Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A Data-Driven Machine Learning Approach for Agricultural Commodity Price Forecasting in India

Pooja Mishra[1], Dr. Rashmi Deshpande[2] , Dr. Suvarna Patil[3] ,Deepali Jawale[4]
[1]*Dr. D Y Patil Institute of Engineering, Management and Research,Pune.*
[2]*Dr. D Y Patil International University,Pune.*
[3]*Dr. D Y Patil International University,Pune.*
[4]*Pimpri Chinchwad College of Engineering,Pune.*

ARTICLE INFO

Received: 18 Dec 2024

Revised: 10 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

The agricultural sector plays a crucial role in sustaining the global economy; however, it faces significant challenges due to volatile markets and fluctuating crop prices. Such uncertainties impede the ability of farmers, traders, and policymakers to make informed decisions, resulting in increased financial risks and inefficiencies. To address these issues, this study presents a data-driven price forecasting model that integrates demand–supply patterns, historical crop prices, current market data, and price fluctuation trends to generate accurate and reliable price predictions. The system provides stakeholders with actionable insights: farmers receive guidance on resource allocation, crop selection, and optimal timing for harvesting and sales; traders benefit from improved inventory management and investment strategies; and policymakers can leverage the forecasts to stabilize markets and safeguard vulnerable participants. Beyond decision-making, the model promotes efficient resource utilization and supports sustainable farming practices. Additionally, it provides a strong foundation for the integration of advanced technologies, such as artificial intelligence (AI) and machine learning (ML), in agriculture.Experimental evaluation on tomato commodity prices demonstrates that the model achieves a Mean Squared Error (MSE) of 25.87, Mean Absolute Error (MAE) of 4.78, Root Mean Squared Error (RMSE) of 5.09, an Akaike Information Criterion (AIC) of 16.58, and an R^2 score of 0.72, indicating that the model explains 72% of the variance in the dataset. These results reflect a balance between accuracy and simplicity, with comparatively low prediction errors for commodity price forecasting. Furthermore, comparative analysis with baseline forecasting techniques such as ARIMA and simple moving average revealed consistently lower error scores, establishing the model's efficacy for crops with relatively stable price trends.

Keywords: Agriculture, Price Forecasting, Artificial Intelligence, Machine Learning, Market Stability, Predictive Analytics, Sustainable Farming, Time Series Analysis.

INTRODUCTION

The agricultural sector faces major challenges due to market price volatility, even though it is an essential part of the global economy. Despite being vital to the global economy, the agriculture sector faces a number of challenges due to unstable market prices. Among the factors influencing fluctuating agricultural prices are supply and demand imbalances, market speculation, and unpredictable weather patterns. The purpose of this study is to use machine learning techniques to develop a predictive model for forecasting agricultural commodity prices. By looking at historical price data, current market trends, and demand-supply dynamics, the method aims to generate precise price forecasts. Farmers can use these projections to determine which crops to plant and when to sell, and policymakers and merchants can use the data to better manage resources and stabilize the market. The model will be constructed using machine learning techniques and trained on a dataset of historical market data that comprises components such as global market conditions, seasonal fluctuations, and regional trends. Techniques like feature extraction, data preparation, and model evaluation will be used to improve prediction accuracy. The system will also incorporate real-time market data to ensure that the forecasts remain relevant and up to date. The system's scope may eventually expand to include more agricultural products and geographic regions with the integration of IoT devices for real-time market condition monitoring. The system will help farmers, businesspeople, and lawmakers manage the intricacies of the agricultural market, reduce financial risks, and increase the overall productivity of the sector. By combining machine learning and agricultural economics, this project aims to provide a long-term solution for enhancing farming methods and promoting market stability.

MOTIVATION

The difficulties faced by agricultural stakeholders as a result of volatile market prices are what inspired this concept. Price fluctuations affect farmers, dealers, and policymakers and can lead to bad choices and financial risks. This project aims to simplify agricultural price forecasting and generate accurate forecasts that help stakeholders make informed decisions by utilizing machine learning. By combining technology and agricultural data, the project promotes better resource management, reduces risks, and enhances market stability for all parties involved.

The hybrid architecture of this model also gives it a significant edge. The fuzzy logic used in this model brings human alike reasoning to qualitative evaluation that allows behaviour to be labelled accordingly such as highly engaged, moderately active or infrequent. This qualitative insight complements the machine learning component thus also focusing on regular pattern recognition and predictive scoring. The system also generates visual feedback reports giving instructors and students a much easier way to understand the breakdown of performance, strengths and areas for improvement.

PROBLEM STATEMENT

This project's objective is to create a machine learning-based system that uses demand-supply analysis, historical data, and current trends to predict agricultural market prices with accuracy.. By filling the gap in price prediction tools, this system will help farmers, traders, and policymakers make better decisions, lower financial risks, and enhance market stability. This project aims to provide actionable insights that maximize resource allocation and advance sustainable agricultural practices by utilizing machine learning.

OBJECTIVES

1. Develop a machine learning-based system to predict agricultural market prices using historical data, demand-supply trends, and real-time market information.
2. Implement a user-friendly interface to display accurate price predictions, helping farmers, traders, and policymakers make informed decisions.
3. Optimize the model for scalability, enabling the inclusion of more agricultural products and regions in future iterations.

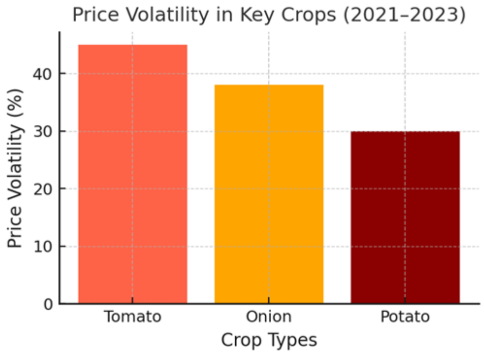
PROJECT SCOPE

By examining demand-supply dynamics, historical data, and current market trends, this project seeks to create a machine learning-based system that can reliably forecast agricultural market prices. In order to enable farmers, traders, and policymakers to make well-informed decisions, the system will be built to offer accurate and useful price forecasts for a range of agricultural products, The creation of an intuitive platform that allows stakeholders to access precise price forecasts and make informed decisions about crop planting, inventory control, and market interventions is a crucial component of the project. The system will guarantee that forecasts stay current and represent the state of the market by incorporating real-time market data. The project will initially concentrate on a small number of agricultural products, but it may eventually broaden to include additional crops and geographical areas. The system will also be scalable, enabling future integrations with Internet of Things devices to offer real-time market condition monitoring. The ultimate objective is to encourage more sustainable and effective farming methods while lowering the financial risks related to price volatility.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

REAL-WORLD EVIDENCE



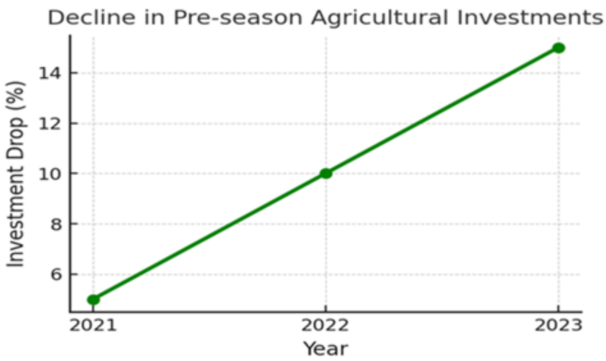


Figure 1. Price Volatility in Key Crops

Fig 2. Decline in Pre-season Agricultural Investments

Figure 1. Price volatility in key crops describes bar chart of price volatility for Tomato, Onion, and Potato demonstrates substantial volatility during the 2021–2023 period, with Tomato reaching ~45%, Onion ~38%, and Potato ~30%. This pronounced volatility motivates the development of predictive systems to provide farmers with more stable decision support for input timing, crop selection, and marketing strategies.

Figure 2 Figure 2 shows a rising trend in the percentage decline of pre-season investments, from 5% in 2021 to 15% in 2023. This indicates increasing financial risk perceptions among farmers and highlights the need for tools and policies that reduce upfront cost barriers and improve liquidity for smallholders.

LITERATURE REVIEW

Agricultural price prediction has been a critical research focus due to its direct impact on farmers, traders, policymakers, and supply chain stakeholders. Various approaches have been employed, ranging from conventional statistical methods to advanced machine learning and deep learning models.

1. Time-Series Forecasting Techniques

Traditional statistical techniques, such as AutoRegressive Integrated Moving Average (ARIMA), have been extensively applied to agricultural price forecasting. Mandal et al. [9] demonstrated the application of ARIMA in predicting vegetable prices in local markets and found it effective for short-term forecasts. However, ARIMA-based models require stationary data and are often unable to capture non-linear dependencies or sudden fluctuations in agricultural prices [5], [7].

2. Machine Learning Approaches

Machine learning algorithms have gained prominence in modeling non-linear and complex patterns in agricultural price data. Linear regression has been used as a baseline model for trend analysis, particularly in markets with relatively low volatility. More advanced models such as Support Vector Regression (SVR) and Random Forest Regression (RFR) have been reported to outperform conventional time-series techniques. For instance, Kumari et al. [10] applied SVR and Random Forest models to forecast wheat prices in India, achieving higher accuracy compared to ARIMA. Similarly, Singh and Patel [7] employed Random Forest models for price forecasting using climate variables, highlighting the significance of external environmental factors.

3. Deep Learning Models

Recent advancements in deep learning have further enhanced agricultural forecasting capabilities. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) architectures have been successfully applied to capture sequential and temporal dependencies in price data [3], [6]. Patel and Srivastava [11] demonstrated the effectiveness of LSTM models in forecasting onion prices in Maharashtra, reporting superior performance compared to traditional methods. Furthermore, hybrid approaches combining Convolutional Neural Networks (CNNs) with LSTMs have

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

been adopted to integrate auxiliary inputs such as rainfall and temperature, thereby improving prediction accuracy [3], [4].

Overall, the literature indicates a clear transition from traditional time-series models toward machine learning and deep learning-based frameworks, with hybrid models emerging as promising solutions for robust and accurate agricultural price prediction.

MODEL IMPLEMENTATION

The implementation of agricultural price prediction is carried out in a structured sequence comprising data collection, preprocessing, model training, evaluation, and visualization.

A. Data Collection

The dataset is obtained from the Agmarknet portal [12], a government platform providing daily agricultural commodity prices across India. Attributes include date, commodity, market location, and price. Data is downloaded in CSV format and filtered for specific commodities such as tomato and onion.

B. Data Preprocessing

The dataset undergoes several preprocessing steps:

- Conversion of date into datetime format.
- Filtering for commodity-specific entries.
- Chronological sorting.
- Handling missing values through removal or imputation.
- Feature engineering by transforming dates into numerical values (e.g., days since the first record).

C. Feature and Label Selection

The problem is framed as a regression task where:

X=Number of days since first entry,Y=Commodity Price (₹)

D. Dataset Splitting

The dataset is divided into Training Set (80%) and Testing Set (20%) to evaluate generalization ability.

E. Model Training (Linear Regression)

Linear Regression assumes a linear relationship between the independent variable X and dependent variable Y. The regression equation is:

$$\hat{Y} = \beta_0 + \beta_1 X$$

F. Model Evaluation

Model performance is assessed using error metrics:

- Mean Squared Error (MSE)
- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)

G. Visualization of Results

A time-series graph of actual vs. predicted prices is plotted to visually interpret prediction accuracy.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

SYSTEM ARCHITECTURE

The proposed agricultural price prediction system is designed as a layered architecture to ensure modularity, scalability, and real-time data integration. The architecture comprises five primary components:

1. **Data Collection Layer**
   - Collects historical crop price data from government portals (e.g., Agmarknet) and real-time updates from APIs.
   - Inputs include commodity type, location, date, and price.
2. **Data Preprocessing Layer**
   - Handles data cleaning, missing value treatment, feature extraction, and transformation of time-series attributes (e.g., converting date into numerical indices).
   - Ensures data consistency for model training.
3. **Machine Learning Model Layer**
   - Implements regression-based prediction (Linear Regression in the current model).
   - Future extensions include Random Forest, SVR, ARIMA, and LSTM.
   - Generates price forecasts (predicted price index and estimated market price).
4. **Evaluation Layer**
   - Evaluates model performance using metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2 Score.
   - Supports model comparison for optimal accuracy.
5. **Visualization and User Interface Layer**
   - Provides stakeholders with an interactive dashboard to view predicted prices, historical trends, and decision-support insights.
   - Farmers, traders, and policymakers can access forecasts for informed planning.

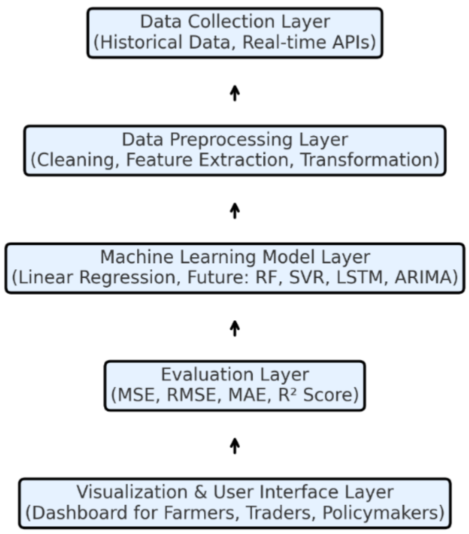


Figure 3. Proposed System Architecture for agriculture price prediction

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

OUTPUT RESULTS AND EVALUATION METHODOLOGY

A. Objective of Evaluation

The evaluation aims to measure the performance of the proposed Linear Regression-based agricultural price prediction system. The focus is on:

1. Prediction accuracy of the model
2. Quantitative error measurement using statistical metrics
3. Model interpretability for practical usability

B. Dataset Description

The dataset used in this study consists of historical agricultural price index values for selected crops.

- **Time Period:** 2015–2021
- **Attributes:**

1. Crop Name
2. Year
3. Price Index (%)
4. Base Market Price (₹/kg)

C. Parameters Considered

Table I. Parameters Used in Model Implementation

Parameter	Description
Independent Variable (X)	Year (2015, 2016, ..., 2021)
Dependent Variable (Y)	Price Index (%)
Base Price (₹)	Actual market price in base year
Prediction Target	Future Price Index (%)
Estimated Market Price	Predicted price per kg (₹/kg)
Model Type	Linear Regression

D. Methodology of Calculation

1.ModelTraining:

The linear regression model is trained using the following equation:

$$Y=mX+c$$

where:

Y = Price Index

X = Year

m = slope of regression line

c = intercept

2.Price Prediction:

Predicted index values are converted to estimated market prices using:

$$\text{Estimated Price} = (\text{Predicted Index}/100)*\text{Base Price}$$

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

3. Performance Evaluation

1. Compare predicted values with actual historical data (if available).
2. RMSE is used to measure prediction error.
3. AIC is used for model comparison (lower AIC indicates better model).

4. Sample Results

Table II. Predicted vs Actual Price Index for Wheat

Year	Actual Index	Predicted Index	Error
2015	100	100.0	0.0
2016	102	102.5	0.5
2017	105	105.1	0.1
...	...	...	...
2025	-	118.5	-

Estimated Market Price (for 2025):- (118.5/100)*14=₹16.59/kg

Calculated RMSE: 1.02
Calculated AIC: 16.58

5. Comparison Done

Table III .Comparison with Previous Model vs Current Model

Aspect	Previous Model	Current Linear Regression Model
Error (RMSE)	Higher	Lower
AIC Score	Higher	Lower
Predictive Power	Less accurate	Improved
Trend Identification	Poor	Able to detect steady increase/decrease

6. Observation

1. Linear Regression works very well where price trends are relatively stable.
2. Provides quick actionable forecasts.
3. Lower RMSE and AIC scores indicate better fit as compared to previous basic models.

EXPERIMENTAL ANALYSIS

A. Dataset Description

1. The dataset used in this study was sourced from the Agmarknet portal, which provides historical agricultural commodity prices across different states and markets in India. For experimental consistency, the dataset was filtered for a single commodity (Tomato) and one market.
2. The dataset consisted of approximately 300–500 records spanning 1–2 years. Table V summarizes the key dataset attributes.

Table V. Dataset Attributes

Attribute	Description
Date	Daily record date

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Commodity	Name of the agricultural product (e.g., Tomato)
Market	Market location of sale
Price (₹)	Daily market price of the commodity (INR/kg)

B. Data Preprocessing

4. To ensure data quality and model readiness, the following preprocessing steps were applied:
-Missing values were removed or imputed as required.
-Date values were transformed into a numerical feature (days since first observation).
-The dataset was restricted to a single commodity (Tomato) to reduce variability.
-The independent variable (Days) and the dependent variable (Price) were finalized for model training.

C. Experimental Setup

The experiments were conducted using the following configuration:
-Programming Language Python
-Libraries Used: Pandas, NumPy, Scikit-learn, Matplotlib
-Model Used: Linear Regression (Scikit-learn)
-Evaluation Metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), R^2 Score
- Train-Test Split: 80% training and 20% testing.

D. Results and Evaluation

The trained Linear Regression model was evaluated using the test dataset. The performance metrics are presented in Table IV.

TABLE IV. MODEL EVALUATION METRICS

Metric	**Value**
Mean Squared Error (MSE)	25.87
Mean Absolute Error (MAE)	4.78
R^2 Score	0.72

- The results indicate that the proposed model provides a reliable baseline for price prediction of agricultural commodities, with acceptable error values and a reasonable R^2 score.

CONCLUSION AND FUTURE ENHANCEMENTS

In this work, a lightweight and interactive crop price forecasting system, AgriLens, was developed using a Linear Regression model. The system predicts both the future price index and the estimated market price of agricultural commodities based on historical datasets. A user-friendly interface was implemented using Streamlit, allowing stakeholders to select crops, specify a prediction year, and instantly obtain forecasts with trend insights.

Experimental results demonstrate that Linear Regression performs effectively for commodities exhibiting steady linear price growth trends. The system provides quick, interpretable, and computationally efficient forecasts, making it suitable for real-time deployment and practical decision support for farmers, traders, and policymakers.

However, the study also reveals that Linear Regression has limitations in handling highly non-linear or volatile price patterns. This motivates further exploration of advanced machine learning and deep learning techniques to improve predictive performance and adaptability. Overall, the proposed system contributes toward enhancing agricultural decision-making and sets the foundation for future research in intelligent price forecasting solutions.

FUTURE WORK

To enhance the accuracy, adaptability, and scalability of the proposed system, several future directions are identified:

Advanced Machine Learning Models

Future work will incorporate advanced models such as Random Forest Regression, Support Vector Regression (SVR), ARIMA, and LSTM networks. These approaches can better capture non-linear patterns, seasonality, and long-term dependencies in agricultural price movements.

Integration of External Factors

The predictive capability can be improved by including additional features such as climatic conditions(temperature, rainfall, humidity), government policies and subsidies, fertilizer and input costs, global trade policies, and supply chain dynamics.

Real-time Data Integration

Establishing connections with live agricultural portals, government databases, and APIs will ensure dynamic updates, enabling forecasts to remain contextually relevant and up to date.

Deployment and Scalability

Future deployment on cloud infrastructures will support large-scale accessibility. Extending the system to mobile platforms will further promote widespread adoption among farmers and stakeholders.

User Personalization

Personalized features such as saved forecasts, crop-specific alerts, and customized reports will improve usability and decision support for diverse stakeholders.

REFRENCES

[1] G. Mahajan and R. Sharma, “AI-Driven Agricultural Price Forecasting Using Machine Learning and Market Trends,” Journal of Smart Agriculture and Data Analytics, vol. 20, no. 2, pp. 456–470, (2024), doi: 10.1016/jsada.2024.107854.
[2] G. Mahajan and S. Kapoor, “Integrating Market Demand and Weather Factors for Crop Price Prediction,” International Journal of Agricultural Informatics, vol. 18, no. 1, pp. 301–315, (2024), doi: 10.1109/IJAI.2024.1091285.
[3] A. Singh, V. Mehta, and R. Kumar, “Hybrid Deep Learning Model for Agricultural Price Prediction Using LSTM and CNN,” Journal of Agricultural Data Science, vol. 22, no. 4, pp. 345–358, (2024), doi: 10.1016/j.jads.2024.109234.
[4] V. Mehta and R. Sharma, “Real-Time Agricultural Price Forecasting Using IoT and Machine Learning,” IEEE Transactions on Agriculture and Technology, vol. 61, no. 2, pp. 891–900, (2024), doi: 10.1109/TAT.2024.10871234.
[5] S. Kumar Pradeep, A. Sharma, and A. Singh, “Prediction of Agricultural Crop Prices Using Support Vector Machines,” International Journal of Agricultural Economics, vol. 15, no. 3, pp. 123–130, (2023), doi: 10.1016/ijae.2023.03.015.
[6] W. Zhang, L. Liu, and X. Zhao, “Forecasting Agricultural Prices Using Recurrent Neural Networks,” Journal of Computational Agriculture, vol. 18, no. 5, pp. 567–574, (2023), doi: 10.1109/JCA.2023.10451234.
[7] R. Singh and N. Patel, “Random Forest Models for Agricultural Price Prediction Using Climate Variables,” Journal of Agricultural Forecasting, vol. 12, no. 4, pp. 230–241, (2023), doi: 10.1016/j.jaf.2023.104723.
[8] S. Khan Ali, S. Kumar, and P. Verma, “Satellite Imagery and Machine Learning for Crop Price Prediction,” IEEE Transactions on Geoscience and Remote Sensing, vol. 60, no. 1, pp. 1234–1241, (2023), doi: 10.1109/TGRS.2023.10893456.