ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2024.0463

PRICE PREDICTION SYSTEM – A PREDICTIVE DATA ANALYTICS USING ARIMA MODEL

K. Vikranth[1], P.S. Nethravathi[2] and K. Krishna Prasad[3]
[1]*Institute of Computer Science and Information Science, Srinivas University, India*
[2,3]*Institute of Engineering and Technology, Srinivas University, India*

Abstract

In India, agriculture represents the primary occupation of more than 60% of the population. In terms of GDP, economic growth, traditional aspects, and social aspects, agriculture is essential for the country's development. The Indian farmers experienced numerous issues that have an impact on their way of life because the expansion in the agronomy business has not been as expected during the past two decades. Price fluctuation is one of the major issues faced by farmers, and as a result, they cannot get a reasonable price for their commodity. Also, it is very problematic to decide today without knowing the future price. So, this paper focused on finding a solution to the uncertainty problem in price faced by farmers that helps them take appropriate decisions during the farming process. The paper mainly concerns predictive data analytics using the ARIMA model, which predicts the price of areca nut products for the next 4 years using the past ten-year price dataset. The ARIMA model is a time series approach and a very appropriate framework for predicting future prices compared to other models. This paper includes a step-by-step procedure for the ARIMA techniques for forecasting price of agriculture commodity, and the outcomes are represented in the form of tables and graphical representations.

Keywords:

ARIMA, Price Prediction, Time Series Approach, Areca Nut, Smart Agriculture

1. INTRODUCTION

In India, the agricultural industry is anticipated to contribute more than 15% of GDP in 2023. Over 60% of the people in India depends on agriculture, making it a primarily agricultural nation. The agricultural sector, which lags the other sectors in every way, has made relatively little progress in terms of technological improvement. The exponential rise in global population drives a need for agricultural and food safety [1]. The world's demand for food will not be satisfied by the present traditional farming practices. Currently, there is a need for the agricultural industry to integrate new emerging technologies that will transform traditional agriculture into smart agriculture. Because it incorporates prediction and recommendation systems, data analytics is crucial in the agriculture sector [2]. IT agriculture, often referred to as smart agriculture, encompasses the utilization of technology and networks in the agricultural sector. This advanced approach, known as the smart agriculture system, integrates sensor expertise, automated control, digital network communication, data storage, and data analysis to provide effective solutions for various farming tasks [3]. Many technologies help to achieve a smart agricultural system in which data analytics plays a significant role, as discussed in the paper. The price prediction system needs a price dataset spanning more than ten years. Due to the nature of the dataset, employing a time series data analytics approach is not only suitable but also highly recommended. The Auto-Regressive Integrated Moving Average (ARIMA) model, a technique for analysing time series data, is utilized for the purpose of forecasting the price of areca nuts within the Puttur Taluk of Karnataka State's Dakshin Kannada district. The ARIMA model is a very suitable model to process time series datasets [4]. This paper demonstrates how the ARIMA model was developed for the price dataset. The research uses data analysis tools and packages to attain and compare the consequences from the proposed methods using the R programming language. The study uses different performance metrics and graphs to analyze the results. Simulation analysis is conducted on seven distinct categories of price data, including the minimum, maximum, and modal prices of the new variety of areca nut dataset. A detailed and comparative analysis is performed between observed values and predicted values for minimum, maximum, and modal prices for areca nuts of the new variety.

1.1 OBJECTIVES

- To offer a price-uncertainty solution that enables stakeholders to make the right choices.
- Predicting the minimum, maximum, and modal prices for areca nuts using the time series model known as the ARIMA model.
- To compare actual and predicted prices, which helps to interpret the forecasted result.

2. METHODOLOGY AND MATERIALS USED

The daily minimum price, maximum price, and modal price data are gathered from 2010 to 2022; the dataset's structure is shown in Table.1. The dataset was received from the government website under the Open Government Data (OGD) platform [5].

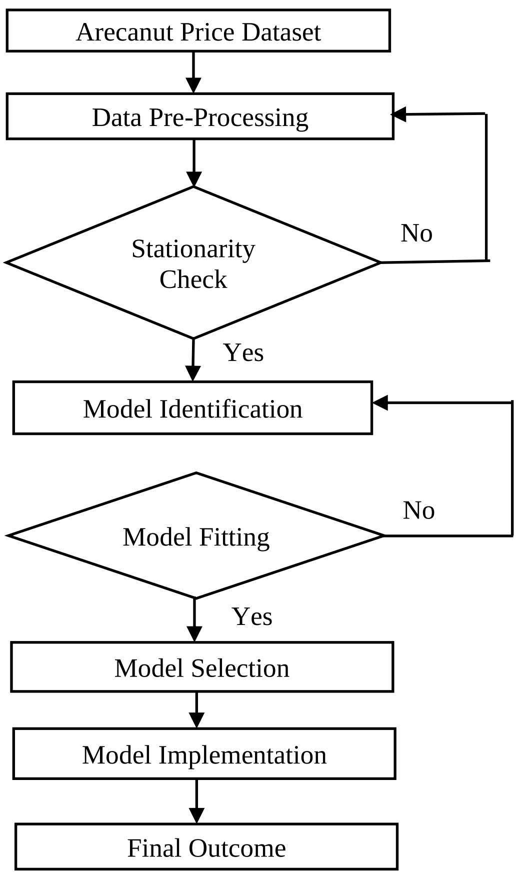
Table.1. Structure of the areca nut price dataset

Attributes	Data type	Description
State_Name	Text	Name of the state
District_Name	Text	Name of the district.
Market	Text	Market_Name within the district.
Commodity	string	Name of the agricultural product
Variety	string	Type of commodity.
Arrival_Date	string	Price date in dd/mm/yyyy format.
Min_Price	double	The minimum price of the commodity.
Max_Price	double	Maximum price of the commodity.
Modal_Price	double	The average price of the commodity.

Class	String	Very High, Very Low, High, Medium, Low

According to Table.1, the pricing dataset consists of nine parameters, five of which are categorical and four of which are numerical. The dataset only includes daily areca nut prices for Puttur Taluk in Dakshin Kannada District in the State of Karnataka [6]. There are five different kinds of areca nuts available, but this research emphases on type called new variety. In order to create an ARIMA model for forecasting, we, therefore, gather a dataset of a new variety of areca nuts.

When an autonomous variable in a linear regression model or multilinear regression model cannot be measured, observed, or has insufficient frequency, the recommendation is to opt for the univariate time sequence model, or ARIMA approach, which uses a systematic development process to forecast the time sequence data [7]. The ARIMA appraoch is a highly suitable, accurate, and popular method for forecasting time sequence data. The proposed research explores various approaches for predicting prices. As suggested by the name, it uses three different techniques: moving average, integration (differencing), and autoregression. Three parameters—*p*, *d*, and *q*—will be used in this component model. In the model, the moving average parameter (*q*) signifies the regression error, while the integrated parameter (*d*) indicates time series differentiation. The autoregressive parameter (*p*) specifies that the variable being modeled is regressed against its own lagged values. Before processing non-stationary data, the ARIMA Model converts it to stationary data. The ARIMA approach is widely recommended to predict linear time sequence data [8]. The below diagram defines the flowchart of a price prediction system using the ARIMA model.



As depicted in Fig.1, the pricing dataset, which was downloaded from the Open Government Data (OGD) website [6], contains 50,000 records of the Puttur taluk's daily areca nut price. The dataset has mislaid values that are filled using the linear regression method during the data pre-processing stage. Several data types are included in the database, all of which need to be further processed through data cleansing and data type validation before being made accessible for processing and analysis. The stationarity is checked in the following stage using the augmented Mickey Fuller test [9]. In this proposition trial, the null proposition considers the series as non-stationary, while the alternate hypothesis assumes the sequence is stationary. For assessing stationarity, the autocorrelation function (ACF) and Partial Autocorrelation Function (PACF) plots of the original series are employed.

Original series must be differentiated if they are not stationary in direction to make them stationary. Once the series is stationary, use the ARIMA model method in the following stage. The Fig.2 depicts the ARIMA model procedure.

The Fig.2 illustrates the three processes of the ARIMA methodology, which include identification, estimation, and diagnostic testing. To determine the p, d, and q values for the ARIMA approach is a stimulating task. The ACF and PACF correlograms of different series are employed to recognize the AR and MA process terms. It is recommended to explore alternative models, which are crucial in constructing the ARIMA model. After model formulation, all potential models are executed to determine important factors, change (instability), record-probability, AIC, BIC, and RMSE values. The resulting statistical values are organized in a table to guide the selection of the most suitable model. Additionally, deposits of the chosen ARIMA approach are estimated. Furthermore, the ACF and PACF correlograms of these residuals are plotted to perform a diagnostic check on the selected model [10].

Fig.1. Flow chart of the price prediction system

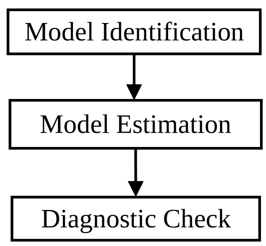


Fig.2. ARIMA model process

3. STATIONARITY AND DIFFERENCING

A time series with stationary characteristics remains unaffected by the time period in which it was observed. Time series displaying seasonality and patterns, however, are not considered stationary. Conversely, white noise within a series is stationary, maintaining a constant appearance across different time points. Additionally, a series exhibiting cyclic behaviour without trend or seasonality is also categorized as stationary.

To change from one state to another state, differencing is employed to stabilize the series' mean by eliminating trends and seasonality. The data after differencing represents the variance of successive observations in the main series. This can be mathematically expressed as:

$$\overline{yx}_t = yx_t - yx_{t-1} \quad (1)$$

As the data after difference only consists of the *t*-1 value, calculating the difference is not feasible.

$$\overline{yx}_1 = yx_1 - yx_{1-1} \quad (2)$$

When the data after difference is white noise, then it can be written as:

$$yx_t - yx_{t-1} = \varepsilon_t \quad (3)$$

Here ε_t represents the white noise component,

Reordering this equation results in a random walk prototypical for differencing:

$$yx_t = yx_{t-1} + \varepsilon_t \quad (4)$$

This model is mainly used for non-stationary data series. It allows to perform integer mean, in which case...

$$yx_t - yx_{t-1} = c + \varepsilon_t \quad (5)$$

$$yx_t = c + yx_{t-1} + \varepsilon_t \quad (6)$$

In this context, the parameter "*c*" represents the modal change of successive data values. If "*c*" is non negative, it signifies an increase in the value of y_t, resulting in an upward drift of yx_t. Conversely, if "c" is negative, it implies a downward drift of yx_t. [11].

3.1 SECOND ORDER DIFFERENCING

Second-order differencing refers to taking the difference between consecutive elements of a time series twice. It is a process used to remove trends and seasonality in a more aggressive manner than first-order differencing. It involves subtracting the previous value from the current one, and then deducting the previous value of the already differenced series from current differenced value. This can be useful when the data still exhibits some form of trend or seasonality after the first-order differencing. Second-order differencing can help to form the stationary in accurate, which is often required for certain statistical analyses and modelling techniques. It can be denoted as:

$$\overline{\overline{yx}}_t = \overline{yx}_t - \overline{\overline{yx}}_{t-1} \quad (7)$$

$$\overline{\overline{yx}}_t = \overline{yx}_t - \overline{yx}_{t-1} \quad (8)$$

$$= (yx_t - yx_{t-1})\ (yx_{t-1} - yx_{t-2}) \quad (9)$$

$$= yx_t - 2yx_{t-1} + 2yx_{t-2} \quad (10)$$

In this scenario, $\overline{\overline{yx}}_t$ consists of *t*-2 values. Therefore, it is recommended to apply differencing twice (second-order differencing) without necessarily going beyond this level [12].

3.2 SEASONAL DIFFERENCING

A seasonal difference is characterized by the disparity in a current value and a prior value within the same seasonal period. This distinction is articulated as:

$$\overline{yx}_t = yx_t - yx_{t-m} \quad (11)$$

Here, "*m*" represents the quantity of intervals, also referred to as lag-*m* changes, signifying the act of extracting observations at a delay of *m* intervals.

If the seasonal data has the characteristics of white noise, then the real series equation can be represented as:

$$yx_t = yx_{t-m} + \varepsilon_t \quad (12)$$

Predictions derived using this structure are equivalent to the most recent values from the corresponding season [13].

If $\overline{yx}_t = yx_t - yx_{t-m}$ represents a seasonally changing values, then the second differenced values is obtained by taking the difference between consecutive elements of the $\overline{y}_t$ series, which can be expressed as:

$$\overline{\overline{yx}}_t = \overline{yx}_t - \overline{yx}_{t-1}$$

$$= (yx_t - yx_{t-m}) - (yx_{t-m} - yx_{t-m-1})$$

$$= yx_t - yx_{t-1} - yx_{t-m} + yx_{t-m-1} \quad (13)$$

It makes no difference in the result if we apply both values, but it is strongly recommended to perform seasonal difference and then first difference.

3.3 BACKSHIFT (LAG) NOTATION

When working with time series lags, the backshift operator B or lag operator *L* is a notation. It is indicated as:

$$\mathrm{B}yx_t = yx_{t-1} \quad (14)$$

When this operator is applied to a time series of length "*yx*," the data is shifted back by one period [12]. Like this, applying *B* twice to a time series y_t causes the data to be shifted back to 2 periods. It is indicated as:

$$B(Byx_t) = \mathrm{B}^2 yx_t = yx_{t-2} \quad (15)$$

This operator is an appropriate tool for explaining the concept change from one state to another state. The first difference can be formulated below notation:

$$\overline{yx}_t = yx_t - yx_{t-1} = yx_t - \mathrm{B}yx_t = (1-B)\, yx_t \quad (16)$$

In this context, the first difference is symbolically expressed as 1-B, Same way, the second difference can be expressed as:

$$\overline{\overline{yx}}_t = yx_t - 2yx_t + yx_{t-2} = (1-2B+B^2)\, yx_t = (1-B)^2 yx_t \quad (17)$$

In a universal context, the nth-order equation can be expressed as: $(1-B)^n yx_t$

3.4 AUTOREGRESSIVE MODEL

An autoregressive model, often abbreviated as AR model, is a approach for data series with time component that predicts a future value based on its own past values. This approach, the current value of a object is assumed to be a direct mixture of its previous values, with some added noise.

Mathematically, an autoregressive model of order "*p*" (AR(*p*)) can be expressed as:

$$yx_t = c + \grave{\Omega}_1 * yx_{t-1} + \grave{\Omega}_2 * yx_{t-2} + \ldots + \grave{\Omega}_p * yx_{t-p} + \varepsilon_t \quad (18)$$

where:

yx_t is the present value of series at time t.

c is a persistent object.

$\grave{\Omega}_1, \grave{\Omega}_2, \ldots, \grave{\Omega}_p$ are the autoregressive factors.

$yx_{t-1}, yx_{t-2}, ..., yx_{t-p}$ are the past values of the time series up to order "*p*".

ε_t represents the white noise.

The model order "*p*" determines how many past values are considered for prediction. The coefficients $\phi_1, \phi_2, ..., \phi_p$ are estimated from the historical data. The larger the order "*p*", the more complex the model becomes, which might capture intricate temporal patterns but also makes it more prone to overfitting.

Autoregressive models are useful for capturing trends and dependencies in time series data, particularly when past values significantly influence future values.

3.5 MOVING AVERAGE MODEL

A Moving Average (MA) model the forecasting model of series data with time component that predicts a future value by means of direct mixture of previous errors (also known as the residual errors) and a white noise term. In contrast with the autoregressive model, which considers previous values of the variable being forecasted, the MA approach uses past forecast errors to make predictions.

Mathematically, a MA approach of order "q" (MA(q)) can be represented as:

$$yx_t = \mho + \varepsilon_t + \grave{\Omega}_1 * \varepsilon_{t-1} + \grave{\Omega}_2 * \varepsilon_{t-2} + ... + \grave{\Omega}_q * \varepsilon_{t-q} \quad (19)$$

where:

yx_t is the present value of series at time *t*.

$\mho$ is the series mean value.

ε_t represents the white noise.

$\grave{\Omega}_1, \grave{\Omega}_2, ..., \grave{\Omega}_q$ are the parameters representing the weights of the past forecast errors up to order "*q*".

The model order "*q*" determines how many past forecast errors are considered for prediction. The parameters $\grave{\Omega}_1, \grave{\Omega}_2, ..., \grave{\Omega}_q$ are estimated from the historical data. The larger the order "*q*", the more complex the model becomes, capturing dependencies in the error terms.

MA models are useful for removing random noise from time series data and can be effective when there is a correlation between forecast errors. Combining MA and autoregressive (AR) components leads to the Autoregressive Moving Average (ARMA) model, which is more flexible and capable of handling a wider range of time series patterns. [13].

3.6 NON-SEASONAL ARIMA MODEL

A Non-Seasonal ARIMA model is a technique to forecast data series with time component that combines autoregressive (AR), differencing (I), and moving average (MA) components to model and predict non-seasonal data series with time component.

The ARIMA approach is expressed as ARIMA (*p*, *d*, *q*), where:

"*p*" represents the order of the AR component, which identifies the link between current value of the series with old value.

"*d*" represents the order of order of change between current and old value, which helps in making the time series stationary by removing trends.

"*q*" represents the direction of MA component, which identifies the link between present and old forecasting errors.

A Non-Seasonal ARIMA model involves:

- AR component: Identifies the direct link between the current value and past values up to order "*p*".
- Differencing (I) factor: Expresses the rate of change in data series with time component to make stationary one, with the order of differencing denoted by "*d*".
- MA component: Explores the link between current and past errors during the forecasting of data series with time component up to order "*q*".

By combining these components, a Non-Seasonal ARIMA model can effectively capture and predict various non-seasonal patterns present in the data series with time component. It is a versatile and massy used approach in time series analysis and forecasting, often applied to economic, financial, and other data exhibiting non-seasonal trends and behaviors.

The mathematical representation of a Non-Seasonal ARIMA approach is expressed as:

$$\overline{\overline{yx}}_t = z + \pi_1 * \overline{yx}_{t-1} + \pi_2 * \overline{yx}_{t-2} + ... + \pi_p * \overline{yx}_{t-(p)} - \acute{\emptyset}_1 * \varepsilon_{t'-1} - \acute{\emptyset}_{12} * \varepsilon_{t'-2} - ... - \acute{\emptyset}_{1q} * \varepsilon_{t'-q} \quad (20)$$

where:

$\overline{\overline{yx}}_t$ is the second-order differenced series (after applying differencing "*d*" times).

z is a constant term.

$\pi_1, \pi_2, ..., \pi_p$ are the autoregressive coefficients, representing the relationship with past values up to order "*p*".

$\overline{y}x_{t-1}, \overline{y}x_{t-2}, \ldots, \overline{y}x_{t-(p)}$ are the past values of the second-order differenced series up to order "*p*".

$\acute{\emptyset}_{11}, \acute{\emptyset}_{12}, ..., \acute{\emptyset}_{1q}$ are the moving average coefficients, representing the relationship with past forecast errors up to order "*q*".

$\varepsilon_{t'-1}, \varepsilon_{t'-2}, ..., \varepsilon_{t'-q}$ are the past forecast errors up to order "*q*".

This equation represents how the current value of the second-order differenced series $\left(\overline{\overline{y}}_t\right)$ is modeled based on its own past values and past forecast errors. The goal of fitting an ARIMA model is to estimate the values of the parameters $\phi_1, \phi_2, ..., \phi_p, \theta_1, \theta_2, ..., \theta_q$, and the constant *c* from historical data in order to make accurate predictions for future values of the time series [14].

4. IMPLEMENTATION AND RESULT ANALYSIS

The modal price of areca nuts of the new variety contains data from January 2014 to December 2022. In the process of selecting a suitable model, it is imperative to initially assess the stationarity of the dataset containing modal prices. The Fig.3 below shows the year-wise time plot of the modal price data series.

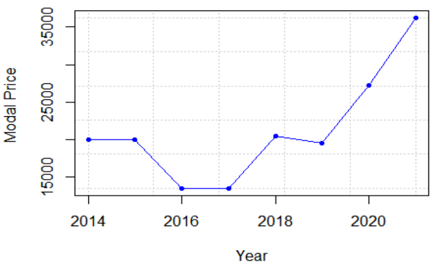
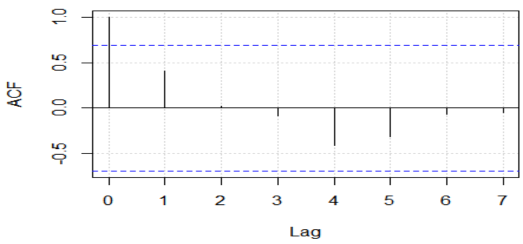
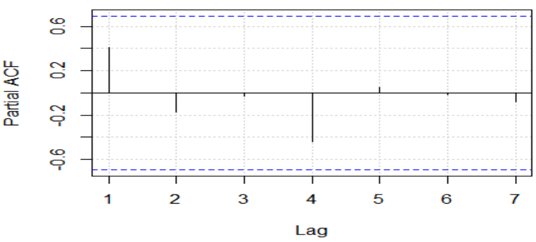


Fig.3. Time plot for minimum price data of type New Variety

The modal price data plot for the new variety indicates a clear rising and declining trend from 2014 to 2022. So, this series has a trend component and is non-stationary series. Furthermore, utilizing the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of the raw data is recommended to examine its stationarity [15]. The Fig.4 below shows the ACF and PACF of the modal price data for the new variety.



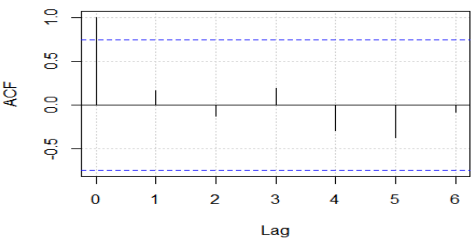
(a)



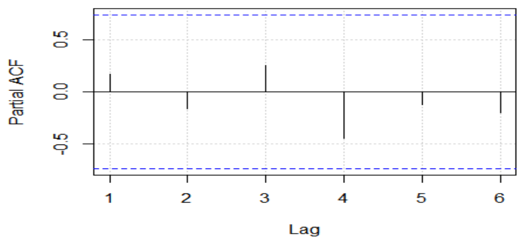
(b)

Fig.4. ACF (a) and PACF (b) of modal price data of type new variety

In the above figure, it is clearly indicated that ACF is gradually decreasing as the number lag increases, and one substantial spike that exceeds the standard error (SE) band shows the modal price data series of the new variety is not stationary [16]. It is a clear sign that we need to perform first differencing to convert the original series into stationary series [17]. Fig.5 illustrates the AutoCorrelation Function (ACF) and Partial AutoCorrelation Function (PACF) of the modal price data for the new variety, after the differencing process.



(a)



(b)

Fig.5. ACF and PACF of modal price data of type new variety after differencing

Based on the information derived from the AutoCorrelation Function (ACF) and Partial AutoCorrelation Function (PACF), the process of differencing the initial series aids in determining the order for the ARIMA model. In this scenario, the value of d is set to 2, signifying the need for the first difference to convert the original series into a stationary form [18]. When examining Fig.5, the ACF displays a diminishing trend at higher lags while presenting a noteworthy peak at lag 0. In relation to the PACF demonstrated in Fig.5, where the first lag of PACF is negative, the order for p is determined to be 0. Additionally, based on the declining pattern observed in the ACF, the order for q is determined to be 2 [19].

To attain more conclusive results and enhance the model's accuracy, various alternatives are taken into account. These alternatives encompass potential configurations such as ARIMA (0,0,0), ARIMA (1,1,0), ARIMA (1,1,1), ARIMA (1,0,1), ARIMA (0,0,1), ARIMA (1,0,0), ARIMA (2,2,2), ARIMA (0,2,2), and ARIMA (2,1,2). This comprehensive exploration aims to refine the model selection process.

The best ARIMA model can be selected for each areca nut's daily minimum, maximum, and modal prices of both CCQA and the new verify type. Multiple performance criteria like Akaike Information Criteria (AIC), Root Mean Square Error (RMSE), Mean Absolute Errors (MAE), Mean Percent Forecast Error (MPFE), Bayesian Information Criteria (BIC), and log-likelihood are used for selecting the best model [20].

To obtain more conclusive evidence and progress the exactness of the model, several alternatives to the ARIMA model are considered [21]. Based on the lowest AIC value, highest log-likelihood, largest significant coefficient, and lowest RMSE value, we compared different alternative models and presented them in Table 3.

Table.3. Comparison of alternative models with estimation criteria for modal price series of the type new variety

Alternative models ssmodelsddmodels /Estimation Criteria	**AIC**	**Log-likelihood**	**RMSE**
ARIMA (0, 0, 0)	168.28	-82.14	6962.792
ARIMA (1, 1, 0)	144.61	-70.3	5173.231
ARIMA (1, 1, 1)	145.65	-69.82	4427.231
ARIMA (1, 0, 1)	167.02	-79.51	4208.925
ARIMA (0, 0, 1)	165.61	-79.81	4534.819
ARIMA (1, 0, 0)	167.05	-80.52	5435.222

ARIMA (2, 2, 2)	126.83	-58.41	1994.187
ARIMA (0, 2, 2)	125.79	-59.89	3521.459
ARIMA (2, 1, 2)	149.63	-69.82	4441.165

In above Table.3, it is very clear that the ARIMA (0, 2, 2) model beats all other models with the lowest AIC value, highest log-likelihood value, and lowest RMSE value. So, all the criteria are in favor of ARIMA (0, 2, 2). So, we take ARIMA (0, 2, 2) for the diagnostic check. In addition, based on the Ljung-Box test, we accepted the null hypothesis that the residuals are white noise [22]. Thus, it can be identified that the suitable model based on ACF and PACF is ARIMA (0, 2, 2).

Once we have selected the suitable model for all types of data series, we proceed with the application of the model to generate the next 3 years' forecasted values using the respective models. The forecasted outcomes for the modal prices of the new variety of areca nut are presented both in tabular format and through graphical representation.

The year-wise predicted modal price of the areca nut of the new variety is calculated from 2016 to 2025. The first two years' dataset is considered training data, and actual prediction is done after the year 2016. Below Table.4 represents the actual and predicted prices for the modal price of areca nuts of the new variety, and the forecast is done based on the ARIMA (0, 2, 2) model for the next 4 years.

Table.4. Prediction of the areca nut modal price of the type new variety

Year	**Actual Price**	**Predicted Price**
2014	20000	
2015	20000	
2016	13500	16833
2017	13500	15667
2018	20500	18833
2019	19500	19833
2020	27250	25417
2021	36250	32667
2022	3554	36607
2023		34832
2024		35422
2025		35498
2026		36646

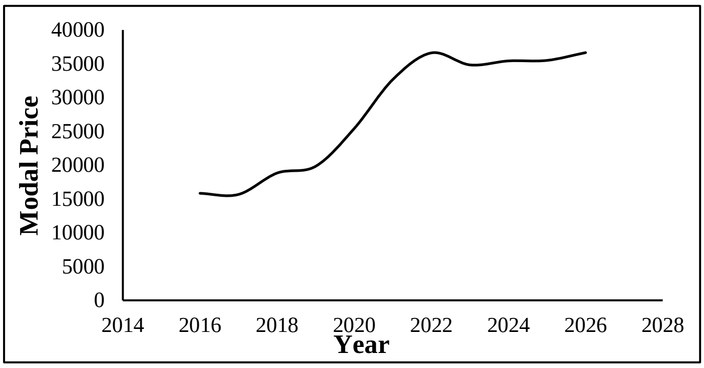


Fig.6. Forecasted result of the modal price of the type new variety

The Fig.6 shows the predicted areca nut modal price from 2016 to 2026. The predicted modal price of areca nut for the years 2023, 2024, 2025, and 2026 is represented graphically in Fig.6 and shown in Table.4.

The Fig.7 depicts the comparison of actual and predicted results of the modal price of areca nut for the new variety.

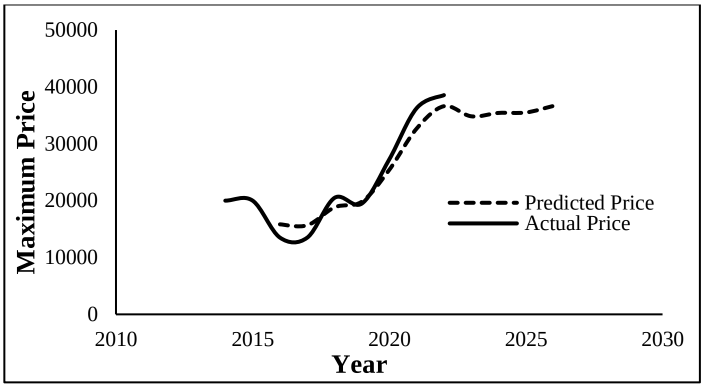


Fig.7. Actual price and predicted price of the areca nut using the ARIMA model

The Fig.7 depicted that from 2016 to 2022, the actual price and predicted price showed similar values for the ARIMA (0, 2, 2). The predicted modal price for the years 2023, 2024, 2025, and 2026 indicates that the future modal price of areca nut will increase slightly.

5. CONCLUSION

The agriculture sector is facing many issues that can be overcome by incorporating smart agricultural system. Data analytics theatres an important role in achieving a smart agricultural system. With the aid of a price prediction system, stakeholders in the areca nut product can make appropriate decisions that lead them to achieve profitability in business. The ARIMA model is a suitable model to achieve short-term prediction for time series datasets. This result will be accurate only in normal conditions, and sometimes price fluctuations will vary significantly under some external conditions. So, we can apply this model to predict the price of any commodity if a dataset is available.

REFERENCES

[1] K.M. Sabu and T.M. Kumar, "Predictive Analytics in Agriculture: Forecasting Prices of Areca Nuts in Kerala", *Procedia Computer Science*, Vol. 171, No. 1, pp. 699-708, 2020.

[2] S. Rajeswari and K. Suthendran, "Developing an Agricultural Product Price Prediction Model using HADT Algorithm", *International Journal of Engineering and Advanced Technology*, Vol. 9, No. 4, pp. 569-575, 2019.

[3] R. Dhanapal and J. Balaji, "Crop Price Prediction using Supervised Machine Learning Algorithms", *Journal of Physics: Conference Series*, Vol. 1916, No. 1, pp. 12042-12056, 2021.

[4] S. Shil, A.K. Sit and G.V. Thomas, "Forecasting of Areca Nut Market Price in North Eastern India: ARIM Modelling

Approach", *Journal of Planation Crops*, Vol. 41, No. 3, pp. 330-337, 2013.

[5] About Variety Wise Market Price, Available at https://data.gov.in/resources/variety-wise-daily-market-prices-areca nutbetelnutsupari-2020, Accessed in 2021.

[6] About Agricultural Commodity Price, Available at http://agriexchange.apeda.gov.in/india%20production/India_Productions.aspx?hscode=1092, Accessed in 2021.

[7] A.K. Kumar, R. Birau and M.E. Loredana, "Forecasting Areca Nut Market Prices using the Arima Model: A Case Study of India", *Annals of Constantin Brancusi University of Targu-Jiu. Economy Series*, Vol. 2, pp. 1-13, 2021.

[8] D.A. Mohan, "Big Data Analytics: Recent Achievements and New Challenges", *International Journal of Computer Applications Technology and Research*, Vol. 5, No. 7, pp. 460-464, 2016.

[9] A.K. Mahto, J. Ahmad and S.I. Alam, "Short-Term Forecasting of Agriculture Commodities in Context of Indian Market for Sustainable Agriculture by using the Artificial Neural Network", *Journal of Food Quality*, Vol. 2021, pp. 1-13, 2021.

[10] S.K. Purohit and S.K. Behera, "Time Series Forecasting of Price of Agricultural Products using Hybrid Methods", *Applied Artificial Intelligence*, Vol. 35, No. 15, pp. 1388-1406, 2021.

[11] About ARIMA Time Forecasting Model, Available at https://otexts.com/fpp2/arima.html, Accessed in 2023.

[12] About Mathematical Theory model of ARIMA, Available at https://people.duke.edu/~rnau/411arim.html, Accessed in 2023.

[13] About time series data analytics, Available at https://www.analyticsvidhya.com/blog/2021/11/performing-time-series-analysis-using-arima-model-in-r/, Accessed in 2023.

[14] Z. Chen and X.Y. Liew, "Automated Agriculture Commodity Price Prediction System with Machine Learning Techniques, *Proceedings of International Conference on Machine and Deep Learning*, pp. 1-14, 2021.

[15] A. Kamei, "Forecasting of Areca Nut (Areca Catechu) Yield using Arima Model for Uttara Kannada District of Karnataka", *International Journal of Science and Research*, Vol. 3, No. 7, pp. 1002-1009, 2014.

[16] L. Narsimhaiah and P. Pandit, "Modelling and Forecasting of Areca Nut Production in India Vision", *International Journal of Current Microbiology and Applied Sciences*, Vol. 8, No. 11, pp. 728-738, 2020.

[17] S.A. Mulla and S.A. Quadri, "Crop-Yield and Price Forecasting using Machine Learning", *International Journal of Analytical and Experimental Modal Analysis*, Vol. 12, pp. 1731-1737, 2020.

[18] P. Samuel and N.A. Kumar, "Crop Price Prediction System using Machine Learning Algorithms", *Quest Journal of Software Engineering and Simulation*, Vol. 23, pp. 1-13, 2020.

[19] P.S. Rachana, N. Shruthi and R.S. Kousar, "Crop Price Forecasting System using Supervised Machine Learning Algorithms", *International Research Journal of Engineering and Technology*, Vol. 6, pp. 4805-4807, 2019.

[20] X. Pham and M. Stack, "How Data Analytics is Transforming Agriculture", *Business Horizons*, Voll. 61, No. 1, pp. 125-133, 2018.

[21] J.A. Brandt and D.A. Bessler, "Price Forecasting and Evaluation: An Application in Agriculture", *Journal of Forecasting*, Vol. 2, No. 3, pp. 237-248, 1983.

[22] M. Kaur and H. Kundra, "Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications", *International Journal of Computer Applications*, Vol. 99, No. 12, pp. 1-3, 2014.