive accuracy and reliability.

Auto Regressive Integrated Moving Average (ARIMA) Model

One of the most important and widely used classical time series models is the Auto Regressive Integrated Moving Average (ARIMA) model. The popularity of the ARIMA model is due to its linear statistical properties as well as the popular Box-Jenkins methodology (Box and Jenkins, 1970) for model building procedure.

The general form of the ARIMA model is represented as ARIMA (p, d, q) where:

- p: Number of lag observations included in the model (Autoregressive term)
- d: Number of times the data has been differenced (Integrated term)
- q: Size of the moving average window (Moving Average term)

The ARIMA (p, d, q) model is expressed as:

$$Y_t = c + \sum_{i=1}^{p} \emptyset_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t$$

Where,

- Y_t is the actual value at time t
- c is a constant
- ϕ_i are the coefficients for the AR terms
- θ_j are the coefficients for the MA terms
- ϵ_t is the white noise error term

The Box-Jenkins ARIMA model building consists of three steps *viz.*, identification, estimation, and diagnostic checking. First step in model building is to identify the model i.e. to determine the model order. Second step is to estimate the parameters of model based on identified model order. Finally, the third step is diagnostic checking of residuals.

Time Delay Neural Network (TDNN)

The TDNN model is a type of artificial neural network that is designed to capture non-linear relationships in time-series data by introducing time delays in the input layer. This network uses past lagged observations as input to model complex patterns in the data.

The TDNN can be mathematically expressed as:

$$\hat{Y}_t = f\left(Y_{t-1,}Y_{t-2,} \dots \dots, Y_{t-p}\, ; W\right)$$

Where,

- $\hat{Y}_t$ is the predicted value at time t
- $f\,(\cdot)$represents the non-linear activation function
- $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}$ are the lagged inputs
- W represents the weights of the network

The TDNN model learns the optimal weights by minimizing the error between predicted and actual values using back propagation.

Hybrid ARIMA-TDNN Model

The hybrid ARIMA-TDNN model is constructed by first fitting the ARIMA model to capture the linear patterns in the data and then applying the TDNN model to the residuals from the ARIMA model to account for non-linear patterns.

The hybrid model can be formulated as:

$$\hat{Y}_t = \hat{Y}_{ARIMA_t} + \hat{Y}_{TDNN_t}$$

Where,

- $\hat{Y}_t$is the final forecasted value
- $\hat{Y}_{ARIMA_t}$ is the ARIMA forecasted value at time t
- $\hat{Y}_{TDNN_t}$ is the TDNN forecasted residual value at time t

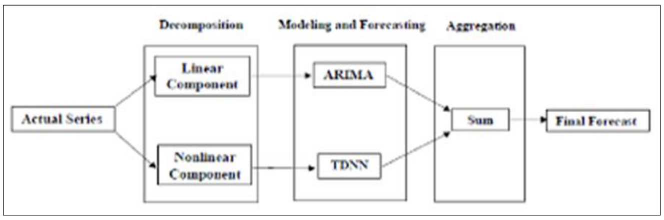

Fig 1: ARIMA-TDNN as a Hybrid model

Performance Criteria

The performance of the models was evaluated using the following criteria.

1) Root Mean Square Error (RMSE): Measures average magnitude of prediction error

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2}{n}}$$

A model having lower value of RMSE is considered as best model.

2) Mean Absolute Error (MAE): Average of absolute errors

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|Y_t - \hat{Y}_t\right|$$

A model having lower value of MAE is considered as best model.

3) Akaike Information Criterion (AIC): Balances model fit and complexity. Penalizes over fitting.

$$AIC=2k-2\ln(L)$$

A model having lower value of AIC is considered as best model.

4) Mean Absolute Percentage Error (MAPE): Average of percentage errors. Gives error in percentage terms

$$MAPE = \left(\frac{100}{n}\right)\sum_{i=1}^{n}\left|\frac{(A_i - F_i)}{A_i}\right|$$

A model having lower value of MAPE is considered as best model.

5) The coefficient of determination (R^2)

$$R^2 = 1 - \frac{(\,\Sigma\,(y_i - \hat{y}_i)^2)}{(\,\Sigma\,(y_i - \bar{y})^2)}$$

An R^2 value closer to 1 indicates a better fit of the model to the data.

Results and Discussion

Descriptive statistics

Table 1: Descriptive statistics of maize production

Maize Production	
Mean	1560.005333
Standard Error	198.9737953
Median	901.35
Mode	
Standard Deviation	1541.244391
Sample Variance	2375434.274
Kurtosis	-0.439690516
Skewness	0.928500887
Range	5354.43
Minimum	7.7
Maximum	5362.13
Sum	93600.32
Count	60

Production was right-skewed, meaning there has been significant expansion in recent decades, but earlier years were low in both metrics. Standard deviations are very high, indicating instability or major growth trends over time possibly influenced by policies, irrigation, or technological interventions. The mean> median pattern in both variables confirms that recent high values are pulling the average up. The lack of mode and non-normal distribution characteristics (negative-kurtosis, skewness) suggest that simple models might not capture the trends well supporting the use of hybrid models like ARIMA-TDNN as optimal.

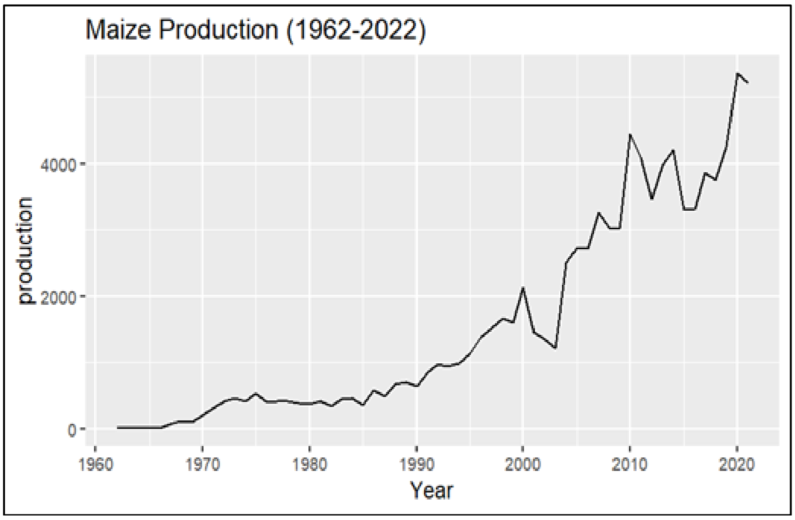

Fig 2: Time series data of Maize Production

Results of ARIMA Model

To analyse the trend in Maize production from 1962 to 2022, ARIMA (Auto Regressive Integrated Moving Average) model were developed. The dataset was split into training and testing sets in a 1:57 and 58:60 ratio, ensuring enough data points for both model training and validation.

Model Selection and Evaluation

The differencing parameter was determined to be d = 1 based on stationarity requirements. Tested using various combinations (p, d, q), and their performance was evaluated using Root Mean Square Error (RMSE), AIC and R^2.

Table 2: ARIMA Model Comparison (based on RMSE)

ARIMA Model	RMSE	AIC	R^2
(2,1,0)	330.19	816.92	0.9398
(2,1,1)	327.20	816.77	0.9409
(3,1,0)	323.65	815.62	0.9421
(3,1,1)	318.26	815.92	0.9441
(3, 1, 2)	312.75	816.28	0.9460

Among all the models, ARIMA (3, 1, 2) showed the lowest RMSE value (312.75), indicating the best performance. This model explained approximately 94.60% of the variation, making it the most suitable for forecasting.

Model Diagnostics

The ADF test confirmed stationarity after first differencing. The estimated coefficients of the selected ARIMA (3, 1, 2) model are shown below:

Table 3: Coefficients of ARIMA (3, 1, 2)

Parameter	Estimate
ar1	-0.7975
ar2	-0.9752
ar3	-0.1064
ma1	0.6841
ma2	0.6699

Forecast vs Actual Comparison

Table 4: Actual vs Forecasted Values using ARIMA model

Year	Actual Value	Forecasted Value
2020	4250.02	3579.2
2021	5362.13	3969.51
2022	5220.7	3842.53

The performance of the ARIMA (3, 1, 2) model was validated using test data. The ARIMA model for maize production forecasting showed moderate accuracy on test data with an RMSE of 1197.14, MAE of 1149.86, and MAPE of 22.7. The actual and forecasted values are presented below:

Results of TDNN Model

Before fitting the forecasting models, a BDS test was conducted on the training dataset to investigate the presence of nonlinear dependence within the series. The test results, revealed p-values of less than 0.05 across all embedding dimensions and epsilon values, providing strong evidence of nonlinearity in the series. This preliminary diagnostic confirmed that the data exhibited complex nonlinear patterns, thereby justifying the adoption of nonlinear modelling techniques. Consequently, a Time Delay Neural Network (TDNN) was employed, given its ability to effectively capture nonlinear and dynamic structures in time series forecasting applications.

Table 5: TDNN model comparison

Particulars	Network structure	RMSE	MAE	MAPE
Production	4-6-1	68.42	54.38	24
	4-7-1	61.83	48.94	21
	4-8-1	57.41	43.55	17
	4-9-1	56.60	43.19	20
	4-10-1	57.93	44.38	28

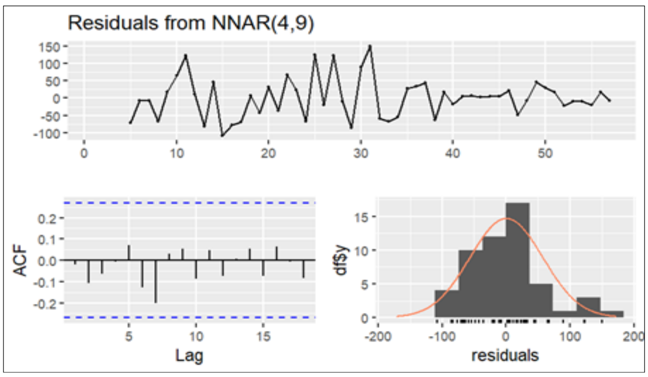

Fig 3: Residual Diagnostic Plot for NNAR (4-9-1) for train data

The model structure that provided the best performance for the training data was a 4-9-1 neural network, indicating 4 input nodes, 9 hidden neurons, and 1 output node with better diagnostic criteria like RMSE, MAPE and MAE values. The analysis was carried out using 57 data points for training and 3 data points for testing.

This figure displays the diagnostic plots for residuals from the NNAR (4, 9) model. The top panel shows the residuals over time, indicating their randomness. The bottom-left panel presents the autocorrelation function (ACF) of residuals to check for any significant lags. The bottom-right panel shows the histogram of residuals with a fitted normal curve to assess normality. The residual diagnostics of the NNAR (4-9-1) model were examined to assess model adequacy. The residual plot exhibited random fluctuations around zero, and the ACF of residuals showed no significant autocorrelations, indicating that the model has captured the essential structure of the series. Additionally, the Ljung-Box test was applied up to lag 10, yielding a Q* statistic of 5.5593 with a p-value of 0.8508. This confirms that the residuals are not significantly autocorrelated and can be treated as white noise. These findings collectively suggest that the NNAR (4-9-1) model provides a good fit for the data. The TDNN model for maize production forecasting showed good performance with an RMSE of 957.29, MAE of 893.93, and MAPE of 18.18 on test data.

Table 6: Actual vs Forecasted values using TDNN model

Year	Actual Value	Forecasted Value
2020	4250.02	3367.96
2021	5362.13	4046.82
2022	5220.7	4744.25

Results of ARIMA-TDNN model

The ARIMA-TDNN hybrid model is a combination of linear and nonlinear modelling techniques used for more accurate time series forecasting. It aims to capture both the linear structure of data (handled by ARIMA) and the nonlinear patterns (modelled by TDNN). ARIMA (Auto Regressive Integrated Moving Average) is effective for capturing linear patterns in time series data. TDNN is a type of feedforward neural network specifically designed to handle time-dependent data with lags or delays, capturing complex nonlinear relationship. Real-world time series often contain both linear and nonlinear components. A single model may not be sufficient. ARIMA handles the linear part, but leaves behind nonlinear residuals. TDNN models on ARIMA residuals to capture the remaining structure and improve accuracy.

Table 7: TDNN Model Performance on ARIMA Residuals

Model	RMSE	MAE	MAPE
4-13-1	19.69	11.75	4
4-14-1	14.98	9.33	3.5
4-15-1	11.35	6.49	2
4-16-1	14.6	7.9	29

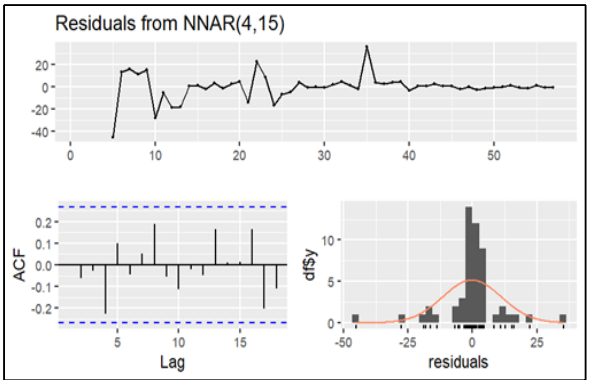

Fig 4: Residual diagnostic plot for NNAR (4-15-1) for ARIMA residuals

The residual diagnostics for the NNAR (4-15-1) model indicated that the residuals fluctuated randomly around zero without any discernible pattern. The autocorrelation function (ACF) plot revealed no significant autocorrelations up to lag 18, suggesting that the model had successfully captured the data's temporal structure. Additionally, the residuals exhibited an approximately normal distribution. The Ljung-Box test further confirmed this, yielding a Q* statistic of 7.6267 with a p-value of 0.6653, thereby failing to reject the null hypothesis of no autocorrelation. These diagnostics collectively validate the adequacy of the NNAR (4, 15) model in modelling the given time series data. The Hybrid ARIMA-TDNN model for maize production forecasting performed well on test data, recording an RMSE of 880.92, MAE of 837.39, and MAPE of 15.

Table 8: Hybrid Model comparison

Hybrid Model	Performance criteria		
ARIMA+ TDNN	RMSE	MAE	MAPE
(3,1,2) + 4-14-1	14.45	8.75	8.6
(3,1,2) + 4-15-1	10.95	6.11	10
(3,1,2) + 4-16-1	14.08	7.43	9.3

Table 9: Actual vs forecasted maize production using Hybrid ARIMA-TDNN Model

Year	Actual Value	Forecasted Value
2020	4250.02	3574.72
2021	5362.13	4139.57
2022	5220.7	4606.4

Table 10: Comparison of forecasting accuracy of ARIMA, TDNN and Hybrid model ARIMA-TDNN on Training data

Criteria	ARIMA	TDNN	ARIMA-TDNN
Maize production			
RMSE	312.75	56.60	10.95
MAE	183.50	43.19	6.11
MAPE	17	20	10.1

Table 11: Comparison of Forecasting Accuracy of ARIMA, TDNN, and Hybrid Models on Test Data

Year	Actual	ARIMA	TDNN	ARIMA-TDNN
2020	4250.02	3579.2	3367.96	3574.72
2021	5362.13	3969.51	4046.82	4139.57
2022	5220.7	3842.53	4744.25	4606.4
RMSE	—	1197.14	957.29	880.92
MAE	—	1149.86	893.93	837.39
MAPE	—	22.7	18.1	15

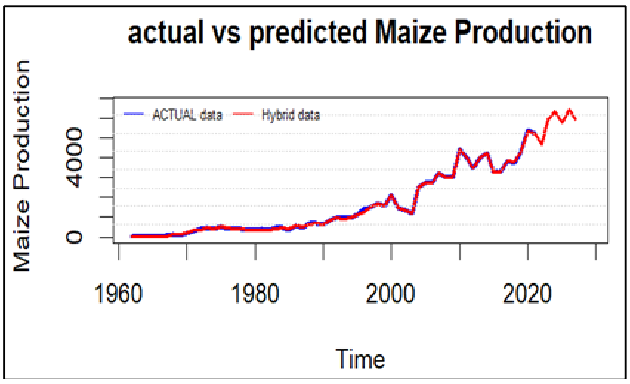

Fig 5: Actual vs predicted Maize production using the Hybrid ARIMA-TDNN model

These projections underscore a steady growth trajectory in maize cultivation, which can be attributed to technological advancements, improved market demand, and favourable agro-economic conditions in Karnataka state, India. The hybrid ARIMA-TDNN model proves to be an efficient and accurate tool for agricultural forecasting and can greatly support data-driven decision-making in policy formulation, resource allocation, and strategic agricultural planning.

Conclusion

The study effectively established a hybrid modelling approach combining ARIMA and Time Delay Neural Network (TDNN) for forecasting maize production in Karnataka state, India. While ARIMA (3, 1, 2) captured the linear trends and demonstrated a high level of accuracy, using the BDS test confirmed the presence of nonlinearity, prompting the application of TDNN (4-15-1) to model for nonlinear patterns. The final hybrid ARIMA-TDNN model significantly outperformed the standalone ARIMA and TDNN models, yielding the lowest error metrics. The hybrid model maintained superior accuracy with a low RMSE and low MAPE, confirming its robustness and reliability. The hybrid model forecasted an increasing trend in maize production for the upcoming years, i.e., from 2022 to 2027, with predicted values of 4668.55, 5925.44, 6302.09, 5784.41, 6423.36, and 5955.62, respectively.

References

1. Ali AM, Mushi J. Investigation of a Suitable Hybrid Time Series Model for Predicting Clove Price. Journal of ICT Systems. 2023;1(2):1-16.
2. Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: Forecasting and control. John Wiley & Sons; 2015.
3. Devra SJ, Patel DV, Shitap MS, Raj SR. Time series forecasting of price for oilseed crops by combining ARIMA and ANN. International Journal of Statistics and Applied Mathematics. 2023;8(4):40-54.
4. Kumar S, Arora K, Singh P, Gupta AK, Sharma I, Vatta K. Performance comparison of ARIMA and time delay neural network for forecasting of potato prices in India. Agricultural Economics Research Review. 2022;35(2).
5. Kumar TM, Prajneshu. Development of hybrid models for forecasting time-series data using nonlinear SVR enhanced by PSO. Journal of Statistical Theory and Practice. 2015;9(4):699-711.
6. Jha GK, Sinha K. Time-Delay Neural Networks for Time Series Prediction: An Application to the Monthly Wholesale Price of Oilseeds in India. Neural Comput Appl. 2014;24(3):563-571.
7. Rathod S, Mishra GC. Statistical models for forecasting mango and banana yield of Karnataka, India. Journal of Agricultural Science and Technology. 2018;20(4):803-816.
8. Rathod Santosha, Singh KN, Patil SG, Naik RH, Ray M, Meena VS. Modelling and forecasting of oilseed production of India through artificial intelligence techniques. Indian J Agric Sci. 2018;88(1):22-27.
9. Ray M, Rai A, Ramasubramanian V, Singh KN. ARIMA-WNN hybrid model for forecasting wheat yield time-series data. J Ind Soc Agric Stat. 2016;70(1):63-70.
10. Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing. 2003;50:159-175.