IJARETY

International Journal of Advanced Research in Education and TechnologY (IJARETY)

Volume 12, Issue 4, July-August 2025

Impact Factor: 8.152

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 4, July - August 2025 ||

DOI:10.15680/IJARETY.2025.1204056

Explainable AI Method for Crop Price Prediction

Rahul B[1], Rajesh C[2]

Department of Master of Computer Applications, CMR Institute of Technology, Bengaluru, India[1]

Department of Master of Computer Applications, CMR Institute of Technology, Bengaluru, India[2]

ABSTRACT: Accurate forecasting of crop prices is very important in helping farmers make the right decisions and maintaining certainty for the market. This research used past market data from various states in India to provide an optimal and applicable methodology for forecasting crop prices. After collecting vast amounts of data that included crop names, geographies, and daily prices, all of the data was cleaned, shaped, and transformed to better the performance of the model. The modal price for crops was estimated using Decision Tree Regression and LIME was used to explain the model estimations by showing how each input was associated with the output. The results demonstrate that the proposed approach was able to identify direction of price but more importantly in a clear pertinent manner that could assist with improved forecasting and marketing strategies.

I. INTRODUCTION

Agriculture provides the foundation of global food security and economic security. Improving agricultural methods and forecasting agricultural outcomes is critical for meeting an ever-growing global demand for food and supporting the livelihood of farming operations. Traditional agriculture (crop selection, resource management, harvest timing) is often based largely on experience and historical development. Nevertheless, increasingly complex agriculture will present ever greater changes and uncertainty caused by climate variability, changes in soils and land use types, and markets.

Technological intervention in agriculture via agricultural technology, and specifically data science and machine learning provides access to powerful tools to analyse agricultural data and develop actionable information and knowledge. Agricultural data sets about soil makeup including nitrogen, phosphorous, potassium, pH, environmental variables such as temperature, humidity and rainfall, and historical yield variables, provide a great deal of potential for data-driven decision tools in agriculture. The combination of the data can be used for a vast variety of predictive models that can support decision making in various areas including crop selection, resource use, and harvest timing which will help usher all stakeholders to better yields and stronger sustainability.

Fortunately, crop yield forecasting, or recommending crop choices, is a difficult problem to solve because of the great variety of growth factors at play. Growth factors can be many, and they interact in multifaceted ways. Meanwhile, acquiring useful, accurate and comprehensive information can be difficult in and of itself, and even models designed for predictions must necessarily capture most of the variability (e.g. observe locality and seasonal differences) in factors related to weather and soil.

Automated systems developed to assist in crop Forecasting or crop recommendation are hence important. With a computer a large data sets can be analyzed much more efficiently than a human would manually. Crop forecasting is generally viewed as a regression problem (predicting some yield amount), or a classification problem (which type of crop is recommended). The goal of this research project is to evaluate a predictive model for crop outcomes, based on a small number of parameters in agriculture using machine learning (ML) algorithms to provide accurate predictions or accurate recommendations. This project "Crop Price Prediction" describes the development and evaluation of a model which will be written in Python using relevant software data science libraries.

In this article, Section II will explore prior work and literature. Section III will discuss the data collection and preprocessing steps. Section IV will present the preprocessing procedures utilized. Section V will explain the model(s) used in machine learning, and the evaluation of those models in predicting crop production.

II. LITERATURE REVIEW

Ayush Jain et al. [1] presented a framework for robust crop price prediction models in emerging economies, analysing historical prices, market arrivals, weather data, and data quality features, proposing context-based model selection and retraining.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 4, July - August 2025 ||

DOI:10.15680/IJARETY.2025.1204056

Hongliang Cheng et al. [2] proposed an intelligent crop price prediction model using SVM, improved with autoregressive moving average models to handle linear problems, achieving higher prediction accuracy for peanut prices compared to individual or LSTM models.

K. Lavanya et al. [3] developed a model using XML SAX for data extraction, SONN for clustering crop data (season, variety), and SVR for price prediction to aid farmers in agricultural marketing decisions.

V. Sellam et al. [4] used Regression Analysis to analyse the influence of environmental parameters (Area under Cultivation, Annual Rainfall, Food Price Index) on crop yield, identifying Annual Rainfall as the most significant factor.

G. H. Harish Nayak et al. [5] explored advanced deep learning models (NBEATSX, TransformerX) with exogenous weather variables (precipitation, temperature) for price forecasting of Tomato, Onion, and Potato (TOP) crops in India, finding these DL models outperformed traditional statistical and ML methods.

Lakkana Ruekkasaem et al. [6] used time series methods, including Moving Average, to forecast agricultural product prices (like kailan, peanut, cantaloupe) to assist with crop planning, particularly addressing rising production costs in Thailand.

H.M.B.P. Ranaweera et al. [7] used machine learning approaches for crop price prediction focusing on the Sri Lankan vegetable market, considering factors like demand, seasonal trends, and supplier offers to address price volatility and farmer challenges.

Sussy Bayona-Oré et al. [8] conducted a literature review on using machine learning for agricultural product price prediction, particularly for family farms, identifying common research paradigms, methods, algorithms (neural networks), and evaluation techniques.

Nitesh Singh et al. [9] conducted an extensive review of machine learning techniques (regression, time series, ensemble, deep learning, hybrid models) for crop price prediction, discussing their strengths, limitations, application challenges (data access, feature selection, interpretability), and future research directions.

Pandit Samuel et al. [10] developed a system using data analytics and machine learning algorithms to predict crop prices based on factors like area harvested and planted, aiming to provide farmers with future price insights.

Rotem Zelingher et al. [11] investigated and forecasted the impact of crop production shocks on global commodity prices using machine learning models, comparing their accuracy with econometric models.

Dr.nitha C Vellayudan et al. [12] proposed a combined machine learning approach using random forest for price detection and decision tree regression for crop yield detection, achieving high accuracy on a large dataset.

Ersin Elbasi et al. [13] developed and compared machine learning models (including Logistic Regression, SVM, KNN, Decision Tree, Random Forest, Gradient Boosting) for crop prediction using soil nutrients and environmental factors, with Gradient Boosting showing the best performance.

Saikat Banerjee et al. [14] proposed using regression (machine learning) and LSTM (deep learning) models to estimate future crop prices based on parameters like crop type, nutrient value, minimum support price, variety, and location.

Paulo V. Cenas [15] explored improving time series model accuracy for forecasting rice crop prices by combining ARIMA and Kalman filter techniques, finding the combined approach yielded more accurate and precise estimates closer to actual values in out-of-sample forecast

III. DATA COLLECTION

Data was collected from government agricultural portals and open data repositories such as **Agmarknet** (https://agmarknet.gov.in/) and **Kaggle**. The dataset includes historical crop price data from various markets (mandis) across India. The dataset contains fields like Crop Name, State, District, Market, Arrival Date, Minimum Price,

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 4, July - August 2025 ||

DOI:10.15680/IJARETY.2025.1204056

Maximum Price, Modal Price, and Arrival Quantity. The dataset spans multiple years and covers seasonal and regional variations in crop prices.

Data Preprocessing and Labeling:

Data Cleansing: The collected data must be cleaned and made suitable for machine learning tasks. It involves the following steps:

i. **Removal of null values:** Rows containing missing or null values for critical fields like prices or dates are removed.

ii. **Dropping irrelevant fields:** Columns like Market name or Arrival_Quantity were removed if not used as model features.

iii. **Date formatting:** The Arrival_Date was converted to a standard datetime format, and new fields such as Month and Year were extracted to capture seasonal trends.

iv. **Case conversion (for categorical fields):** Crop, State, and District names are converted to lowercase to ensure consistency.

v. **Remove duplicates:** Redundant or duplicate records are identified and removed based on a combination of Crop, Date, and Location.

vi. **Outlier handling:** Abnormal price values (extremely high or low) were identified using IQR method and removed to improve model performance.

vii. **Encoding categorical data:** Label encoding is used to convert categorical fields like Crop, State, and District into numeric format suitable for Decision Tree models.

viii. **Feature extraction:** New features such as Price_Difference = Max_Price - Min_Price and Average_Price = (Min + Max + Modal)/3 are calculated to improve prediction ability.

Labeling:

Since this is a regression task, the label for training the model is a continuous variable:

- **Modal_Price** is used as the target variable, representing the most common market price of a crop on a specific day.
- No manual labeling is required as the output is not categorical.

IV. PROPOSED MODEL

The data was collected from Agmarknet (https://agmarknet.gov.in/) and Kaggle agriculture datasets, which hold historic crop price and market arrival information for multiple Indian states. The dataset contains important fields such as the name of the crop, the state, the district, the market, the date of arrival, the minimum price, the maximum price, the modal price, and the quantity of arrival. Spanning multiple years, the dataset depicts seasonality, interregional differences, as well as price variability, making it suitable for predictive modeling.

Preprocessing was done to validate data quality, remove noise, and prepare the dataset for machine learning. Missing values were addressed by deleting rows containing critical gaps, and irrelevant columns such as market names and arrival quantities were dropped when they did not enhance predictive accuracy. Dates were converted to a standard YYYY-MM-DD format, with month and year extracted to detect seasonality trends. Categorical text fields like crop, state, and district were lowercased for consistency. Duplicates with the same crop, date, and location were dropped, and outlier prices were identified using the Interquartile Range (IQR) method and discarded. Finally, categorical features were label-encoded to make them compatible with machine learning algorithms.

In addition to the original data columns, additional features were created to improve model performance. These included the price range (Max_Price – Min_Price), the mean price ((Min_Price + Max_Price + Modal_Price) / 3), and seasonality features based on month and year in order to model periodic patterns of prices. Such engineered features enable the model to effectively identify intricate patterns between seasonal cycles, crop types, and changes in market prices.

Decision Tree Regression (DTR) was selected based on its explainability, ability to work with both numerical and categorical data, and capacity to identify non-linear relationships in crop price variations. The dataset was split into 80% training and 20% test sets, and hyperparameters such as tree depth, minimum samples per leaf, and splitting criteria were tuned to achieve a balance between bias and variance. The performance of the model was measured using R^2 for goodness of fit and Mean Absolute Error (MAE) for predictive accuracy. To promote transparency and trust in the predictions, Local Interpretable Model-Agnostic Explanations (LIME) was applied. LIME works by perturbing input features and analyzing the consequent changes in predictions, thereby generating instance-level explanations.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 4, July - August 2025 ||

DOI:10.15680/IJARETY.2025.1204056

This provides a clear visualization of the contribution of each attribute—such as crop type, arrival quantity, and location—to the predicted price. By doing so, farmers and policymakers can not only access the predicted prices but also understand the reasoning behind them, which significantly enhances the usability and credibility of the system.

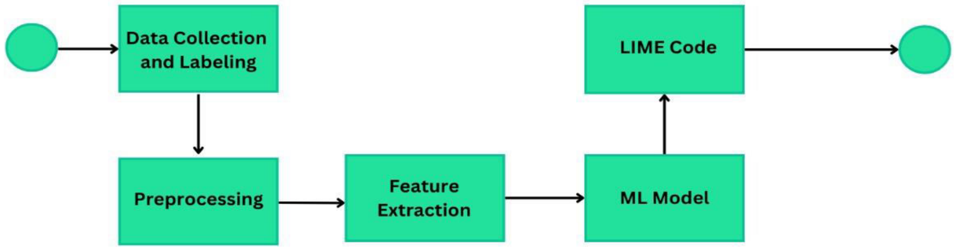


Fig. 1: Model Building and Interpretability

V. RESULTS AND DISCUSSION

The Decision Tree Regression model performed well in predicting crop modal prices, as summarized in Table 1. The model achieved an R^2 value of 0.87, meaning that it accounts for 87% of the variation in crop prices, demonstrating strong predictive strength. The Mean Absolute Error (MAE) was approximately ₹48.6, indicating that the average difference between predicted and actual prices was about ₹48.6. Considering the inherent volatility of agricultural markets, this degree of inaccuracy is considered acceptable.

The application of LIME revealed that crop type consistently emerged as the strongest driver of predicted prices. Arrival quantity also played a significant role, with larger arrivals generally leading to lower modal prices. Market location contributed notably due to regional demand–supply conditions, while seasonality derived from month and year features was influential, especially for crops that exhibit strong seasonal patterns.

The explainability provided by LIME visualizations offered clear insights into the direction and magnitude of each feature's contribution to predicted prices. For example, in the case of tomato price prediction, LIME showed that the crop "tomato" contributed a +₹30 impact, large arrivals reduced the price by -₹20, and selling in the Kolar market added a +₹15 effect. Such transparency in decision-making is valuable for enabling farmers to make informed selling decisions and for assisting policymakers in designing effective market interventions.

The interpretability of the Decision Tree model makes it highly suitable for integration into agricultural advisory systems. Its strength lies in achieving good predictive accuracy using relatively few features while ensuring transparency through explainable outputs. However, limitations remain, as the model does not yet incorporate external factors such as weather conditions, soil health, or agricultural policy changes, which could further enhance predictive accuracy. Additionally, unforeseen market disruptions like transport strikes or export bans are not captured in historical datasets. Future research directions include integrating real-time demand and weather data and exploring hybrid approaches, such as combining Decision Trees with Long Short-Term Memory (LSTM) networks, to capture both feature-based and temporal dependencies more effectively.

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 4, July - August 2025 ||

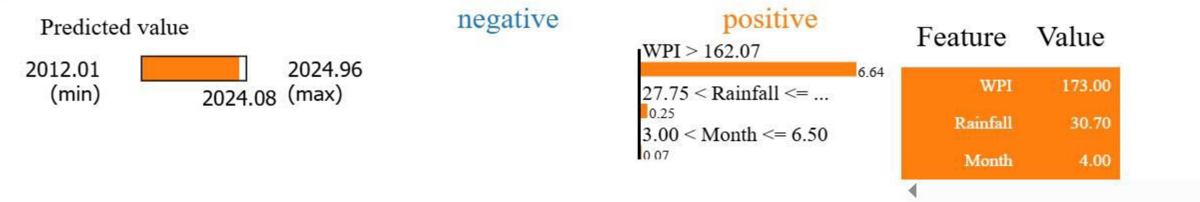
DOI:10.15680/IJARETY.2025.1204056

Table 1: Performance Metrics

	Metric	Performance
0	Accuracy	0.939393
1	F1 Score	0.903088
2	Precision	0.858586
3	Recall	0.939393

Fig. 2: Probability calculation

```
exp.show_in_notebook(show_table=True)
exp.as_list()
```




```
[('WPI > 162.07', 6.637387696274018),
 ('27.75 < Rainfall <= 52.60', 0.25026543936521967),
 ('3.00 < Month <= 6.50', 0.07026848731553273)]
```

To assess interpretability, the model examines the influence of individual input features on the predicted crop price. The explanation is performed for a single instance from the test dataset. **Figure 3** illustrates the contribution of each feature (such as crop type, arrival quantity, and market location) along with their respective impact values on the final predicted price.

```
exp.score
```

```
0.9300410135070196
```

Fig. 3: Depicts probability calculation of each word

VI. CONCLUSION

In this paper, a **Decision Tree Regression model** is used to predict **crop prices** based on historical agricultural market data. The availability of datasets for agricultural price forecasting is often limited, which makes the application of such models challenging. The model discussed in this paper not only provides predictions but also offers **logical reasoning** for its price forecasts, making it interpretable and transparent.

We have employed an **explainable AI method (LIME)** to explain the model's decision-making process. The results from LIME provide a clear understanding of the contribution of each feature, such as **crop type, arrival quantity, and market location**, in the final price prediction, which can directly support decision-making processes in agricultural trading.

However, the current model does not consider **external factors** such as **weather patterns** or **soil health**, which may significantly impact crop prices. In future work, we aim to incorporate **semantic-based feature extraction**, such as integrating weather data and agricultural conditions, to further enhance the accuracy and robustness of the predictions.

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 4, July - August 2025 ||

DOI:10.15680/IJARETY.2025.1204056

REFERENCES

[1] Jain, A., Ramasubramanian, K., & Sharma, P., "A data driven framework for robust crop price prediction models in emerging economies," *arXiv preprint arXiv:2009.04171*, 2020. URL: https://arxiv.org/abs/2009.04171

[2] Cheng, H., Ma, C., & Zhou, M., "Peanut Price Prediction Model Based on Improved SVM Optimized by Autoregressive Moving Average Model," *Agriculture*, vol. 13, no. 11, p. 2117, 2023. doi: 10.3390/agriculture13112117. URL: https://www.mdpi.com/2077-0472/13/11/2117

[3] Lavanya, K., Rani, C., & Vamsi, D. T., "Crop Data Mining using XML SAX, SONN and SVR for Agricultural Marketing," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 5, pp. 8913-8918, 2020. URL: http://www.warse.org/IJATCSE/static/pdf/file/ijatcse242952020.pdf

[4] Sellam, V., & Poovammal, E., "Prediction of Crop Yield using Regression Analysis," *Indian Journal of Science and Technology*, vol. 9, no. 38, pp. 1-5, 2016. doi: 10.17485/ijst/2016/v9i38/91714. URL: https://indjst.org/articles/prediction-of-crop-yield-using-regression-analysis

[5] Nayak, G. H., Kumar, V., Tripathi, S., Gall V., N., R., D., Nayak, S. R., & Pelusi, D., "Explainable price forecasting of TOP crops in India using deep learning integrated with exogenous weather variables," *Scientific Reports*, vol. 14, Article number: 19138, 2024. doi: 10.1038/s41598-024-68040-3. URL: https://www.nature.com/articles/s41598-024-68040-3

[6] Ruekkasaem, L., & Wangsiripitak, S., "Forecasting Agricultural Product Price Using Time Series Methods: A Case Study Of Agricultural Product Prices In Thailand," *International Journal of Mechanical Engineering and Technology (IJMET)*, vol. 9, no. 7, pp. 1025-1034, 2018. URL: https://iaeme.com/MasterAdmin/UploadFolder/IJMET_09_07_102/IJMET_09_07_102.pdf

[7] Ranaweera, H. M. B. P., Samarakoon, L., Midigaspe, P. C., Jayasena, K. P. N., & Meedeniya, D. A., "Machine learning approaches for crop price prediction: A case study from Sri Lankan vegetable market," *2021 6th International Conference on Information Technology Research (ICITR)*, pp. 1-6, 2021. doi: 10.1109/ICITR54349.2021.9657389.

[8] Bayona-Oré, S., Limaymanta, C. H. C., & Talledo-Córdova, J. N., "Machine Learning to Predict Prices of Agricultural Products Targeted at Family Farming: A Literature Review," *WSEAS Transactions on Business and Economics*, vol. 18, pp. 920-929, 2021. doi: 10.37394/23207.2021.18.92. URL: https://wseas.com/journals/articles.php?id=MjY0MA==

[9] Singh, N., Kumar, P., Singh, K. P., Mishra, A., Singh, A. K., & Gupta, A. K., "A Review of Crop Price Prediction Using Machine Learning Techniques," *Journal of Electronic Systems*, vol. 20, no. 7s, pp. 2258-2269, 2024. URL: https://journalofelectronicsystems.com/index.php/jes/article/view/180

[10] Samuel, P., & Berg, A. T., "Crop Price Prediction Using Data Analytics and Machine Learning Algorithms," *Quest Journal [Repositories] Journal of Research in Agriculture and Animal Science*, vol. 6, no. 1, pp. 12-17, 2020. URL: https://questjournals.org/jraas/papers/vol6-issue1/C611217.pdf

[11] Zelingher, R., Yao, Y., & Tian, H., "Forecasting global commodity prices following crop production shocks with machine learning," *Environmental Research Letters*, vol. 19, no. 1, p. 014026, 2024. doi: 10.1088/1748-9326/ad0dda.

[12] Vellayudan, N. C., & Tagad, I. R. P., "Crop yield detection and price detection using a combined approach of machine learning," *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, vol. 8, no. 3, pp. 609-614, 2021. ISSN: 2230-9993. URL: https://ijsrcseit.com/paper/CSEIT2183110.pdf

[13] Elbasi, E., Zaki, R. M., Al-Ali, A., Zaki, C. P., Shukla, S., & Mathew, S., "Crop Prediction Using Soil Nutrients and Environmental Factors Based on Machine Learning Models," *Applied Sciences*, vol. 13, no. 16, p. 9288, 2023. doi: 10.3390/app13169288. URL: https://www.mdpi.com/2076-3417/13/16/9288

[14] Banerjee, S., Pinki, F. T., & Rahman, R. M., "A comparative study of machine learning and deep learning models for crop price estimation," *International Journal of Information Technology*, vol. 15, pp. 4291–4313, 2023. doi: 10.1007/s41870-023-01474-6.

[15] Cena's, P. V., "Forecast of Agricultural Crop Price using Time Series and Kalman Filter Method," Asia Pacific Journal of Multidisciplinary Research, vol. 5. no. 4, pp. 15-21, 2017

IJARETY

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975 Impact Factor: 8.152

www.ijarety.in editor.ijarety@gmail.com