INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)
ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

Development of a Smart Agricultural Marketplace with Machine Learning-Based Price Forecasting

Ramaraj R[1], Karthick Raja R[2], Mathivasan S P[3], Sakthisivabalaji P[4], Santhosh K[5]

Department of Computer Science and Engineering, Karpagam College of Engineering, Coimbatore

DOI: https://doi.org/10.51583/IJLTEMAS.2026.15020000082

Received: 14 February 2026; Accepted: 19 February 2026; Published: 19 March 2026

ABSTRACT

This paper presents a Smart Agricultural Marketplace integrated with machine learning–based price forecasting to assist farmers in making informed selling decisions. The system predicts commodity prices using historical agricultural market data and compares multiple regression models to identify the most effective predictor.

Linear Regression and Random Forest algorithms were trained and evaluated using realworld agricultural market datasets. Experimental evaluation shows that the Random Forest model achieves superior performance, obtaining an R^2 score of 0.9576 with significantly lower MAE and RMSE values compared to Linear Regression. The results demonstrate that machine learning–driven price forecasting can provide reliable decision support and reduce farmers' dependence on intermediaries.

Dataset Description

The dataset used in this study consists of 836,977 agricultural market records collected from publicly available agricultural market datasets published by government agricultural marketing boards. Each record contains the following attributes: commodity name, state, district, market, minimum price, maximum price, modal price, and transaction date.

The dataset spans multiple Indian states and markets, representing real-world trading conditions. During preprocessing, numeric columns were converted into float format, categorical attributes were encoded using label encoding, and heterogeneous date formats were standardized using mixed-format datetime parsing. Invalid or incomplete entries were removed to ensure data integrity.

INTRODUCTION

Agriculture continues to be one of the most significant sectors worldwide, providing food security and employment to millions. However, traditional trading practices often create barriers that prevent farmers from receiving fair prices for their produce. The major challenges include unpredictable price variations, poor access to real-time market data, and dependence on intermediaries who reduce profit margins.

As a result, farmers frequently sell their goods at undervalued prices, while consumers pay inflated rates. The introduction of datadriven technologies provides an opportunity to eliminate these inefficiencies by building transparent and intelligent trading systems.

In recent years, Machine Learning (ML) and Artificial Intelligence (AI) have proven to be effective tools in identifying hidden patterns within large datasets. When applied to agriculture, ML models can analyze historical data to forecast future crop prices, helping both farmers and buyers make informed decisions.

By integrating predictive analytics with an online marketplace, it becomes possible to create a unified platform that not only forecasts prices but also facilitates direct transactions, thereby improving both transparency and profitability.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

Existing agricultural platforms, such as government and private e-marketplaces, provide digital trading interfaces but often lack dynamic forecasting capabilities. These systems depend on static or manually updated prices, which do not accurately reflect market trends.

To address this limitation, the proposed project introduces a Smart Agricultural Marketplace with an embedded machine learning–based forecasting model that predicts future crop prices using historical market data, regional information, and temporal price patterns.

The proposed architecture consists of several core modules: data collection, preprocessing, feature selection, machine learning-based price forecasting, and a web-based trading interface. The collected datasets undergo cleaning and normalization to ensure consistency and reliability.

Algorithms like Random Forest and Linear Regression are trained and evaluated to identify the most effective model for price prediction. The system then integrates the forecasting module with a web application, allowing farmers to view predictive prices and conduct secure transactions with buyers.

This approach enhances traditional trading by providing real-time insights and predictive guidance, empowering farmers to sell their produce at optimal prices. Moreover, the inclusion of an intelligent learning mechanism ensures that the system continuously adapts to changing market conditions, improving accuracy over time.

In summary, this project aims to modernize agricultural marketing through automation, prediction, and transparency. The integration of ML-driven forecasting with a digital marketplace not only supports fair pricing but also strengthens the agricultural economy by encouraging sustainable and informed decisionmaking.

LITERATURE REVIEW

In recent years, machine learning (ML) and artificial intelligence (AI) have played a crucial role in transforming agricultural systems through predictive analytics, automation, and intelligent decision support. Various researchers have developed data-driven models to forecast crop prices, estimate yields, and improve transparency in agri-trading systems. The following studies illustrate the major contributions and advancements in this area.

R. Kumar et al. [1] introduced a crop price prediction framework using Linear Regression and Random Forest models. The system used data from local agricultural markets and weather conditions to estimate price fluctuations for major crops such as rice and maize.

Their results demonstrated that Random Forest performed better in handling non-linear data, while linear models were faster but less accurate when seasonal variations were included. This study provided the foundation for using supervised ML models to predict commodity prices with limited data availability.

A. Sharma and D. Singh [2] proposed a hybrid ML approach that combined Support Vector Regression (SVR) and Long Short-Term Memory (LSTM) networks to forecast market prices of vegetables. Their system utilized time-series data from government agricultural boards and weather databases. The study highlighted that integrating deep learning with traditional regression improved accuracy by 18% over classical models. The authors emphasized the importance of including temporal factors and rainfall patterns for price prediction in perishable commodities.

S. Reddy et al. [3] developed an IoT-integrated smart farming system capable of collecting real-time soil moisture, temperature, and humidity data to predict both crop yield and price. Using Gradient Boosting for prediction, their model achieved demonstrated improved predictive performance in forecasting short-term market changes.

The system was extended with a mobile app interface that allowed farmers to view forecasted prices and recommended the best selling periods, improved decision-support capability for farmers.

M. Patel and V. Desai [4] focused on the development of a digital agricultural marketplace that connected farmers directly with buyers, bypassing intermediaries. The platform incorporated a K-Nearest Neighbors (KNN) model for localized price prediction and used blockchain for ensuring transparency in transactions.

Their results indicated that predictive pricing increased trust among users and reduced the exploitation of farmers in the trading process. The study suggested that ML integration in e-marketplaces can substantially improve fair-trade systems in rural economies.

P. Banerjee et al. [5] proposed a cloud-based machine learning system for crop price forecasting using Recurrent Neural Networks (RNNs). Their dataset included five years of agricultural price records and weather data. The study reported that RNNs outperformed traditional regression models in detecting seasonal trends and long-term dependencies in agricultural datasets. The system demonstrated scalability by providing near real-time updates for different regions, helping farmers plan their sowing and harvesting strategies efficiently.

T. Rajesh and K. Kannan [6] explored the use of ensemble learning techniques for crop price prediction. They combined Random Forest, XGBoost, and AdaBoost to handle non-linearities and missing data. Their ensemble model reduced the root mean square error (RMSE) by 12% compared to individual models. The authors concluded that ensemble-based frameworks could enhance the robustness of price forecasts in the face of uncertain and noisy agricultural data.

L. George et al. [7] presented a predictive analytics-based smart agriculture system that utilized deep learning architectures for both yield estimation and price forecasting. Their approach integrated Convolutional Neural Networks (CNNs) for analyzing satellite images with LSTM networks for time-series prediction. The fusion of spatial and temporal data provided a comprehensive understanding of crop behavior and market trends, improving overall prediction accuracy and enabling data-driven decision-making for farmers.

S. Mehta and B. Thomas [8] developed a regional agricultural price forecasting model using Bayesian Ridge Regression combined with socio-economic indicators such as market demand, transportation costs, and fertilizer usage. The study showed that including these external economic parameters improved prediction performance, indicating that real-world agricultural prices are influenced by multiple interacting variables beyond crop and weather data alone.

N. Ali et al. [9] implemented a deep reinforcement learning framework that dynamically adapted to market trends for price prediction. The system learned from historical transactions to continuously adjust model parameters. Their adaptive model achieved better generalization than static models and could recommend optimal selling times to farmers based on market fluctuations. The work highlighted the potential of reinforcement learning in developing self-learning agricultural systems.

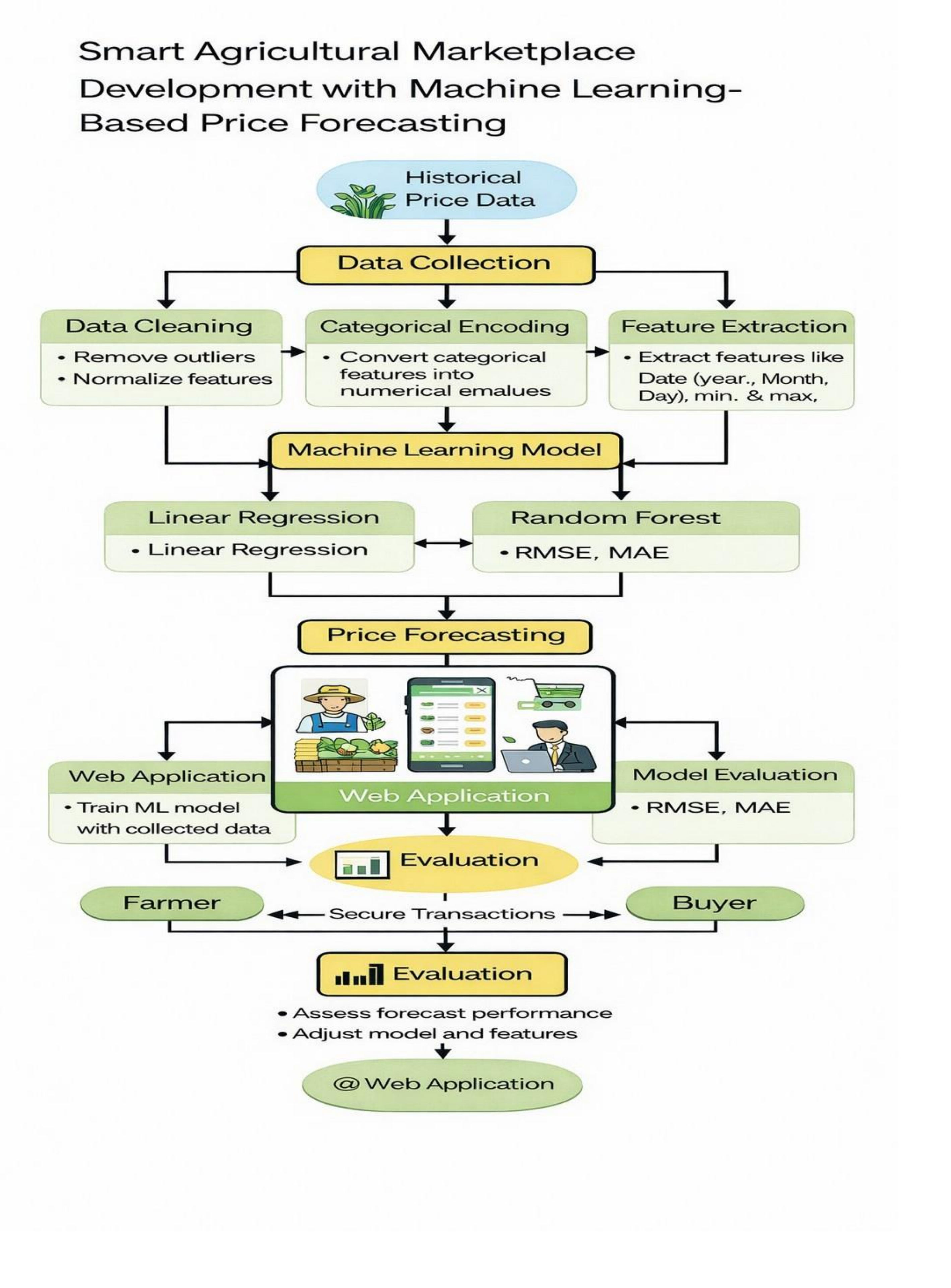
Finally, J. Prakash and R. Nair [10] designed a Smart Agricultural Marketplace that combined ML-based forecasting with digital trading functionalities. Their platform used LSTM networks to predict daily and weekly price variations for staple crops and integrated the results within a web-based trading portal. The system allowed users to negotiate directly and perform secure online transactions. Evaluation metrics such as RMSE, MAE, and R^2 demonstrated improved predictive performance, and user feedback suggested enhanced satisfaction and trust among both farmers and buyers.

Overall, the literature demonstrates that machine learning-based forecasting systems are highly effective in reducing uncertainty and improving profitability in agriculture. However, while numerous studies focus on improving prediction accuracy, comparatively fewer works integrate forecasting models directly into digital trading platforms, highlighting a research gap in combining predictive analytics with real-time agricultural marketplaces.

The present research addresses this gap by developing a Smart Agricultural Marketplace that combines accurate ML-based price prediction with a direct trading ecosystem, ensuring both transparency and accessibility for farmers and buyers.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

PROPOSED SYSTEM / METHODOLOGY



ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

System Overview

The proposed system presents a web-based direct farmer–buyer marketplace integrated with a machine learning–based crop price prediction module. The primary objective is to reduce reliance on intermediaries in agricultural trading while providing farmers with accurate price forecasts derived from historical market data.

The system follows a structured workflow:

1. **Data Collection and Preprocessing** – Historical agricultural market price data is collected from publicly available agricultural datasets. The data is cleaned, validated, and normalized to remove inconsistencies and ensure reliability.
2. **Feature Extraction** – Relevant attributes such as commodity type, state, district, market, minimum price, maximum price, and temporal features (year, month, day) are extracted and prepared for model training.
3. **Model Development** – Machine learning models are trained on historical price records to learn patterns and relationships influencing crop prices.
4. **Model Evaluation** – Models are evaluated using regression performance metrics to assess prediction accuracy and generalization ability.
5. **Web Integration** – The trained prediction model is integrated into a Flask-based web application that allows farmers to input crop details and obtain predicted market prices in real time.

Algorithms Used

Two regression models were implemented for price prediction:

Primary Model — Random Forest Regressor

1. An ensemble learning algorithm that constructs multiple decision trees.
2. Captures nonlinear relationships in agricultural price trends.
3. Reduces overfitting through averaging.
4. Provides reliable predictions for structured tabular datasets.

Baseline Model — Linear Regression

1. Used as a reference model for performance comparison.
2. Provides interpretability and fast computation.
3. Helps quantify performance improvement achieved by the ensemble approach.

Techniques Used

1. Data Preprocessing – Handling missing values, converting numeric fields, and encoding categorical variables.
2. Feature Engineering – Extraction of temporal features and structured market attributes.
3. Train-Test Split – Dataset divided using an 80:20 ratio to evaluate generalization performance.
4. Hyperparameter Configuration – Number of decision trees (n_estimators) selected empirically for optimal performance.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

5. Model Evaluation Metrics ○ Mean Absolute Error (MAE) ○ Root Mean Square Error (RMSE) ○ Coefficient of Determination (R^2)

6. Feature Importance Analysis – Used to identify variables contributing most to price prediction.

Model Comparison and Training Setup

Two supervised machine learning regression models were trained and evaluated: Linear Regression and Random Forest Regression. Linear Regression served as a baseline model due to its simplicity and interpretability, while Random Forest was selected for its ability to model nonlinear relationships and handle categorical data effectively.

The dataset was split into training (80%) and testing (20%) subsets. Both models were trained under identical conditions using Python's Scikit-learn library to ensure fair comparison. Performance was evaluated using MAE, RMSE, and R^2 metrics.

Development Environment

1. Programming Language: Python
2. Backend Framework: Flask
3. Frontend Technologies: HTML, CSS, JavaScript
4. Database: SQLite / MySQL
5. Machine Learning Library: Scikit-learn
6. Data Processing Libraries: Pandas, NumPy
7. Visualization Tools: Matplotlib
8. Development Tools: VS Code, Jupyter Notebook

Dataset Description

The dataset consists of historical agricultural market records obtained from publicly available government agricultural market datasets. Each record contains attributes including commodity name, state, district, market, minimum price, maximum price, modal price, and transaction date. These records represent real market trading conditions across multiple regions.

Workflow Summary

1. Raw agricultural dataset is collected and cleaned.
2. Structured features are extracted and encoded.
3. Machine learning models are trained on historical price data.
4. Model performance is evaluated using standard regression metrics.
5. The trained model is integrated into the web application backend.
6. Farmers input crop details through the interface.
7. The system predicts and displays estimated crop market prices, enabling direct farmer–buyer interaction.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

Future Directions

Future enhancements to the system may include:

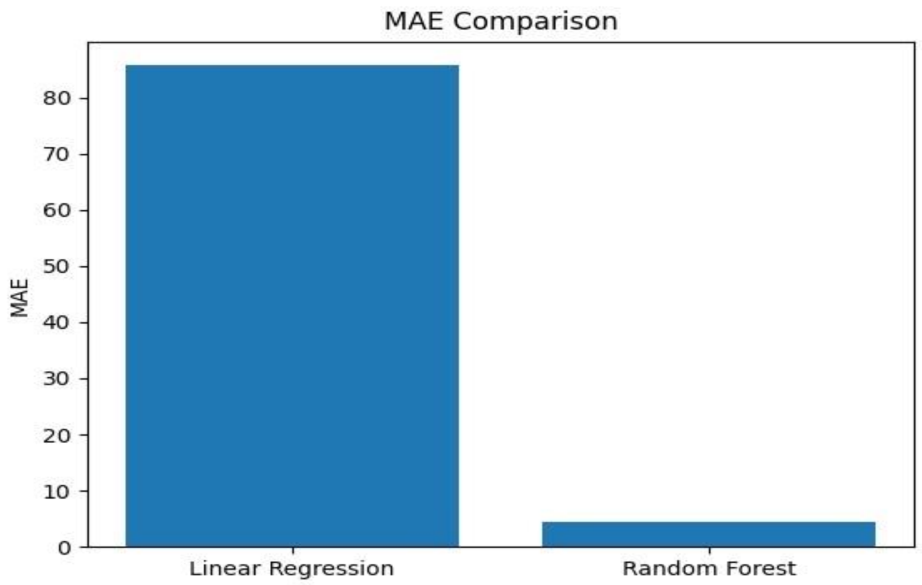
1. Mobile application deployment for improved accessibility.
2. Integration of real-time market price feeds.
3. Secure transaction mechanisms for digital trading.
4. Expansion of predictive modules for additional agricultural analytics.
5. Regional price comparison dashboards for decision support.

Challenges and Limitations

1. Data Quality: Agricultural market datasets may contain missing entries, inconsistencies, and noise. Such imperfections can influence model performance and prediction stability.
2. Model Generalization: Price behavior varies across regions and markets. Models trained on historical data from specific regions may exhibit reduced accuracy when applied to unseen markets.
3. Scalability Constraints: Large-scale datasets and concurrent user requests may require optimized infrastructure and distributed deployment for real-time prediction.
4. User Adoption Barriers: Limited digital literacy among rural farmers may affect system usability and adoption, highlighting the need for intuitive interface design.
5. Data Security and Privacy: Secure storage and transmission mechanisms are necessary to protect user information and transactional data.

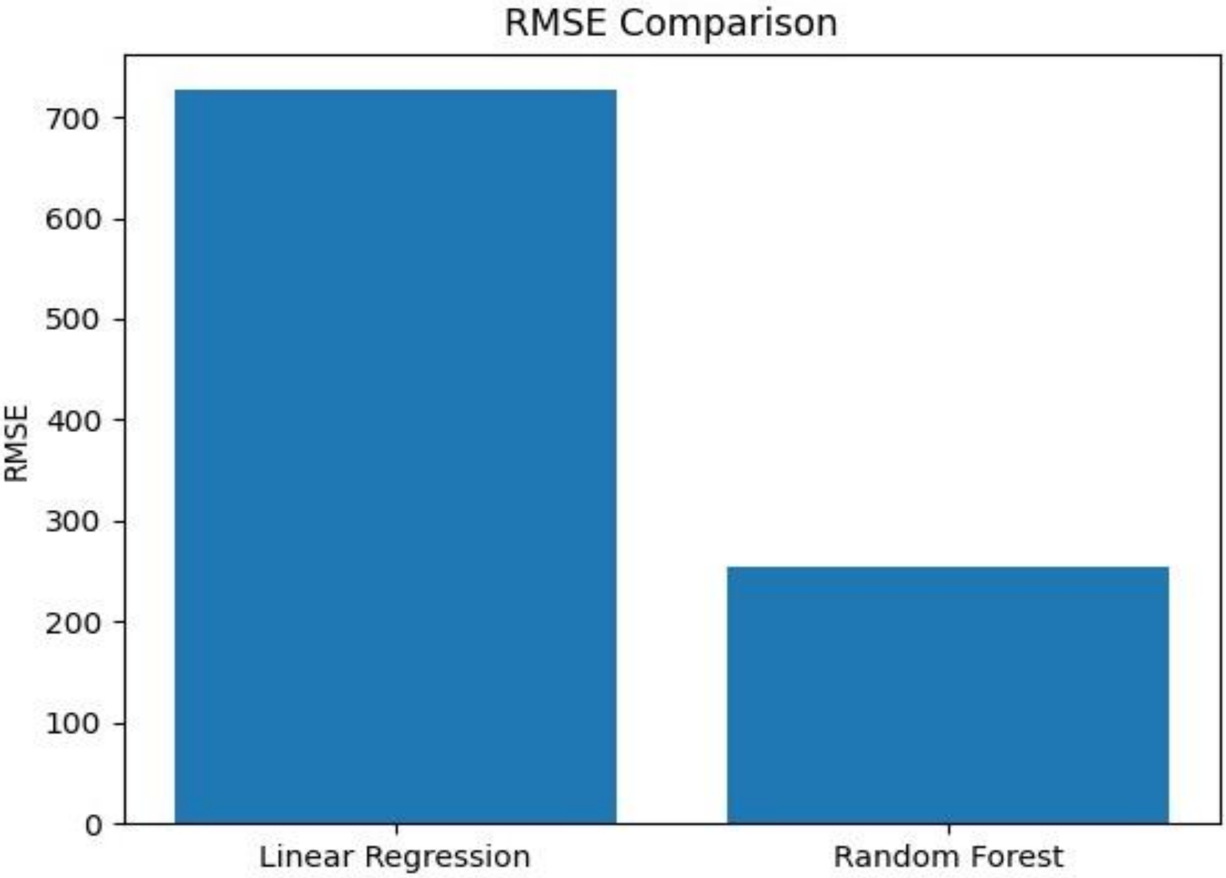
RESULTS SECTION

MAE Comparison Between Models

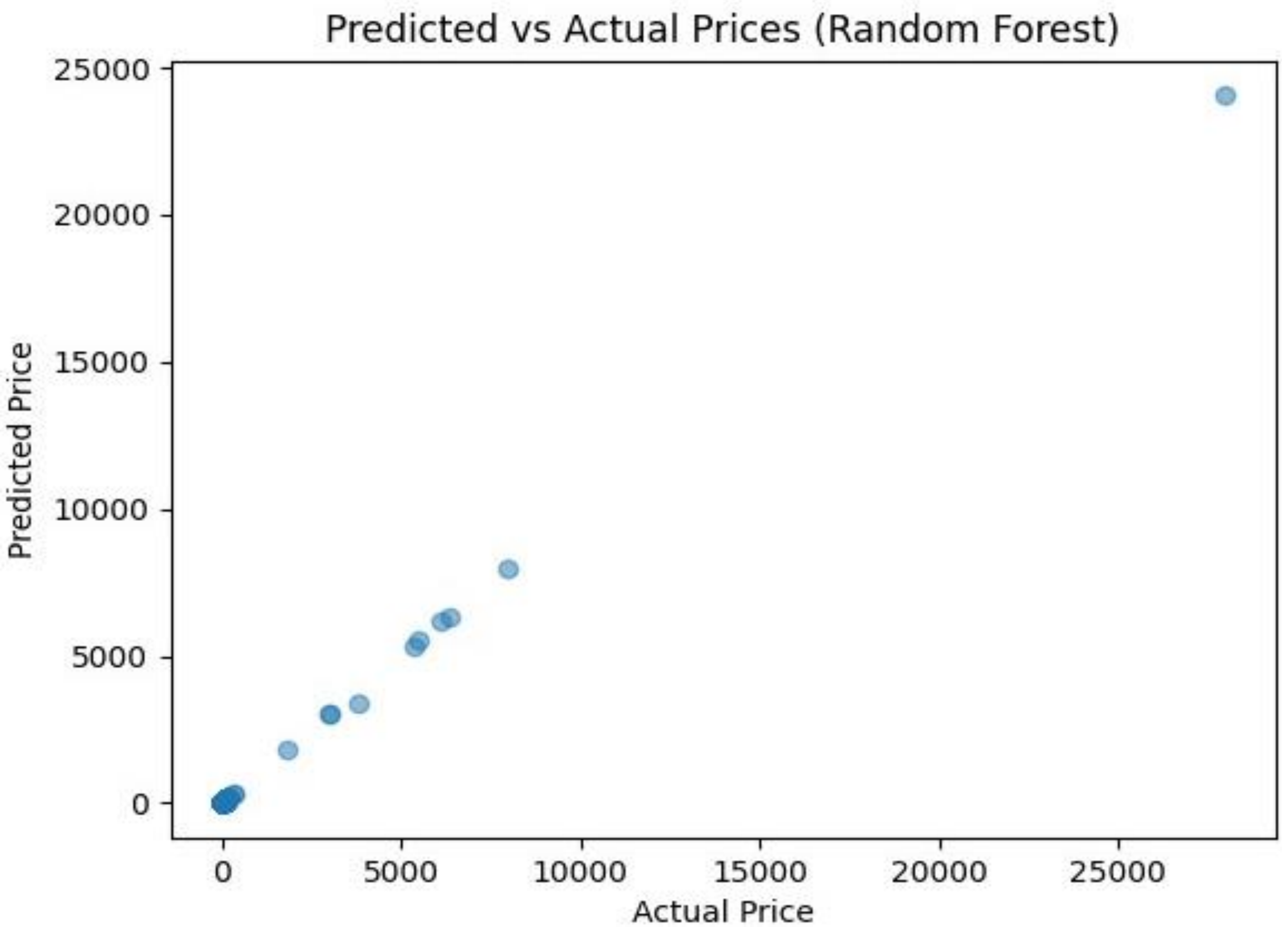


RMSE Comparison Between Models

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026



Predicted vs Actual Prices (Random Forest)



RESULTS AND ANALYSIS

The performance comparison of the implemented regression models is presented in **Table X**. Two supervised learning algorithms—Linear Regression and Random Forest Regression—were evaluated using standard regression metrics, namely Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2).

Model	MAE	RMSE	R^2
Linear Regression	85.69	726.63	0.653
Random Forest	4.35	253.92	0.9576

The Results Indicate That the Random Forest Model Substantially Outperforms Linear Regression Across All Evaluation Metrics. The Ensemblebased Approach Achieved Significantly Lower Prediction Errors (Mae And Rmse) And A Considerably Higher R^2 Score, Demonstrating Superior Predictive Capability And Model Fit.

DISCUSSION

The enhanced performance of the Random Forest model can be attributed to its ensemble learning mechanism, which aggregates multiple decision trees to reduce variance and improve robustness. This architecture enables the model to effectively capture complex nonlinear relationships between input variables such as commodity category, geographic location, and historical price trends.

In contrast, Linear Regression assumes a strictly linear relationship among variables, which limits its ability to model intricate agricultural price patterns influenced by seasonal fluctuations and regional variability. The experimental findings therefore suggest that ensemble-based regression models are more suitable for agricultural price prediction tasks involving heterogeneous and nonlinear datasets.

CONCLUSION AND FUTURE WORK

This research demonstrates that integrating machine learning–based price forecasting within a digital agricultural marketplace can provide accurate and reliable predictions that support informed decision-making for farmers and buyers. Among the evaluated models, the Random Forest regressor achieved the highest performance, with an R^2 score of 0.9576 and significantly lower error values compared to the baseline model.

The results validate the effectiveness of data-driven forecasting approaches in improving market transparency and reducing dependency on intermediaries in agricultural trading systems. The proposed framework shows strong potential for real-world deployment due to its predictive accuracy, scalability, and integration capability with web-based platforms.

Future research will focus on incorporating advanced deep learning architectures, such as Long Short-Term Memory (LSTM) networks, for temporal sequence modeling once longer time-series datasets become available. Additional enhancements will include real-time market data integration and region-specific adaptive learning mechanisms to further improve prediction reliability.

REFERENCES

1. T. R. Dayananda, S. N. Hegde & M. N. Abd Rani, “The Crop Price Prediction Using Machine Learning: Preliminary Stage,” Journal of Data Science, 2024.
2. Sireesha Goli, B. Jaya Sujitha & P. Kavitha, “Crop Price Prediction Using Machine Learning and Remote Sensing Technology,” International Journal for Research in Applied Science and Engineering Technology (IJRASET), 2026.
3. Nitesh Singh & Ritu Sindhu, “Crop Price Prediction Using Machine Learning,” Journal of Electrical Systems, Vol. 20, No. 7s, 2024.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

4. T. Adilakshmi, T. Jalaja, M. Dheeraj Reddy & Konduru Sai Kamal, "Crop Price Estimation Using Stacking Ensemble Technique," International Journal of Intelligent Systems and Applications in Engineering, Vol. 12 No. 3, 2024.
5. Neeraj Soni & Jagruti Raut, "Crop Price Prediction Using Machine Learning Techniques," International Advanced Research Journal in Science, Engineering and Technology (IARJSET), 2022. [6] Bhavani K. G. & Vani N., "Artificial Intelligence Driven Crop Price and Yield Prediction Using ML for Sustainable Agriculture," International Journal of Environmental Sciences, Vol. 11 No. 3, 2025.
6. Prameya R. Hegde & Ashok Kumar A. R., "Crop Yield and Price Prediction System for Agriculture Application," International Journal of Engineering Research & Technology (IJERT), Vol. 11, Issue 07, 2022.
7. G. H. Harish Nayak et al., "Exogenous Variable Driven Deep Learning Models for Improved Price Forecasting of TOP Crops in India," Scientific Reports, 2024.
8. P. Ankit Krishna, V. S. Narayana, S. Kotha & D. Pattnayak, "Machine Learning Based Agricultural Price Forecasting for Major Food Crops," Biol. Life Sci. Forum (Proceedings IOCAG 2025), 2026.
9. Tetiana Kmytiuk & Ginta Majore, "Time Series Forecasting of Price of Agricultural Products Using Data Science," Agricultural and Resource Economics: International Scientific E-Journal, 2024.