

## The Crop Price Prediction Using Machine Learning: Preliminary Stage

Ramya T K <sup>1\*</sup>, Shreedhara N Hegde <sup>1\*</sup>, Mohd Norshahriel Abd Rani<sup>2</sup>

<sup>1</sup>Dayananda Sagar Academy of Technology and Management, Karnataka, India.

<sup>2</sup>Faculty of Data Science and Information Technology, INTI International University, Malaysia

**Emails:** <sup>1</sup>ramyatk196@gmail.com, <sup>1</sup>shreedhara-mca@dsatm.edu.in

### Abstract

The objective of our research is mostly concerned with agriculture. Farmers are the key players in agriculture. Knowing how much a crop will cost will enable you to make smarter judgments, which will reduce the losses and lower the risk of price changes. An ML model that forecasts agricultural prices in advance while accurately analyzing the crop may be able to solve this issue. A predictive system, a statistical method combining machine learning and data collecting, is used in many applications, including healthcare, retail, education, and government sectors. Its usage in the agriculture sector has comparable relevance. The back-end predictive model for this project is created utilizing machine learning algorithms. The steps involved in creating a predictive model are data collecting, data cleaning, data mining, and validation. The goal is to give farmers an intuitive user interface, and this model should correctly forecast crop market value given the real-time variables provided.

### Keywords

Crop Price Prediction, Random Forest Algorithm, Machine Learning Model.

### Introduction

India is regarded as an agrarian nation because around 55% of its people rely on agriculture or closely related industries to make a living. An important component of India's first sector of the economy is agriculture (Ghutake, I., et.al., 2021). Farmers experience significant losses due to barren soil, global climate change, an abundance of products on the market, and numerous other unknowns. To manage and sell the products at the optimum time to maximize revenue and minimize loss, it is imperative to predict projected prices in the future (Rachana, P. S., et.al., 2019).

In the past, farmers' experience with specific crops and fields was used to anticipate prices (Sabu, K. M., & Kumar, T. M., 2020). Assume that we have access to historical data with

**Submission:** 19 May 2024; **Acceptance:** 23 June 2024



**Copyright:** © 2024. All the authors listed in this paper. The distribution, reproduction, and any other usage of the content of this paper is permitted, with credit given to all the author(s) and copyright owner(s) in accordance to common academic practice. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license, as stated in the website: <https://creativecommons.org/licenses/by/4.0/>

various matching price forecasts recorded and that we will utilize this data to categorize potential future price projections. Farmers need accurate price forecasting as well since they use it to guide their marketing and production choices, which may have an impact on their finances months down the road. This forecast is based on data from official sources and employs machine learning methods. The price of a crop can also be predicted via data mining. Data mining, in general, is the process of examining data from various angles and condensing it into helpful information (Rohith, R., et.al, 2020). Crop price forecasting is a significant agricultural issue. Every farmer strives to understand how much of his expected price he will receive (Darekar, A., & Reddy, A. A., 2017). With regard to machine learning/deep learning-based models, updating the models could result in stability problems due to quality concerns with the crop price data since new price data is entered every day (Peng, Y. H., 2015). Other agriculture-based enterprises may find this website useful for planning the cost-effective sourcing of raw materials (Shakoor, M. T., 2017). Based on machine learning, the research will give farmers and those in related fields access to predictions at any time that will show them expected price trends for different crops 12 months in advance. As a result, our feeding farmers may find an ML model with outstanding prediction accuracy to be a tremendous benefit.

In this study, crop price predictions are made using machine learning techniques. Additionally, a thorough study of the crop was provided, and a future scenario was projected so that farmers could choose the best crop for their specific situation (Shakoor, M. T., et.al 2021).

This essay seeks to forecast the price and profit of the specified crop before sowing. Rainfall, Maximum-trade, Minimum Support Price, and Yield are the factors taken into account for price prediction. Crop price, yield, cultivation costs, and seed costs are the factors taken into account for profit prediction. The K-Nearest Neighbour (KNN) algorithm was used to forecast crop profit while the Naive Bayes Algorithm was utilized to forecast crop pricing (Ghutake, I., et.al. 2021).

In this study, machine learning models and time-series models are used to forecast arecanut prices every month in Kerala. The arecanut dataset with prices from 2007 to 2017 was used to examine the performance of the model SARMIA, Holt-Winter's Seasonal approach, and the LSTM neural network. The model that best fits the data was determined to be the LSTM neural network model (Guruprasad, R. B, et. al., 2019).

Forecasts in great detail for the upcoming six months were made in this study. A supervised learning approach called Decision Tree Regressor was applied. Data on rainfall and crop prices were taken into account as prediction criteria (Thomas, D., 2023).

In this study, future paddy price predictions were made using the Auto Regressive Integrated Moving Average (ARIMA) model. The effectiveness of the model was evaluated by computing various goodness of fit metrics. According to the findings, the ARIMA model is the most effective model for predicting (Dhanapal, R. et.al., 2021).

From local markets and internet questionnaires, the concerned crops' data are gathered for this system's dataset for machine learning Models undergo training. Algorithms like Artificial Neural Networks, Partial Least Squares, and Autoregressive Integrated Moving

Averages are used to predict prices (Ikram, A. et.al., 2022). Partial least square and artificial neural networks are superior when compared to other algorithms for both short- and long-term prediction, according to results from the use of the aforementioned algorithms using recent data.

This study attempted to assist farmers in making decisions by rating the compatibility of a crop to the area in question. supervised machine learning methods, such as the K closest neighbor regression algorithm, and decision tree learning, are used for prediction and ranking.

### **Research Methodology**

The following methodology is applied to our research, and it is illustrated in Figure 1.

1. **Data Collection:** The first step is to collect historical data on crop prices and other relevant factors that can impact the prices, such as weather, market demand, and supply chain information. The data can be obtained from various sources, including government agencies, commodity exchanges, and private data providers. It includes year, month, state, district, crop, and type of crops.
2. **Data Preprocessing:** Once the data is collected, it needs to be pre-processed. It involves cleaning, transforming, and organizing the data to make it suitable for further analysis. Here are some common steps for preprocessing crop price data.
3. **Data Cleaning:** Remove any duplicate records from the dataset. The rows with missing values are eliminated.
4. **Transformation:** The dataset includes categorical values. These categorical values are converted into numerical values.
5. **Feature Selection:** Feature selection involves selecting the most relevant features from the dataset that can impact crop prices. This can be done using various techniques such as correlation analysis.
6. **Model Training:** After feature selection, the Random Forest algorithm is trained on the pre-processed data. The algorithm creates multiple decision trees on randomly selected subsets of data and combines them to generate the final prediction model. The algorithm uses techniques like bagging and feature importance to improve the model's performance and interpretability.
7. **Model Evaluation:** Once the model is trained, it needs to be evaluated to determine its accuracy and robustness. This can be done using various metrics such as mean squared error and mean absolute error.
8. **Prediction:** After model evaluation, the trained Random Forest algorithm can be used to predict future crop prices. The algorithm takes as input the relevant features for the prediction and generates a prediction for the crop price

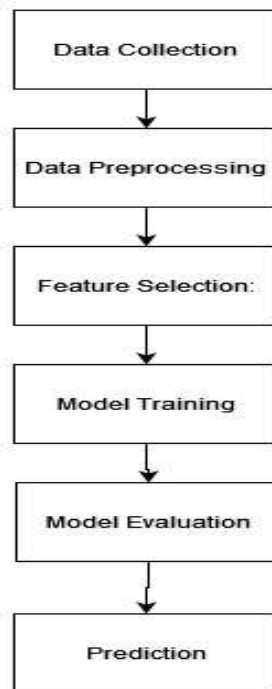


Figure 1. Flow Chart of Methodology

### Results and Analysis

This manuscript offers a succinct overview of crop yield forecasts made for the chosen region using multiple linear regression. The agricultural study of organic and nonorganic farming, timely crop cultivation, data on profits and losses, and analysis of local business land in a specific location are its main areas of concentration.

It concentrates on the real estate, organic, and inorganic data sets that will provide agricultural projections. The use of data mining techniques to forecast agricultural yields based on input factors has been demonstrated in this research. Numerous agro-climatic input characteristics have an impact on crop output. The input of climatic parameters into a system designed to predict crop yields shows a tendency for each crop to be predominately influenced by a specific climatic parameter.

To address numerous agricultural issues, this paper presents several applications of data mining. Because it combines the works of numerous authors in one location, scholars can use it to learn more about the status of data mining techniques and applications specific to the agriculture industry.

From the algorithms that we run to measure the better prediction of retail crop prices, the results of prediction are tabulated using bar charts like those in Figures 2 and 3. The graphs show the retail sale price of all the states in January the states in a year for the rice crop.

From the experiments, we can conclude that the Random Forest algorithm can be a better tool for crop price prediction in agriculture. The algorithm combines multiple decision trees to improve the accuracy and robustness of the prediction model and uses techniques like bagging and feature importance to enhance the model's performance and interpretability. By training the algorithm on historical data on crop prices and other relevant factors, stakeholders can use the model to predict future crop prices and make informed decisions about buying, selling, or storing their crops.

Using Random Forest for crop price prediction can also provide various benefits, such as improving market efficiency, reducing risks, and helping policymakers make informed decisions. However, it is essential to note that the accuracy of the predictions depends on the quality and quantity of the data, the feature selection, and the model's training and evaluation. Therefore, it is crucial to collect high-quality data, use appropriate feature selection techniques, and regularly update and monitor the model to ensure its accuracy and relevance.

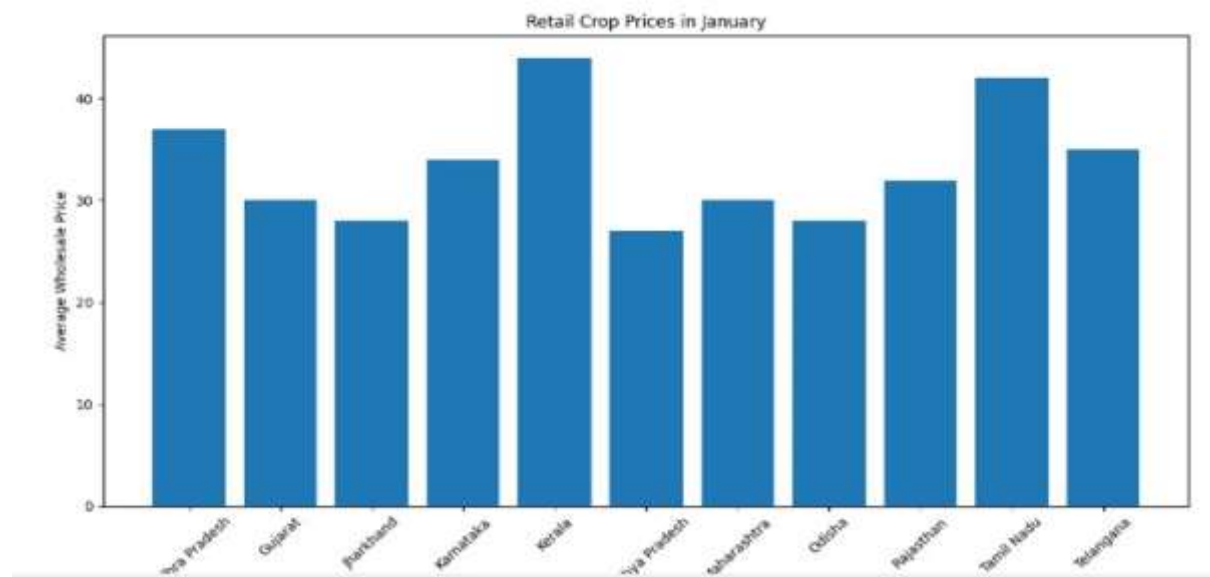


Figure 2: The graph shows the retail price of all states

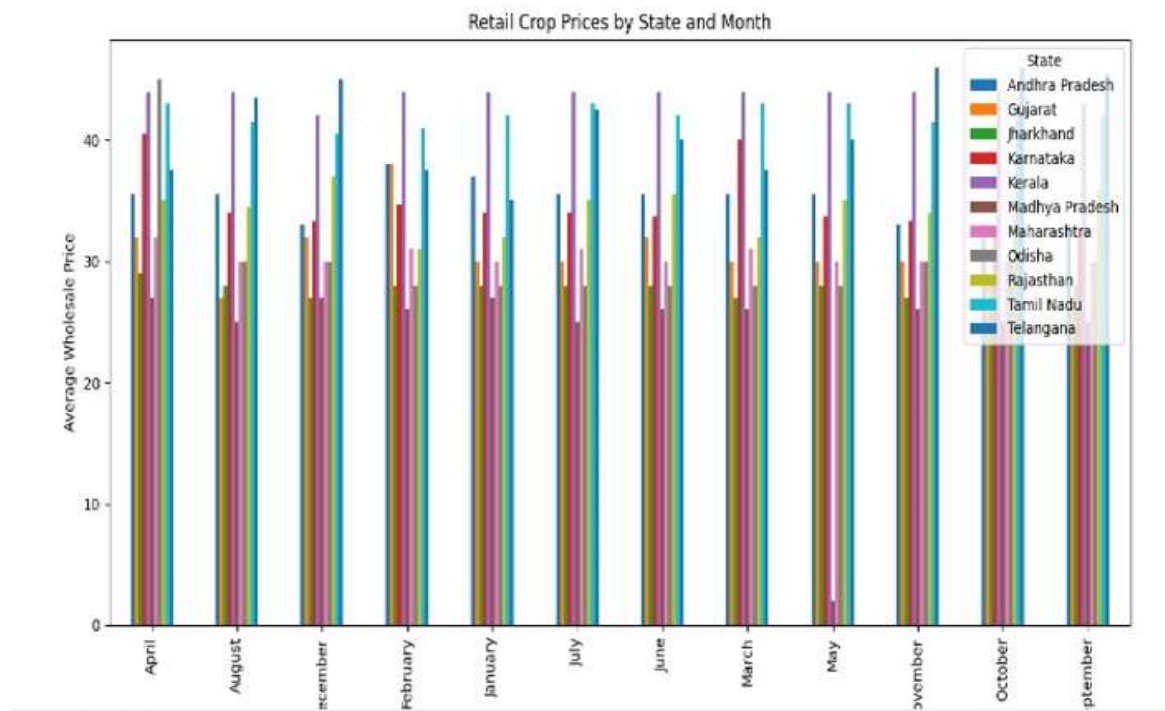


Figure 3: The retail sale price of all the states in January

### Conclusion

In summary, Random Forest can be a valuable tool for crop price prediction in agriculture, providing accurate and reliable predictions that can help stakeholders make informed decisions and mitigate risks. In addition to these advantages, using Random Forest for crop price prediction can also lower risks, increase market efficiency, and support decision-making by policymakers. It is imperative to acknowledge that the precision of the forecasts is contingent upon the caliber and volume of the data, the feature selection process, and the training and assessment of the model. To guarantee the accuracy and applicability of the model, it is therefore essential to gather high-quality data, apply suitable feature selection approaches, and update and monitor it regularly.

### Acknowledgments

The authors would like to express our heartfelt gratitude to Dayananda Sagar Academy of Technology and Management (DSATM) for providing us with the necessary resources and facilities to conduct this research project on the “Crop Price Prediction System”. The support and encouragement from the institution have been instrumental in the successful completion of this endeavor.

## References

- Ghutake, I., Verma, R., Chaudhari, R., & Amarsinh, V. (2021). An intelligent crop price prediction using suitable machine learning algorithm. In *ITM web of conferences* (Vol. 40, p. 03040). EDP Sciences.
- Rachana, P. S., Rashmi, G., Shravani, D., Shruthi, N., & Kousar, R. S. (2019). Crop price forecasting system using supervised machine learning algorithms. *International Research Journal of Engineering and Technology (IRJET)*, 6, 4805-4807.
- Sabu, K. M., & Kumar, T. M. (2020). Predictive analytics in Agriculture: Forecasting prices of Arecanuts in Kerala. *Procedia Computer Science*, 171, 699-708.
- Rohith, R., Vishnu, R., Kishore, A., & Chakkarawarthy, D. (2020). Crop price prediction and forecasting system using supervised machine learning algorithms. *International Journal of Advanced Research in Computer and Communication Engineering*, 9(3).
- Darekar, A., & Reddy, A. A. (2017). Forecasting of common paddy prices in India. *Journal of Rice Research*, 10(1), 71-75.
- Peng, Y. H., Hsu, C. S., & Huang, P. C. (2015, November). Developing crop price forecasting service using open data from Taiwan markets. In *2015 Conference on Technologies and Applications of Artificial Intelligence (TAAI)* (pp. 172-175). IEEE.
- Shakoor, M. T., Rahman, K., Rayta, S. N., & Chakrabarty, A. (2017, July). Agricultural production output prediction using supervised machine learning techniques. In *2017 1st international conference on next generation computing applications (NextComp)* (pp. 182-187). IEEE.
- Ghutake, I., Verma, R., Chaudhari, R., & Amarsinh, V. (2021). An intelligent crop price prediction using suitable machine learning algorithm. In *ITM web of conferences* (Vol. 40, p. 03040). EDP Sciences.
- Guruprasad, R. B., Saurav, K., & Randhawa, S. (2019, July). Machine learning methodologies for paddy yield estimation in India: a case study. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium* (pp. 7254-7257). IEEE.
- Thomas, D. (2023, May). A review of crop yield prediction based on Indian agriculture sector using machine learning. In *AIP Conference Proceedings* (Vol. 2773, No. 1). AIP Publishing.

Dhanapal, R., AjanRaj, A., Balavinayagapragathish, S., & Balaji, J. (2021, May). Crop price prediction using supervised machine learning algorithms. In *Journal of Physics: Conference Series* (Vol. 1916, No. 1, p. 012042). IOP Publishing.

Ikram, A., Aslam, W., Aziz, R. H. H., Noor, F., Mallah, G. A., Ikram, S., ... & Ullah, I. (2022). Crop Yield Maximization Using an IoT-Based Smart Decision. *Journal of Sensors*, 2022, 1-15.