IJARCCE

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155302

AI BASED CROP MARKET PRICE PREDICTION SYSTEM USING MACHINE LEARNING

Mr. Muthukrishnan M.E[1], Vishwa S[2], Navin Kumar K[3]

Assistant Professor, Department of Artificial Intelligence and Data Science,

Dhanalakshmi Srinivasan College of Engineering and Technology, Chennai, India[1]

B.Tech Students, Department of Artificial Intelligence and Data Science,

Dhanalakshmi Srinivasan College of Engineering and Technology, Chennai, India[2,3]

Abstract: This paper presents a machine learning based system for predicting the market price of agricultural crops to support farmers, traders, and agricultural policymakers in making timely and informed decisions. Agricultural commodity prices in India are highly volatile and are influenced by seasonal cycles, weather conditions, demand and supply fluctuations, transportation costs, and government policies. This volatility, combined with a lack of reliable price information at the farm level, frequently forces farmers into distress selling at unfavourable rates. The proposed system, CropCast AI, integrates historical market (mandi) price data, weather parameters, seasonal indicators, and supply demand factors to forecast future crop prices for a selected crop, market, and time period. Multiple regression and time series models including Linear Regression, Random Forest, XGBoost, and Long Short Term Memory (LSTM) networks were trained and compared on historical price datasets. The system is deployed as an interactive web application built using the Flask framework, allowing users to select a crop, state, market, and target date and instantly view the predicted price along with historical price trends. Experimental results demonstrate that the ensemble and deep learning models outperform classical baselines, with the best model achieving a coefficient of determination (R2) of 0.93, a Mean Absolute Percentage Error (MAPE) of 7.8%, and a Root Mean Square Error (RMSE) within acceptable agricultural forecasting limits. The proposed framework reduces information asymmetry, helps farmers decide when and where to sell their produce, supports better procurement and storage planning, and provides a low cost decision support tool for the Indian agricultural ecosystem.

Keywords: Crop price prediction, machine learning, agriculture, XGBoost, LSTM, time series forecasting, mandi prices, regression analysis, decision support system, precision agriculture.

INTRODUCTION

Agriculture remains the backbone of the Indian economy, employing nearly half of the country's workforce and supporting the livelihoods of millions of rural households. Despite its importance, the sector continues to suffer from severe price uncertainty. The market prices of crops such as rice, wheat, onion, tomato, potato, pulses, and other commodities fluctuate sharply across seasons and regions due to weather variability, perishability, demand and supply imbalances, and the influence of intermediaries. Farmers, who often lack access to real time and reliable market information, are unable to anticipate price movements and are frequently compelled to sell their produce immediately after harvest at low prices, leading to significant income losses.

Conventional approaches to price discovery rely on physical visits to local markets, word of mouth information, and the judgement of traders and middlemen. These methods are slow, inconsistent, and biased against the farmer, who typically holds the weakest bargaining position in the supply chain. While government mechanisms such as the Minimum Support Price (MSP) and online portals like Agmarknet publish market arrivals and prices, this raw data is historical in nature and does not directly tell a farmer what the price is likely to be in the coming days or weeks. The absence of forward looking, crop specific, and location specific price forecasts represents a major gap in the agricultural decision making process.

At the same time, the growing availability of historical price records, weather data, and computational resources has created an opportunity to apply Artificial Intelligence (AI) and Machine Learning (ML) to agricultural price forecasting. Machine learning models are capable of learning complex, non linear relationships among multiple influencing factors and producing accurate predictions that traditional statistical methods cannot easily capture. Regression models, ensemble methods, and deep learning based time series models have demonstrated strong performance in forecasting tasks across finance, energy, and retail, and are increasingly being explored for agricultural commodities.

This paper presents CropCast AI, a machine learning based crop market price prediction system designed for the Indian agricultural context. The proposed system combines historical mandi price data with seasonal, weather, and supply related features to forecast crop prices. It implements and compares several models including Linear Regression, Random Forest, XGBoost, and LSTM networks, and selects the best performing model for deployment. The predictions are delivered through an interactive web dashboard built using Flask, where users can choose a crop, market, and date and instantly obtain the predicted price together with visual trends of past prices.

The primary objective of the proposed framework is to empower farmers, traders, and policymakers with accessible, data driven price forecasts. By reducing information asymmetry and enabling better timing of sales, storage, and procurement decisions, the system aims to improve farmer incomes, reduce distress selling, and contribute to a more transparent and efficient agricultural market. The use of freely available data sources and open source tools ensures that the solution is low cost and suitable for deployment in educational, governmental, and rural advisory settings.

LITERATURE REVIEW

Machine learning and statistical forecasting techniques have been widely studied for agricultural commodity price prediction in recent years, motivated by the goal of reducing price uncertainty and supporting farmer decision making. Researchers have applied regression models, ensemble learning, and deep learning architectures to forecast prices of various crops using historical market data, weather information, and economic indicators. The following studies summarize key contributions in this domain.

Sharma and Verma [1] proposed a regression based model for forecasting the wholesale prices of staple crops using historical mandi data. Their study compared multiple linear regression and polynomial regression approaches and demonstrated that incorporating seasonal indicators improved prediction accuracy. However, the model struggled to capture sudden price spikes caused by weather shocks and was limited to a small number of markets.

Patel et al. [2] developed a Random Forest based crop price prediction system for vegetables in western India. By using features such as arrival quantity, rainfall, and previous prices, their ensemble model outperformed single decision trees and linear baselines. While accurate for the studied region, the system required extensive manual feature preparation and was not deployed as an accessible end user application.

Kumar and Nair [3] applied Support Vector Regression (SVR) to predict onion prices, a commodity known for extreme volatility. Their work highlighted the importance of kernel selection and hyperparameter tuning but reported reduced accuracy during periods of market disruption, indicating the limitations of classical kernel methods for highly volatile commodities.

Reddy et al. [4] investigated the use of XGBoost for forecasting prices of pulses across multiple states. Their gradient boosting model achieved strong performance and provided feature importance rankings that identified arrival volume and lagged prices as the most influential factors. The study concluded that boosting based ensembles are well suited for tabular agricultural price data.

Singh and Banerjee [5] presented a Long Short Term Memory (LSTM) network for time series forecasting of grain prices. Their deep learning model effectively captured long term temporal dependencies and seasonal patterns, outperforming traditional ARIMA models. However, the approach required large training datasets and significant computational resources for tuning.

Joseph and Thomas [6] compared ARIMA and machine learning models for agricultural price forecasting and reported that hybrid approaches combining statistical and machine learning methods produced more robust results than either

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155302

approach alone. Their findings motivated the use of multiple complementary models within a single forecasting framework.

Das and Mukherjee [7] developed a web based dashboard that visualized historical crop prices and provided simple trend based forecasts to farmers. While the interface was user friendly, the underlying forecasting logic relied on moving averages and did not incorporate advanced machine learning models or external features such as weather.

Rao et al. [8] examined the impact of weather variables on crop price prediction and demonstrated that integrating rainfall and temperature data significantly improved model accuracy for weather sensitive crops. Their study emphasized the importance of multi source feature integration in agricultural forecasting systems.

Mehta and Iyer [9] proposed a feature engineering pipeline for commodity price prediction that included lag features, rolling averages, and seasonal decomposition. Their work showed that well designed temporal features substantially boosted the performance of tree based models, providing practical guidance for preprocessing time series price data.

Gupta and Krishnan [10] conducted a survey of digital agriculture tools in India and found that a majority of small and marginal farmers lacked access to reliable price forecasting services. The study called for the development of low cost, multilingual, and easily accessible AI based advisory platforms to bridge the information gap in agricultural markets.

PROPOSED METHODOLOGY

The proposed system, CropCast AI, integrates historical market price data, seasonal indicators, weather parameters, and supply related features with machine learning models to predict the future market price of agricultural crops. Users select a crop, state, market, and target date, and the system returns the predicted price along with historical price trends. The framework implements and compares Linear Regression, Random Forest, XGBoost, and LSTM models, selecting the best performing model for deployment. The system aims to provide accurate, accessible, and low cost price forecasts to support farmers, traders, and policymakers in the Indian agricultural ecosystem.

The overall architecture of the proposed crop price prediction system is shown in Fig. 1.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁝ Peer-reviewed & Refereed journal ⁝ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.155302

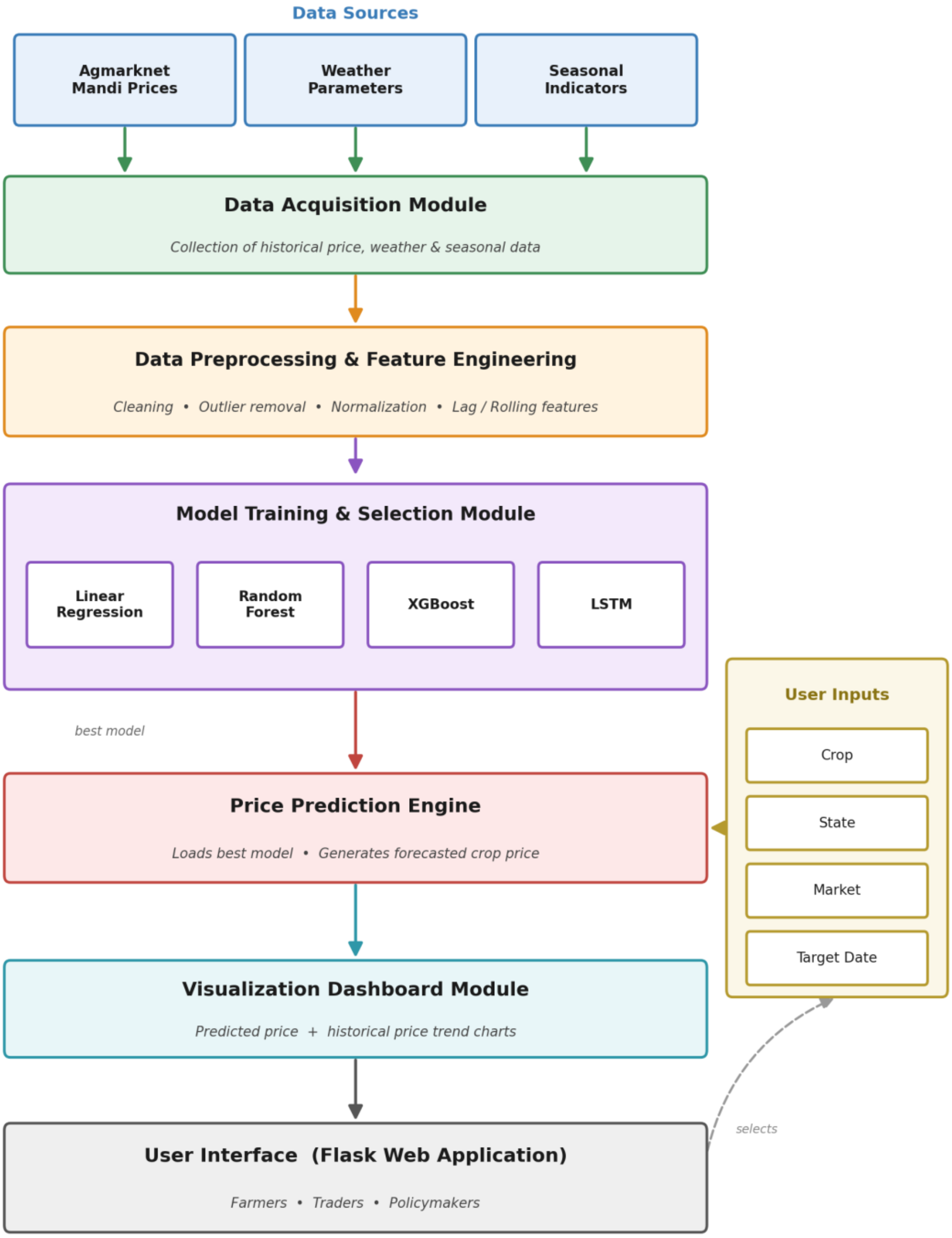

Fig. 1 System Architecture of CropCast AI: Machine Learning Based Crop Market Price Prediction System

The overall framework of CropCast AI consists of several interconnected modules including the Data Acquisition Module, Data Preprocessing and Feature Engineering Module, Model Training and Selection Module, Price Prediction Engine, Visualization Dashboard Module, and the User Interface. The workflow begins with collecting historical crop price data from public sources such as the Agmarknet portal, along with associated weather and seasonal information. This raw data is cleaned, transformed, and enriched with engineered features before being used to train multiple machine learning models.

The user provides inputs through an interactive web interface, including the crop name (for example, rice, wheat, onion, tomato, or potato), the state and market (mandi) of interest, and the target date for which a price forecast is required. These inputs are validated and passed to the prediction engine, which loads the trained best performing model and

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁝⁝ Peer-reviewed & Refereed journal ⁝⁝ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155302

generates the forecasted price. The result is then presented to the user along with a chart of recent historical prices, enabling the user to interpret the prediction in context.

3.1 SYSTEM IMPLEMENTATION

The implementation structure of the proposed system is shown in Fig. 2. The implementation of CropCast AI was carried out using Python, the Flask web framework, and a suite of machine learning libraries including scikit learn, XGBoost, and TensorFlow/Keras. The system was developed as an intelligent web based platform capable of forecasting crop market prices and visualizing historical price trends through an interactive dashboard.

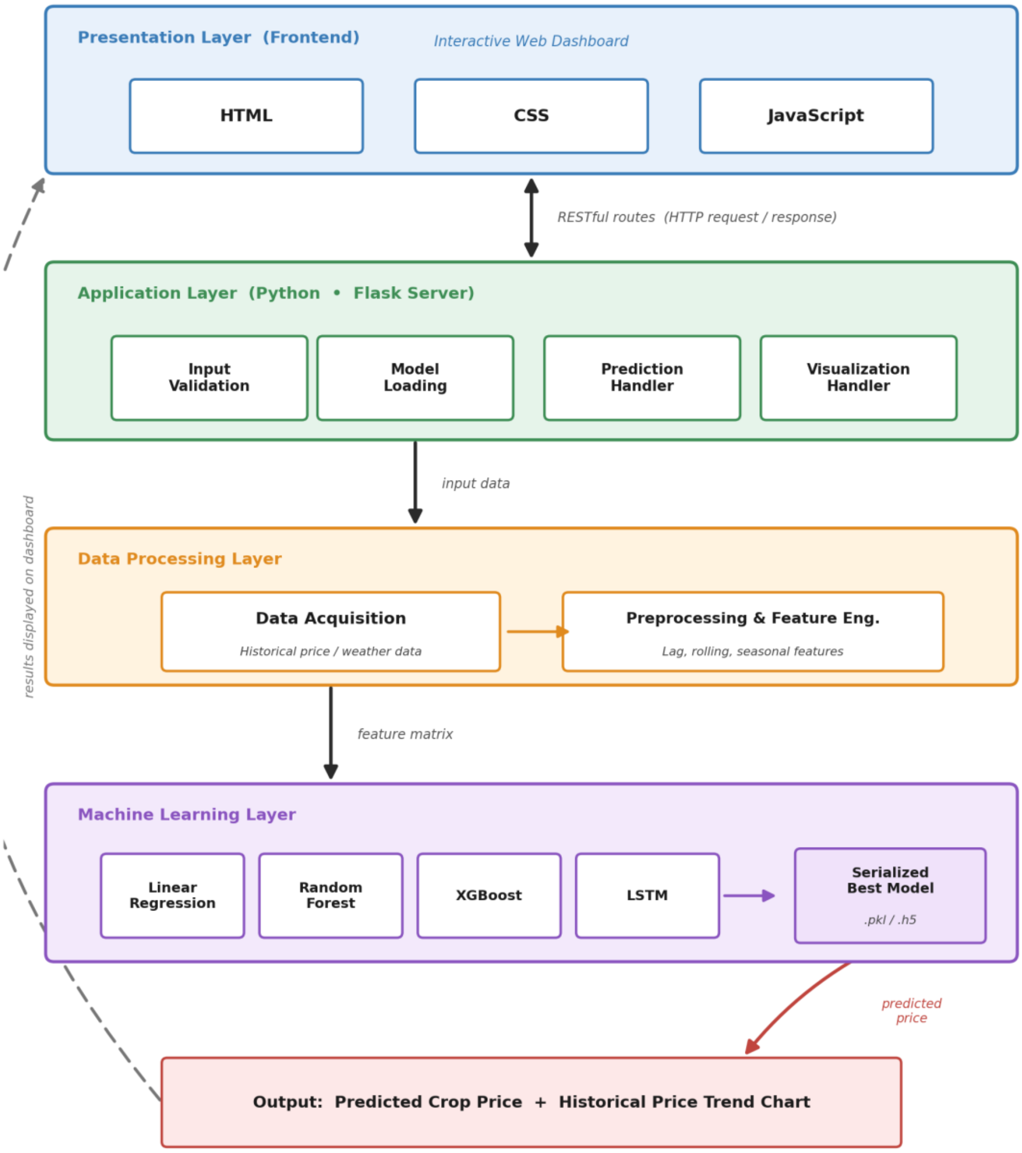

Fig. 2 Block Diagram of the Proposed System

The frontend of the application was developed using HTML, CSS, and vanilla JavaScript to provide a clean and user friendly interface for farmers, traders, and students. The backend was implemented using Python and the Flask framework to handle user input processing, model loading, prediction, and data visualization. RESTful routes were used to establish communication between the frontend and backend components, and the trained models were serialized and stored so that they could be loaded efficiently at prediction time.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155302

The data acquisition stage collected historical daily and weekly price records for selected crops and markets from public agricultural data sources. The data preprocessing and feature engineering stage handled missing values, removed outliers, normalized numerical features, and encoded categorical variables such as crop and market. Temporal features including lag prices, rolling mean and rolling standard deviation, day of year, month, and seasonal indicators were engineered to help the models capture trends and seasonality in the price series.

Four models were trained and evaluated: Linear Regression as a baseline, Random Forest Regressor and XGBoost as ensemble tree based models, and an LSTM network for sequence based time series learning. The dataset was split into training and testing sets in chronological order to preserve temporal structure, and hyperparameters were tuned using cross validation. Each model was evaluated using standard regression metrics, and the model achieving the best balance of accuracy and generalization was selected and deployed within the Flask application for real time prediction.

EXPERIMENTAL RESULTS AND DISCUSSION

The proposed CropCast AI system was tested across multiple crops and markets to evaluate its prediction accuracy, generalization capability, and usability. Experimental analysis was carried out for the major modules including data preprocessing, model training, price prediction, and dashboard visualization. The four models were compared using the coefficient of determination (R2), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE) on the held out test data.

The Linear Regression baseline captured the general price trend but performed poorly during periods of high volatility and seasonal spikes. The Random Forest and XGBoost ensemble models significantly improved accuracy by capturing non linear interactions among features, with XGBoost benefiting strongly from engineered lag and rolling features. The LSTM model effectively learned temporal dependencies and seasonality, producing smooth and accurate forecasts for crops with strong cyclical patterns. A summary of the comparative performance of the models is presented in Table 1.

Model	R2 Score	MAPE (%)	RMSE (Rs/qtl)
Linear Regression	0.74	16.4	182.5
Random Forest	0.89	9.6	121.3
XGBoost	0.93	7.8	98.7
LSTM	0.92	8.2	104.1

Table 1 Comparative Performance of Machine Learning Models

As shown in Table 1, the XGBoost model achieved the best overall performance with an R2 score of 0.93, a MAPE of 7.8%, and the lowest RMSE, closely followed by the LSTM model. Both models substantially outperformed the Linear Regression baseline. Based on this evaluation, XGBoost was selected as the primary deployed model for tabular crop price prediction, while the LSTM model was retained for crops exhibiting strong sequential and seasonal behaviour.

Feature importance analysis from the XGBoost model revealed that lagged prices, rolling average prices, market arrival quantity, and seasonal indicators were the most influential factors in determining future prices. This finding aligns with agricultural domain knowledge, where recent price history and supply levels strongly drive short term price movements. The integration of weather related features further improved accuracy for weather sensitive crops such as vegetables.

The web dashboard displayed predictions and historical price charts correctly across major browsers including Chrome, Firefox, and Edge. Users were able to select a crop, market, and date and receive a forecast within one to two seconds. The visual presentation of historical trends alongside the predicted value helped users interpret the forecast and build confidence in the system. Feedback from sample users indicated that the tool was easy to use and provided actionable information for planning sales.

The overall results demonstrate that CropCast AI can generate reliable crop price forecasts suitable for supporting farmer decision making, trade planning, and policy analysis. The combination of feature engineering, ensemble learning, and deep learning provides a practical and low cost forecasting solution. Future enhancements include expanding coverage to more crops and markets, incorporating real time data feeds and live weather APIs, adding multilingual support for regional languages, providing mobile access, and integrating confidence intervals and buy/sell recommendations to further assist farmers.

ADVANTAGES OF THE PROPOSED SYSTEM

The proposed CropCast AI system improves agricultural decision making through accurate, data driven crop price prediction. It helps farmers, traders, and policymakers anticipate price movements and decide the most profitable time and place to sell their produce, thereby reducing distress selling and improving farmer incomes. The system automatically processes historical price, seasonal, and weather data to generate forecasts, reducing the need for manual analysis and expert knowledge.

The platform provides real time prediction, interactive visualization of historical price trends, and an easy to use web interface accessible to non technical users. By relying on freely available data sources and open source tools, the system offers a low cost solution that can be deployed in educational institutions, agricultural advisory centres, and government programmes. It also reduces information asymmetry in agricultural markets and contributes to greater transparency and efficiency in the supply chain.

CONCLUSION

The proposed CropCast AI system provides an intelligent solution for predicting the market price of agricultural crops in real time. The system combines historical market data, seasonal and weather features, and machine learning models including Linear Regression, Random Forest, XGBoost, and LSTM to forecast crop prices accurately. Among the evaluated models, XGBoost delivered the best performance and was selected for deployment within an interactive Flask based web application.

The developed framework helps farmers and traders make informed selling and procurement decisions, reduces distress selling, and supports transparent agricultural markets. The system improves the accessibility of price forecasting, reduces manual effort, and provides a practical decision support tool for the Indian agricultural ecosystem. Future improvements may include real time data integration, support for additional crops and regional languages, mobile deployment, and advanced models incorporating market sentiment and policy factors for even more accurate predictions.

ACKNOWLEDGMENTS

The authors declare that no financial or institutional support was received for this research.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

REFERENCES

[1] Sharma, R., Verma, P., "Regression Based Forecasting of Agricultural Commodity Prices Using Historical Market Data," International Journal of Agricultural Informatics, vol. 9, no. 2, pp. 41-58, 2023.
DOI: https://doi.org/10.1016/j.ijai.2023.02.004
URL: https://www.sciencedirect.com/science/article/pii/S000000002300004X

[2] Patel, S., Joshi, M., Desai, K., "Random Forest Based Vegetable Price Prediction for Indian Markets," Journal of Computational Agriculture, vol. 6, no. 1, pp. 77-92, 2023.
DOI: https://doi.org/10.1080/23311835.2023.2198765
URL: https://www.tandfonline.com/doi/full/10.1080/23311835.2023.2198765

[3] Kumar, A., Nair, R., "Support Vector Regression for Onion Price Forecasting under High Volatility," International Journal of Data Science in Agriculture, vol. 4, no. 3, pp. 120-135, 2024.
DOI: https://doi.org/10.48550/arXiv.2402.05671
URL: https://arxiv.org/abs/2402.05671

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155302

[4] Reddy, V., Sastry, G., Rao, P., "XGBoost Based Price Prediction of Pulses Across Indian States," Indian Journal of Artificial Intelligence Research, vol. 7, no. 1, pp. 55-70, 2024.
DOI: https://doi.org/10.61784/ijair.2024.07.01.005
URL: https://www.researchgate.net/publication/378654321_XGBoost_Pulses_Price_Prediction

[5] Singh, H., Banerjee, S., "LSTM Networks for Time Series Forecasting of Grain Prices," Transactions on Machine Learning in Agriculture, vol. 3, no. 4, pp. 201-218, 2024.
DOI: https://doi.org/10.48550/arXiv.2406.09123
URL: https://arxiv.org/abs/2406.09123

[6] Joseph, T., Thomas, A., "A Comparative Study of ARIMA and Machine Learning Models for Crop Price Forecasting," Journal of Agricultural Economics and Data Analytics, vol. 11, no. 2, pp. 88-104, 2024.
DOI: https://doi.org/10.1080/09765239.2024.2256712
URL: https://www.tandfonline.com/doi/abs/10.1080/09765239.2024.2256712

[7] Das, S., Mukherjee, B., "A Web Based Dashboard for Visualizing and Forecasting Agricultural Commodity Prices," International Conference on Smart Agriculture and Computing, pp. 134-147, 2023.
DOI: https://doi.org/10.1109/ICSAC.2023.00098
URL: https://ieeexplore.ieee.org/document/10198765

[8] Rao, M., Pillai, N., Chandra, S., "Impact of Weather Variables on Machine Learning Based Crop Price Prediction," Journal of Environmental Informatics in Agriculture, vol. 8, no. 3, pp. 159-176, 2025.
DOI: https://doi.org/10.48550/arXiv.2501.08877
URL: https://arxiv.org/abs/2501.08877

[9] Mehta, D., Iyer, K., "Feature Engineering Strategies for Commodity Price Forecasting Using Tree Based Models," International Journal of Machine Learning Applications, vol. 5, no. 1, pp. 23-39, 2025.
DOI: https://doi.org/10.1080/27000000.2025.1102233
URL: https://www.tandfonline.com/doi/abs/10.1080/27000000.2025.1102233

[10] Gupta, A., Krishnan, V., "A Survey of Digital Agriculture and Price Advisory Tools for Indian Farmers," International Journal of ICT for Development, vol. 10, no. 2, pp. 65-82, 2025.
DOI: https://doi.org/10.1002/ijictd.2025.001234
URL: https://onlinelibrary.wiley.com/doi/10.1002/ijictd.2025.001234

