



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 4, April 2023

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.423

 9940 572 462

 6381 907 438

 [ijirset@gmail.com](mailto:ijirset@gmail.com)

 [www.ijirset.com](http://www.ijirset.com)



# Cotton Price Prediction Using ARIMA Model

Prof. Anushree Deshmukh<sup>1</sup>, Dhiraj Kumar Shastri<sup>2</sup>, Abhishek Pal<sup>3</sup>, Harsh Palan<sup>4</sup>, Tejas Kharat<sup>5</sup>

Assistant Professor, Department of Information Technology, Rajiv Gandhi Institute of Technology (Mumbai University), Mumbai, Maharashtra, India<sup>1</sup>

U.G. Student, Department of Information Technology, Rajiv Gandhi Institute of Technology (Mumbai University), Mumbai, Maharashtra, India<sup>2,3,4,5</sup>

**ABSTRACT:**The cotton market is a global market that involves the buying and selling of cotton, cotton is a fibrous material that has a soft and fluffy texture, and it grows within a protective capsule called a boll around the seeds of cotton plants. Cotton is a widely used crop, and it is used in the production of clothing, textiles, and various other products. There are various factors that can impact the prices of cotton, such as global supply and demand, weather patterns, government regulations, and geopolitical circumstances. In recent years, cotton prices have fluctuated significantly due to a range of factors. In this research work, the cotton prices are predicted using the Auto Regressive Integrated Moving Average (ARIMA) Model. Cotton price predictive models have been developed and run on published cotton data acquired from AGMARKNET. Based on the experimental findings, it can be inferred that the ARIMA model is capable of providing reasonably accurate predictions for cotton prices over a short-term period.

**KEYWORDS:**ARIMA Model, Cotton Price Analysis, Time Series Forecasting, Machine Learning, Predictive Analysis.

## I. INTRODUCTION

India holds a significant position in the cotton industry as it is one of the leading producers and exporters of cotton worldwide. Cotton is a crucial crop for India, both in terms of production and export. After China, India is the second-largest producer of cotton globally, and it is also the largest exporter of cotton. Cotton is grown in several states across India, including Gujarat, Maharashtra, Telangana, Punjab, Haryana, Telangana, and Rajasthan.

Cotton is a vital crop for the Indian economy, as it provides livelihoods to millions of farmers and farm laborers. The cotton sector also has significant linkages with the textile industry, which is a major contributor to India's GDP and exports. The Indian government has taken several steps to support the cotton sector, including providing subsidies and price support to farmers, improving access to credit, and investing in research and development to improve crop yields and quality in the recent years.

However, the cotton sector in India also faces several challenges, including climate change, pest attacks, and low productivity. These challenges can lead to fluctuations in production and prices and can affect the livelihoods of cotton farmers and the competitiveness of the Indian cotton sector in the global market.

The fluctuation in produce prices is mostly caused by variations in market arrivals. There is a need to have a full understanding of the price behavior over time and over space in order to develop suitable strategies and means of decreasing the price volatility of agricultural commodities. Overall, cotton is a critical crop for India, and the government and private sector continue to invest in the sector to support its growth and sustainability. Price projections for the next markets serve to further enhance this information. Farmers can utilize such an analysis to determine the ideal time to dispose of their produce in order to maximize profit. With careful planning of their produce's disposal, farmers can significantly increase their income without incurring substantial additional expenses.

This study aims to forecast the cotton prices in the major cotton-producing states of India, including Gujarat, Maharashtra, Telangana and Karnataka.

Time series analysis is referred as a statistical method that involves the study of data collected over a defined period to identify patterns, trends, and correlations. It is a popular technique for analysing various types of data, including financial data, economic data, climate data, and others. The primary purpose of conducting time series analysis is to generate forecasts about forthcoming values using past data. Time series analysis involves the identification of patterns in



data and the use of statistical models to make predictions. The ARIMA (Autoregressive Integrated Moving Average) model is a powerful and dynamic univariate forecasting technique that can accurately predict future values of a time series. Due to the significance of selecting an appropriate model to analyze cotton price trends and obtain adequate information for decision-making purposes, the study suggests the utilization of the ARIMA model for predicting cotton prices [5,6].

## II. RELATED WORK

Previous research on predicting cotton prices has relied on cotton futures data. However, farmers often struggle to earn a considerable profit from their crops due to additional costs like transportation, intermediaries, and traders [11]. To address this issue and provide farmers with valuable insights, market rates, and accurate price predictions, this study suggests employing Data Mining, Big Data Analytics and Machine Learning techniques [12].

In reference [1], the study aimed to create a crop price prediction system employing Machine Learning (ML) algorithms such as Naïve Bayes and KNN. The research employed past years' crop data and used the training datasets produced by the model to generate valuable insights to predict approximate demand and prices in the market.

In reference [2], the study utilized Data Mining techniques to improve agricultural production of India. To predict prices, the model employed Data Science algorithms [8]. Utilizing a Machine Learning approach, specifically the Naïve Bayes Algorithm, the model made forecasts for prices by taking into account multiple factors including but not limited to precipitation, date, fertilizer usage, trading figures (both minimum and maximum), humidity levels, crop yield, soil composition, wind velocity, cloud cover, sunlight exposure, temperature readings, and pesticide application. The ultimate goal was to eliminate the middleman and prevent farmers and consumers from incurring losses [9].

Reference [3] focuses on predicting crop prices and forecasting them through a web application that employs a supervised Machine Learning algorithm, the Decision Tree Regression algorithm [10]. In order to generate predictions, the model took into account a variety of variables including but not limited to the wholesale price index and precipitation levels. By identifying patterns within the training dataset, the algorithm was able to discern and utilize the inputs provided to it.

Reference [4] used ARIMA models to forecast jute prices in markets of the Murshidabad district of West Bengal. Reference [7] forecasted onion prices in Western Maharashtra using ARIMA models. Reference [13] forecasted cotton prices in selected markets of Karnataka, India using ARIMA models. Reference [14] used ARIMA model to make predictions about the price fluctuations of cotton in the primary states of production within India, as well as groundnut and maize prices in another study. Reference [15] The prices of potatoes in India were predicted by analysing time series data from six different states over a span of eighteen years. Upon conducting a thorough review of the available literature, it has been observed that there exists a dearth of research studies that utilize the ARIMA model for forecasting cotton prices in India. Hence, the objective of this research is to bridge this gap in the current literature by conducting a study on predicting cotton prices in the major states that produce cotton.

## III. METHODOLOGY

In the proposed system, the focus is on predicting cotton prices in major cotton-growing states such as Gujarat, Maharashtra, Telangana and Karnataka. Data pertaining to the monthly prices for these states has been gathered from the AGMARKNET website which falls under the purview of the Ministry of Agriculture, Government of India. The ARIMA model is used to predict the prices and various Python packages are implemented to carry out the analysis.

Time series analysis is a statistical methodology that is utilized to analyse data which is arranged in a sequential, time-based manner. It involves analysing the data over time to identify trends, patterns, and other relationships that can help in forecasting future values. The application of time series analysis can be observed in a diverse array of fields, including economics, finance, engineering, and environmental science. The data in a time series could be collected at regular or irregular intervals, and it could include trend, seasonal, and cyclical components. The primary objective of conducting time series analysis is to develop statistical models that can effectively characterize the data and facilitate the creation of precise predictions regarding future values.



#### A. Dataset Description:

The AGMARKNET website provided the datasets containing historical data on cotton, including Arrival Date, Maximum Price, Minimum Price and Modal Price. Among these components, Arrival Date and Modal Price are the most relevant to the model. The predictor variable is used to estimate the target variable, which is the variable that will be predicted. In this study, Arrival Date serves as the predictor variable, while Modal Price is target variable for each datasets of states. Modal Price refers to the most frequently occurring price of cotton in a specific market during a particular period.

#### B. Pre-processing Dataset:

The process of data pre-processing is an essential step that is undertaken to ensure that the dataset is appropriately formatted and cleaned prior to its utilization for forecasting purposes. Data pre-processing encompasses a series of tasks, one of which involves eliminating redundant and irrelevant data points so that the dataset utilized for forecasting purposes is accurate and devoid of any errors. The conventional methods of data pre-processing include eliminating null values and deleting duplicate data. After checking the datasets, it was found that a few NaN (null) values.

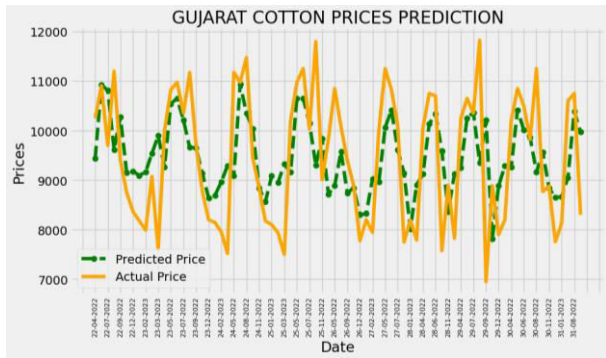
#### C. Methodology:

The Box-Jenkins methodology is a widely used technique to construct ARIMA models for analyzing time series data. This method involves several steps:

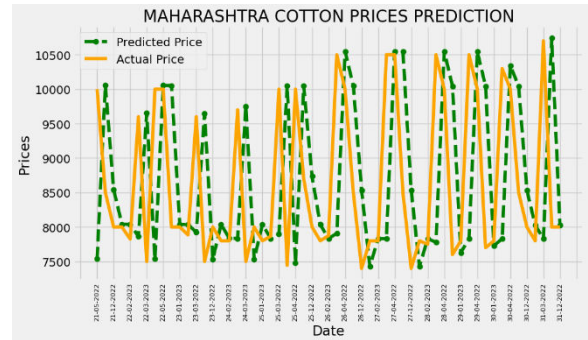
1. Identification: In this step, you analyze the autocorrelation and partial autocorrelation plots of the data to determine the order of the autoregressive (p), the appropriate order of differencing (d) and moving average (q) terms in the ARIMA model.
2. Estimation: Once the sequence of the ARIMA model has been identified, the model parameters are estimated using maximum likelihood estimation.
3. Diagnostic checking: In order to assess the appropriateness of the model, it is necessary to inspect the residuals of the ARIMA model. The residuals are expected to exhibit normal distribution and be uncorrelated.
4. Forecasting: Once a suitable model has been acquired, it can be employed to forecast future values of the time series.

#### D. ARIMA Model:

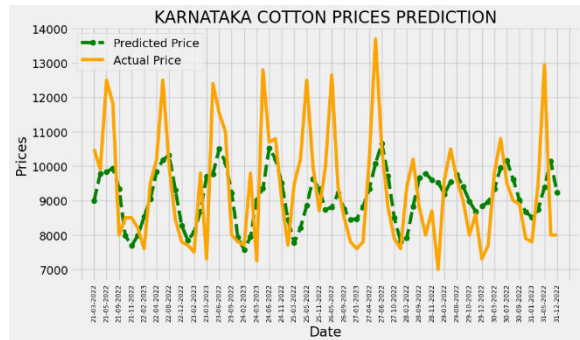
After identifying the optimal p, d, and q values for the ARIMA model, the subsequent stage involves the training and testing of the model. The dataset is partitioned into two segments, with 70% of the data being employed to train the model and the remaining 30% reserved for testing purposes. Once the model has been fit to the training dataset, a `model_prediction` object is created and utilized for generating future forecasts. In order to facilitate comparison, a plot is created wherein the predicted prices are represented in green while the actual prices are denoted by the colour orange. Figures 1 and 2 show the visualization for the Gujarat and Maharashtra datasets, respectively, covering a period of one year. Figures 3 and 4 depict the visualization for Karnataka and Telangana (1 year) datasets respectively.



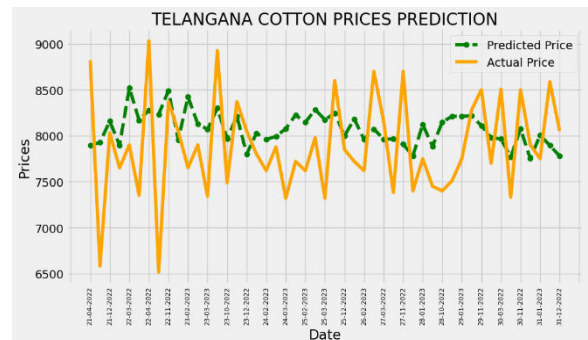
(a)



(b)



(c)



(d)

Fig.1,2,3,4:Forecasting of cotton prices for (a) Gujarat (b) Maharashtra (c) Karnataka (d) Telangana

#### IV. EXPERIMENTAL RESULTS

##### A. Future Price Prediction:

Once the ARIMA model has been trained and tested, the subsequent stage involves utilizing it to forecast future prices for cotton in each of the four distinct regions. This is achieved by calling the `model_fit.forecast()` function, which uses the trained model to forecast the cotton prices for the next 12 days. The results of the forecasts are then plotted using visualization plots to give a clear understanding of the predicted prices. In Figure 5 and 6, the predicted prices for Gujarat and Maharashtra, respectively, are shown using the 1-year dataset. Similarly, Figure 7 and 8 show the predicted prices for Karnataka and Telangana using the same dataset. These predictions can be used by cotton producers, traders, and buyers to make informed decisions about when to buy or sell cotton in these regions. However, it is essential to acknowledge that the precision of the forecasts is contingent upon a multitude of factors, including the quality of the data used to train the model and any unforeseen external factors that may affect cotton prices. As such, it's recommended that the predictions are regularly updated with new data to ensure their reliability.

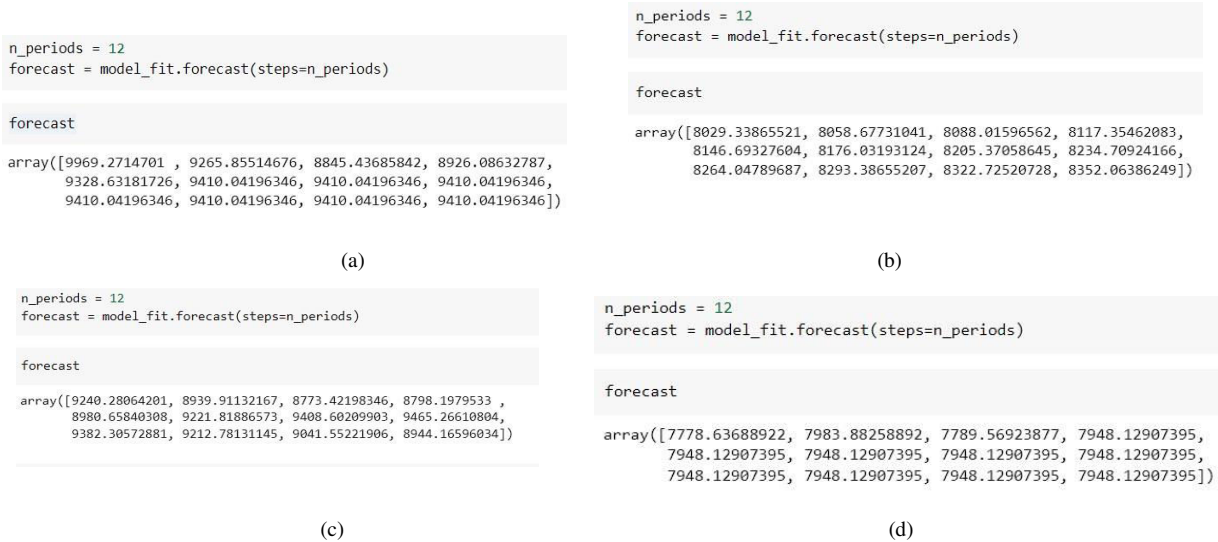


Fig.5,6,7,8:Future Predictionof cotton prices for (a) Gujarat (b) Maharashtra (c) Karnataka (d) Telangana

B. Model Performance

To evaluate the accuracy of the ARIMA model's predictions, several evaluation metrics are employed. The Mean Absolute Error (MAE) calculates the average absolute difference between the actual and predicted values, while the Root Mean Squared Error (RMSE) computes the square root of the mean of the squared differences between the actual and predicted values. The Mean Absolute Percentage Error (MAPE) estimates the percentage difference between the actual and predicted values. These metrics offer a quantitative measure of the model's performance in forecasting future values. By comparing the values of these metrics for different models or different sets of parameters, one can determine the best model or parameters for accurate predictions. The evaluation metrics can also help in identifying any shortcomings of the model and suggest improvements that can be made to enhance the model accuracy.



Fig.9,10,11,12:Model Performance of 1 year dataset of (a) Gujarat (b) Maharashtra (c) Karnataka (d) Telangana



## VI. CONCLUSION

Cotton price forecasting is a challenging task due to the volatility and fluctuations in the market. Notwithstanding, the outcomes of this study propose that the ARIMA model can function as a trustworthy and precise instrument for predicting forthcoming cotton prices. The study validates this hypothesis by applying the ARIMA model to the cotton data of four different states in India. By analyzing the trends from the previous year's data, the ARIMA model can forecast future cotton prices with reasonable accuracy. This information holds significant value for both the cotton industry and the broader Indian economy. However, it is important to note that the ARIMA model's short-term forecast accuracy is higher than its long-term forecast accuracy. Overall, the ARIMA model can be a useful tool for predicting future cotton prices, but it is important to consider its limitations and use it in combination with other forecasting methods.

## REFERENCES

- [1] M. Rachana P S, Rashmi G, Shravani D, Shruthi N, Seema Kousar R, "Crop Price Forecasting System Using Supervised Machine Learning Algorithms", International Research Journal of Engineering and Technology, Volume 06, Issue 04, 2019.
- [2] Varun R, Neema N., Sahana H. P., Sathvik A., Mohammed Muddasir, "Agriculture Commodity Price Forecasting using ML Techniques", International Journal of Innovative Technology and Exploring Engineering, Volume-9 Issue-2S, December 2019.
- [3] Rohith R, Vishnu R, Kishore A, DeebanChakkarawarthy, "Crop Price Prediction and Forecasting System using Supervised Machine Learning Algorithms", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 9, Issue 3, March 2020.
- [4] Kumar, R. R., Gupta, A. K., & Patra, C. (2020). Jute price forecasting in Murshidabad market of west Bengal using ARIMA technique. *Journal of Pharmacognosy and Phytochemistry*, 9(1), 1802-1807.
- [5] Wint Nyein Chan, "Time Series Data Mining: Comparative Study of ARIMA and Prophet Methods for Forecasting Closing Prices of Myanmar Stock Exchange", *Journal of Computer Applications and Research*, vol. 1, pp.75-80, 2020.
- [6] A. A. Ariyo, A. O. Adewumi and C. K. Ayo, "Stock Price Prediction Using the ARIMA Model," 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, Cambridge, UK, 2014, pp. 106-112, doi: 10.1109/UKSim.2014.67.
- [7] Darekar, A. S., Pokharkar, V. G., & Datarkar, S. B. (2016). Onion price forecasting in Kolhapur market of Western Maharashtra using ARIMA technique. *International Journal of Information Research and Review*.
- [8] D. Zhang, S. Chen, "Forecasting Agricultural Commodity Prices Using Model Selection Framework With Time Series Features and Forecast Horizons", *IEEE*, Feb. 2020.
- [9] Yung-Hsing Peng, ChinShun Hsu, and Po-Chuang Huang, "Developing Crop Price Forecasting Service Using Open Data from Taiwan Markets", 2017 *IEEE*.
- [10] Md. Tahmid Shakoor, Karishma Rahman, Sumaiya Nasrin Rayta, Amitabha Chakrabarty, "Agricultural Production Output Prediction Using Supervised Machine Learning Techniques", 2015 *IEEE*.
- [11] Rahman. N. M. F. "Forecasting of Boro Rice Production in Bangladesh: An ARIMA Approach", *J. Bangladesh Agril. Univ.*, 2010; 8(1), pp. 103–112.
- [12] Wankhade. R, Mahalle. S, Gajbhiye. S, Bodade.V.M., "Use of the ARIMA Model for Forecasting Pigeon Pea Production in India", *International Review of Business and Finance*, 2010; ISSN 0976–5891, 2(1), pp. 97–102.
- [13] Nagaraja, Y. (2019). "Behavior of Arrivals and Prices of Cotton in Selected Markets of Karnataka-A Statistical Analysis", Doctoral dissertation, UNIVERSITY OF AGRICULTURAL SCIENCES GKVK, BENGALURU.
- [14] Darekar, A. and Reddy, A.A. (2017). "Cotton price forecasting in major producing states", *Economic Affairs*, 62(3), pp.1-6.
- [15] Kumar, R.R. (2020). "Forecasting of Potato Prices in India: An Application of ARIMA Model", *Economic Affairs*, 65(4), pp.335-190.
- [16] Achudhan, A.M., Sankar, S.S., and Saikumar, K. (2018). "Experimental study on coir fibre mixed concrete", *International Journal of Pure and Applied Mathematics*, 118(20), pp.2913-2929.



**INNO SPACE**  
SJIF Scientific Journal Impact Factor  
**Impact Factor: 8.423**



**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

**9940 572 462** **6381 907 438** **ijirset@gmail.com**



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details