A Survey of Data Analytical Techniques on Commodity Price Prediction for Smart Agricultural System

K. Vikranth [1, 2] & Nethravathi P. S. [3] & Krishna Prasad K. [4]

[1] Research Scholar, Institute of Computer Science and Information science, Srinivas University, Mangaluru, Karnataka, India,

[2] Assistant Professor, Institute of Computer Science & Information Science, Srinivas University, Mangaluru, Karnataka, India

OrcidID: 0000-0001-9549-1743; E-mail ID: vikranth.kadya@gmail.com

[3] Professor, Institute of Computer Science & Information Science, Srinivas University, Mangaluru, Karnataka, India,

OrcidID: 0000-0001-5447-8673; E-mail ID: nethrakumar590@gmail.com

[4] Professor, College of Computer Science & Information Science, Srinivas University, Mangaluru, Karnataka, India,

OrcidID: 0000-0001-5282-9038; E-mail ID: krishnaprasadkcci@srinivasuniversity.edu.in

Area of the Paper: Computer Science.
Type of the Paper: Conceptual Research.
Type of Review: Peer Reviewed as per |C|O|P|E| guidance.
Indexed In: OpenAIRE.
DOI: https://doi.org/10.5281/zenodo.7066670
Google Scholar Citation: IJCSBE

How to Cite this Paper:

Vikranth, K., Nethravathi, P. S., & Krishna Prasad, K., (2022). A Survey of Data Analytical Techniques on Commodity Price Prediction for Smart Agricultural System. *International Journal of Case Studies in Business, IT, and Education (IJCSBE)*, *6*(2), 259-267. DOI: https://doi.org/10.5281/zenodo.7066670

International Journal of Case Studies in Business, IT and Education (IJCSBE)
A Refereed International Journal of Srinivas University, India.

Crossref DOI : https://doi.org/10.47992/IJCSBE.2581.6942.0194

Paper Submission: 10/08/2022
Paper Publication: 31/08/2022

International Journal of Case Studies in Business, IT, and Education (IJCSBE), ISSN: 2581-6942, Vol. 6, No. 2, August 2022

SRINIVAS PUBLICATION

A Survey of Data Analytical Techniques on Commodity Price Prediction for Smart Agricultural System

K. Vikranth [1,2] & Nethravathi P. S. [3] & Krishna Prasad K. [4]

[1] Research Scholar, Institute of Computer Science and Information science, Srinivas University, Mangaluru, Karnataka, India,

[2] Assistant Professor, Institute of Computer Science & Information Science, Srinivas University, Mangaluru, Karnataka, India

OrcidID: 0000-0001-9549-1743; E-mail ID: vikranth.kadya@gmail.com

[3] Professor, Institute of Computer Science & Information Science, Srinivas University, Mangaluru, Karnataka, India,

OrcidID: 0000-0001-5447-8673; E-mail ID: nethrakumar590@gmail.com

[4] Professor, College of Computer Science & Information Science, Srinivas University, Mangaluru, Karnataka, India,

OrcidID: 0000-0001-5282-9038; E-mail ID: krishnaprasadkcci@srinivasuniversity.edu.in

ABSTRACT

Purpose: *Food or agricultural products are one of the most basic needs of people. The population of India and the rest of the world is growing at an exponential rate, as is the demand for food commodities. As a result, there should be a proper and convenient way to increase food production, as well as the introduction of efficient technologies in all aspects of the agriculture sector. Commodity prices are an important factor in agriculture because they determine the former's economic status and wealth. The farmer's income and profit are determined by the current and future price of the commodity. Farmers are losing a lot of money because they don't know what the price of their products will be in the future. As a result, there should be a proper approach that provides future information about agricultural products, allowing farmers to make decisions ahead of time before cultivating any product.*

Design/Methodology/Approach: *Developing a theoretical concept based on model building using the secondary sources and focus group interaction method and analysis of the model using the ABCD listing framework.*

Findings/Result: *A method known as price prediction uses historical and current data from a database to estimate future agricultural commodity prices.* **.** *This paper was primarily concerned with identifying the appropriate data analysis techniques for implementing price prediction systems, particularly for agricultural products. Also conducts a survey on various predictive analytics approaches related to agricultural datasets. Finally, we used our own suggested model to implement a price prediction system with the help of a smart agricultural system.*

Paper Type: *Conceptual Research.*

Keywords: Data Analytics, Smart Agricultural System, Price Prediction, Agricultural Dataset.

1. INTRODUCTION :

Agriculture is the pillar of a country's economy and prosperity. Agricultural product price fluctuations have an impact on a country's GDP. Before farming a certain crop type, crop price estimation and appraisal are undertaken to help farmers make educated decisions [1]. Making smarter judgments, minimizing losses, and managing the risk of price swings will all be made easier with the help of crop price forecasting. Estimating the amount of profit they can expect from their chosen crop is a major problem for farmers [2]. In recent years, price forecasts have been made by analyzing farmers' experiences with a particular crop and field. Today, price prediction is a crucial agricultural issue that can only be resolved with the help of the data at hand [3]. To assure farmers' incomes, our research strives to more successfully resolve the issue of crop price forecast. Because of the use of contemporary

technologies, the growing population, and increased productivity, it is anticipated that the agricultural analytics industry would expand [4]. This work is based on the discovery of suitable data models that aid in the achievement of high accuracy and generality in price prediction. A contemporary research area that is being used to examine agricultural crop prices is big data predictive analytics [5].
Arecanut is a significant commercial tree crop that is mostly grown in the Indian states of Karnataka, Tamil Nadu, Kerala, Assam, West Bengal, and Meghalaya [6]. Arecanut is one of the crucial plantation crops of the coastal region, particularly in Karnataka and Kerala, and it also significantly affects people's quality of life. According to the Indian Horticulture Database 2010 (National Horticulture Board, 2010), India produces 478 thousand metric tons (MT) of processed nuts (Chahal from the areca nut, which takes up around 400 thousand hectares of land. Areca nut consumption and production are dominated by India, which accounts for more than half of global output [7]. Karnataka accounts for more than 60 percent of the total areca nut production [8]. The growers' biggest problem is the fluctuating price of these commodities. Despite the existence of numerous online platforms like commodities online that provide information on the pricing of areca nuts from different regions and taluks, The use of data analytics to forecast areca nut prices in Karnataka has not received much attention from researchers or practitioners [9]. The price prediction algorithm takes into account the agricultural crop price statistics from the Dakshin Kannada District in Karnataka, India. In this work, different data analytical technique has to be surveyed especially to forecast the price of areca nuts in the markets of Karnataka [10]. An efficient forecasting model would enable the farmers and traders in Meghalaya and Assam to generate a sizable income. The ARIMA approach is the most popular strategy for univariate time series forecasting [11]. The different approaches such as SARIMA, Box-Jenkins autoregressive ARIMA time-series methodology, Holt-Winter's Seasonal method, and LSTM neural network, Hybrid Association rule-based Decision Tree algorithm (HADT), and decision tree regressor (Supervised machine learning algorithm) and to determine which of these models suited the data the best, performance was assessed. The government can plan agricultural growth to stabilize the pricing of relevant products by employing these predictive models [12].

2. RELATED WORKS :

To predict commodity prices, many time series techniques have been developed over the previous ten years. The kind and amount of the data set have a significant impact on the validity and reliability of these techniques. In the past ten years, a number of techniques have gained popularity, including ARIMA (Auto-Regressive and Integrated Moving Average), Holt-Winters Seasonal Method, Artificial Neural Network (ANN), Hybrid Association Rule-Based Decision Tree Algorithm (HADT), Decision Tree Regressor, and others. Table 1 shows the most popular price prediction methodologies utilized and their benefits by various writers over the previous ten years. Depending on the type of commodity and the nature of the dataset, authors will utilize one of these approaches.

Table 1: Approaches used by different authors on the agriculture dataset

S. No.	Author	Approach Used	Benefits
1	Sabu, K. M., & Kumar, T. M. (2020). [1]	LSTM neural network, SARIMA, and Holt Winter's seasonal approach	The most efficient model is the LSTM neural network. A lot of data is required for the LSTM algorithm, a deep learning system.
2	Rajeswari, S., & Suthendran, K. (2019). [18]	Algorithm for Decision Trees based on Hybrid Association Rules (HADT)	In comparison to other current decision tree models, this model is more encouraging and precise in predicting agricultural product prices.
3	Dhanapal et al. (2021) [5].	Decision tree regressor (Supervised machine learning algorithm)	Crop price forecasting and crop price prediction for the next 12 months
4	Sandip Shil et al. (2013). [7]	ARIMA (1, 0, 1), ARIMA (1, 1, 1), ARIMA (0, 1, 1), log	ARIMA approach should only be used for short-term forecasting.

		ARIMA (0, 1, 1), log ARIMA (1, 0, 1)	
5	Kumar et al (2021). [10]	Box Jenkins ARIMA methodology	The ARIMA (3, 1, 3) model is thought to be a good fit for forecasting the price of a new variety of Arecanuts in Karnataka.
6	Sadiq A. Mulla & S. A. Quadri (2020). [20]	Decision tree algorithm	Predict the price and gain over the next twelve months based on the last twelve months' data.
7	Kurumatani, K. (2020). [21]	Direct upcoming time sequences forecasting, time-alignment of period point forecasting, and recurrent neural network (RNN) forecasting (DFTS).	TATP of LTSM outperforms TATP of other RNNs not only in terms of accuracy but also in terms of conservation.
8	Majumdar et al. (2017). [6]	PAM, CLARA, and DB, SCAN clustering methods.	DBSCAN outperforms PAM and CLARA in terms of clustering quality; CLARA outperforms PAM in terms of clustering quality.

3. OBJECTIVES :

(1) Determine which data sets are required to meet the analytical requirements.
(2) Examine the use of data analytics in many aspects of agriculture.
(3) Understanding the significance of data analytics in agriculture.
(4) To achieve a smart agricultural system.
(5) Using data analytic tools to forecast commodity prices.
(6) Implementation of a smart agricultural system based on price prediction using the best technique available.

4. METHODOLOGY :

To collect the data and information for this research, we employed a literature review and an online approach. All of the material and data obtained came from secondary sources such as websites, research papers, books, journals, and magazines. Following the evaluation process, we are able to identify the required analytical method that was appropriate for the suggested model, and we developed our model to predict commodity prices. The qualitative technique is applied in this study, which is concerned with attitudes, opinions, and behavior. The generated result is also in a non-quantitative format. As a result, the research is based on my research insight and in-depth understanding.

5. AGRICULTURE DATASET AND DIMENSIONS :

The information for this study was gathered from the open government data (OGD) Platform India's official website [13]. From January 2004 to December 2021, in Dakshin Kannada district a new variety areca nutst's monthly price information was downloaded. Puttur, bantwala, sullia, belthangadi, kadaba, and moodabodri are the seven taluks that make up the Dakshin Kannada district. Areca nuts are available in five different types [14]. Because the majority of farmers lack proper storing facilities, the new variety is the most well-known of those five varieties. As a result, we've created a price series for a new variety of areca nut. The dataset has nine attributes, five of which are categorical attributes and four of which are numerical attributes. This information covers the minimum, maximum, and modal prices of Arecanut, broken down by market, district, and state. This dataset covers day-by-day arecanut prices for the years 2014 to 2021. The dataset has 50000 records and ten attributes. Table 2 shows the structure of the dataset.

Table 2: The structure the of areca nut price dataset [Author]

Attributes	Data type	Description
State	string	Name of the state
District	string	Name of the district within state
Market	string	Name of the market within district.
Commodity	string	Name of the agricultural product
Variety	string	Type of commodity.
Arrival_Date	string	Price date in dd/mm/yyyy format.
Min_x0020_Price	double	Minimum price of the commodity.
Max_x0020_Price	double	Maximum price of the commodity.
Modal_x0020_Price	double	The average price of the commodity.
Class	String	Very High, Very Low, High, Medium, Low

6. VARIOUS PRICE PREDICTION APPROACHES :

A survey offers a variety of strategies for forecasting agricultural commodity prices, including statistical and intelligent methods. No single technique is suitable for all agricultural commodities, according to the "no free lunch theory" [15]. Some of the prediction methodologies used for forecasting areca nut commodities are included in this literature.

6.1 Model for Time Series:

In agricultural markets, period series forecasting is critical. Good's price instability brands it tough to make timely pronouncements based on intuition. Period sequence models like SARIMA and Holt-Winter seasonal models are secondhand to forecast the price of areca nuts. Statistical techniques like ARIMA and Holt-seasonal Winter's approach are frequently used to forecast data containing trends and seasonality [16].

6.1.1 ARIMA (Auto-Regressive Integrated Moving Average) approach:

This is a type of Box Jenkins model that can be utilized to predict or analyze time-series data, whether they are stationary or not. AR, lag-difference (integrated), and MA are the three components that makeup ARIMA. These three components are represented as ARIMA (p, d, q), where p is the numeral of AR relations, d is the direction of delay modification, and q is the numeral of MA relationships. The lagged value will serve as the predictor variable because the predicted value of AR will be dependent on it. While the integrated value is utilized to make the period-series stationery. For constraint assessment and prototypical choice on non-stationary period sequence, the interval change is used. The predictor value is MA, and the forecaster standards are past lagged fault values [17].

The following estimations can be used to choose the right model: the greatest quantity of significant coefficients, the least amount of volatility, the greatest amount of log-likelihood, the least amount of AIC and BIC, and the greatest amount of significant coefficients. Model (3, 1) is superior to all other models since it has the lowest volatility and the most significant coefficients. The model is also favored by the final two criteria, which have the highest log-likelihood of statistics and lowest AI favors [18].

6.1.2 Holt-Winter's Seasonal Method:

The Holt-Winter Periodic Technique is one of the Exponential Flattening alternates used for cyclic data. It involves an estimate calculation and three flattening equations, each with the proper smoothing parameters, one for the level lt, one for the trend bt, and one for the seasonal component st. The symbol m stands for seasonality frequency or the number of periods in a year. This method comes in two variations, each with a distinct seasonal component. The multiplicative method is recommended when seasonal variations vary depending on the level of the series; the additive method is favored when seasonal variations are generally constant across the series [19].

6.2 Artificial Neural Network:

The system can learn from the fluctuations and provide improved forecasting results thanks to ANN. The cavernous education model-LSTM was the contraption education process used for predicting. The LSTM model was used to forecast prices with seasonality. Due to LSTM's ability to take into account composite progressive dependencies, the archetypal was able to suitable both fixed and non-fixed data. Their outcomes were equated using the RMSE value [20]. The LSTM model was determined to be the most economical one for fitting the data and predicting the price of areca nuts. The only drawback will be a shortage of training data because LSTM is a cavernous education technique that needs a proportion of data. An embedded model, an inherited algorithm-constructed neural network model, an RBF neural network model, and a BP neural network model are all used for prediction. The observation results show that the BP neural network model provided the worst results of the four. The RBF model is less flawless than the genetic algorithm-based model [21]. For market price estimation, a fuzzy neural network technique is used. A new hyper cubic training technique has enabled the inter-layer and feed-forward design of this fuzzy neural network. On an hourly basis, it calculates market clearing rates. The results show that this technique beats other neural networks including ARIMA time series, wavelet-ARIMA, MLP, and RBF [22].

6.3 Machine Learning Algorithms:

6.3.1 Hybrid Association rule-based Decision Tree algorithm (HADT):

When it comes to forecasting the pricing of agricultural products, the HADT forecast model is more promising and accurate than other modern decision tree models. The findings of the monthly price forecast for agricultural items are generated using the HADT algorithm. It gives the farming community more authority over their decisions, enabling them to thrive. Focused on creating association rules that only supply one attribute, the Hybrid Association Decision Tree approach (HADT) [23].
The following are the benefits of the Hybrid Association rule-based Decision Tree algorithm (HADT):

- Aside from the size of the data set, training the data is extremely capable.
- High-dimensional training sets can be efficiently handled without making assumptions about dependent and free qualities [24].
- The classification process was swiftly completed.
- The generated rules are simple to understand for humans.
- In order to anticipate trial information and assess the accuracy of the classifier model, classification techniques are used. [25].

6.3.2 Decision tree regressor:

The system uses the decision tree regressor (supervised machine learning algorithm) to analyze historical data, anticipate prices for the upcoming twelve months, and forecast prices for recent data. Employs the decision tree method to efficiently forecast the results and proved to be the best choice for the study project. The information gathered is evaluated and cleansed in order to forecast crop prices [26].

7. IMPLEMENTATION :

The data module, data pre-processing module, data analysis module, data processing (price prediction algorithm) module, and user interface module make up the architecture of the crop price prediction system. This data module offers the necessary database for forecasting prices for the following twelve months [27]. The database contains a variety of data kinds that must be processed further through data cleaning and data type validation before being made available for further processing and analysis. Then, using a suitable analytical technique, predict the commodity's price based on the available data. The system's ultimate result will be displayed in a user interface that is convenient for the end user. The figure 1 depicts the crop prediction system's step-by-step method [28].

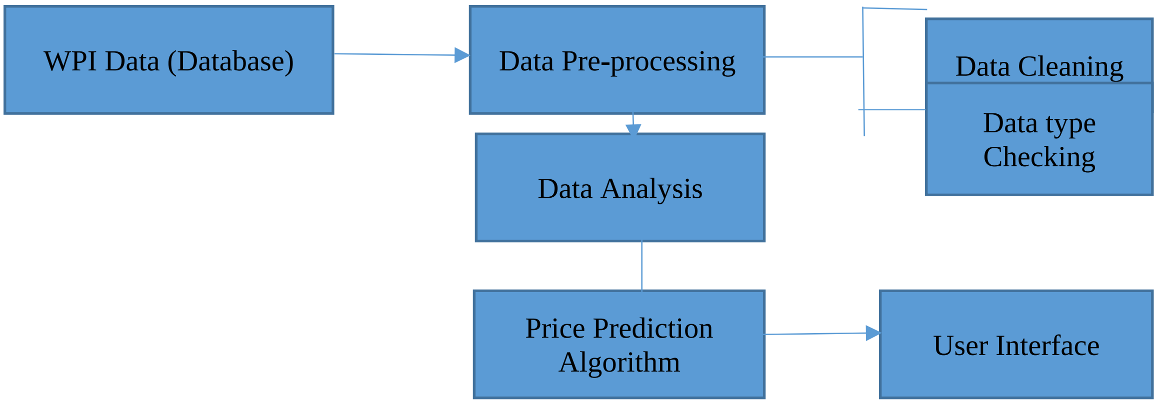

Fig 1: Crop Price Prediction System.

8. CONCLUSION :

Agriculture is the backbone of the country and more than seventy percent of the people depend on agriculture as their main profession. The main problem faced by agriculturists is uncertainty about ads yield and price. These problems can be easily solved using available data and data analytical techniques. This paper mainly focused to survey different available data analytical techniques for price prediction and also surveyed different research papers that work on to predict prices of different commodities. These data analytical techniques will help to force the price of the commodity for the next twelve months and are very useful to farmers to take appropriate decisions regarding agriculture activity. Both the public sector and the industrial sector may find this kind of market price forecasting useful in predicting future policy choices.

REFERENCES :

[1] Sabu, K. M., & Kumar, T. M. (2020). Predictive analytics in Agriculture: Forecasting prices of Arecanuts in Kerala. *Procedia Computer Science*, *171(1)*, 699-708. Google Scholar CrossRef

[2] Darekar, A., & Reddy, A. (2017). Forecasting oilseeds prices in India: Case of groundnut. *Forecasting Oilseeds Prices in India: Case of Groundnut (December 14, 2017). J. Oilseeds Res*, *34*(4), 235-240. Google Scholar CrossRef

[3] Aithal, P. S., & Aithal, S. (2015). Ideal technology concept & its realization opportunity using nanotechnology. *International Journal of Application or Innovation in Engineering & Management (IJAIEM)*, *4*(2), 153-164. Google Scholar CrossRef

[4] Vikranth, K. (2021). An Implementation of IoT and Data Analytics in Smart Agricultural System–A Systematic Literature Review. *International Journal of Management, Technology and Social Sciences (IJMTS)*, *6*(1), 41-70. Google Scholar CrossRef

[5] Dhanapal, R., AjanRaj, A., Balavinayagapragathish, S., & Balaji, J. (2021, May). Crop price prediction using supervised machine learning algorithms. In *Journal of Physics: Conference Series*, *1916*(1), 012042-012056. Google Scholar CrossRef

[6] Majumdar, J., Naraseeyappa, S., & Ankalaki, S. (2017). Analysis of agriculture data using data mining techniques: application of big data. *Journal of Big data*, *4*(1), 1-15. Google Scholar

[7] Shil, S., Acharya, G. C., Jose, C. T., Muralidharan, K., Sit, A. K., & Thomas, G. V. (2013). Forecasting of areca nut market price in north eastern India: ARIM modeling approach. *Journal of Plantation Crops*, *41*(3), 330-337. Google Scholar

[8] Hegde, S. A., & Deal, J. (2014). Areca nut farming in southern India: A case study. *International Journal of business and social science*, *5*(10), 6-17. Google Scholar

[9] Aithal, P. S., & Aithal, S. (2018). Study of various general-purpose technologies and their comparison towards developing sustainable society. *International Journal of Management, Technology, and Social Sciences (IJMTS)*, *3*(2), 16-33. Google Scholar CrossRef

[10] Kumar, A. K., Pinto, P., Hawaldar, I. T., Spulbar, C., Birau, R., & Loredana, M. E. (2021). Forecasting Areca Nut Market Prices Using the Arima Model: A Case Study Of India. *Annals-Economy Series*, *2(1)*, 4-18. Google Scholar CrossRef

[11] Luo, C., Wei, Q., Zhou, L., Zhang, J., & Sun, S. (2010, October). Prediction of vegetable price based on Neural Network and Genetic Algorithm. In *International Conference on Computer and Computing Technologies in Agriculture* (pp. 672-681). Springer, Berlin, Heidelberg. Google Scholar CrossRef

[12] Amjady, N., & Keynia, F. (2010). Electricity market price spike analysis by a hybrid data model and feature selection technique. *Electric Power Systems Research*, *80*(3), 318-327. Google Scholar CrossRef

[13] data.gov.in (2017). 11 About variety wise market price [online] Available at: https://data.gov.in/resources/variety-wise-daily-market-prices-arecanutbetelnutsupari-2020[Accessed 30 Mar. 2021].

[14] agriexchange.apeda.gov.in (2017). 11 About Agricultural commodity price [online] Available at: http://agriexchange.apeda.gov.in/india%20production/India_Productions.aspx?hscode=1092 [Accessed 30 Mar. 2021].

[15] Junrui, Y., Lisha, X., & Hongde, H. (2012, August). A classification algorithm based on association rule mining. In *2012 International Conference on Computer Science and Service System* (pp. 2056-2059). IEEE. Google Scholar CrossRef

[16] Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, *1*(1), 67-82. Google Scholar CrossRef

[17] Zhang, D., Chen, S., Liwen, L., & Xia, Q. (2020). Forecasting agricultural commodity prices using model selection time-series with time series features and forecast horizons. *IEEE Access*, *8*(1), 28197-28209. Google Scholar CrossRef

[18] Rajeswari, S., & Suthendran, K. (2019). Developing an Agricultural Product Price Prediction Model using HADT Algorithm. *International Journal of Engineering and Advanced Technology*, *9*(4), 569-575. Google Scholar CrossRef

[19] Raghavan, V., & Baruah, H. K. (1958). Arecanut: India's popular mastichistory—history, chemistry, and utilization. *Economic Botany*, *12*(4), 315-345. Google Scholar CrossRef

[20] Mulla&, S. A., & Quadri, S. A. Crop-yield and Price Forecasting using Machine Learning. *International journal of analytical and experimental modal analysis*, *12*(8), 1731-1737. Google Scholar

[21] Kurumatani, K. (2020). Time series forecasting of agricultural product prices based on recurrent neural networks and its evaluation method. *SN Applied Sciences*, *2*(8), 1-17. Google Scholar CrossRef

[22] Palanivel, K., & Surianarayanan, C. (2019). An approach for prediction of crop yield using machine learning and big data techniques. *International Journal of Computer Engineering and Technology*, *10*(3), 110-118. Google Scholar↗ CrossRef↗

[23] Vijayabaskar, P. S., Sreemathi, R., & Keertanaa, E. (2017, March). Crop prediction using predictive analytics. In *2017 International Conference on Computation of Power, Energy Inf Communication Communcation (ICCPEIC)* (pp. 370-373). IEEE. Google Scholar↗ CrossRef↗

[24] Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R. L., & Mouazen, A. M. (2016). Wheat yield prediction using machine learning and advanced sensing techniques. *Computers and electronics in agriculture*, *121*(1), 57-65. Google Scholar↗ CrossRef↗

[25] Darekar, A., & Reddy, A. (2017). Forecasting of common paddy prices in India. *Journal of Rice Research*, *10*(1), 71-75. Google Scholar↗ CrossRef↗

[26] Verma, V. K., Kumar, P., Singh, S. P., & Singh, H. P. (2016). Use of ARIMA modeling in forecasting coriander prices for Rajasthan. *International Journal of Seed Spices*, *6*(2), 40-45. Google Scholar↗

[27] Rather, A. M., Agarwal, A., & Sastry, V. N. (2015). Recurrent neural network and a hybrid model for prediction of stock returns. *Expert Systems with Applications*, *42*(6), 3234-3241. Google Scholar↗ CrossRef↗

[28] Co, H. C., & Boosarawongse, R. (2007). Forecasting Thailand's rice export: Statistical techniques vs. artificial neural networks. *Computers & industrial engineering*, *53*(4), 610-627. Google Scholar↗ CrossRef↗

****