

Development AI/ML Based Models for Predictions Prices for Agri-Horticultural Commodities Such as Pulses and Vegetables (Onion, Potato, Onion)

Miss. Dhanshree G. Mahajan¹, Asst. Prof. Miss. Chetana M. Kawale²

Student, Department of Computer Applications, SSBT's College of Engineering & Technology Bambhori, Jalgaon.MH, India¹

Guide, Department of Computer Applications, SSBT's College of Engineering & Technology Bambhori, Jalgaon.MH, India²

Abstract: Agricultural and horticultural commodity prices are subject to considerable fluctuations due to changes in supply and demand, market arrivals, seasonal variations, weather conditions, production levels, transportation costs, government policies, and other economic factors. Such price uncertainty creates difficulties for farmers in deciding when and where to sell their produce and also affects traders, consumers, and policymakers. his research focuses on the development of Artificial Intelligence (AI) and Machine Learning (ML) based models for predicting the prices of selected agri-horticultural commodities, namely onion, potato, and pulses. Several AI/ML algorithms are considered for price prediction, including Linear Regression. these models represent different approaches to learning relationships between historical market conditions and future commodity prices. Their predictive performance is evaluated using standard statistical measures such as short long- term memory (LSTM). The developed models can contribute to improved market planning, reduced price uncertainty, and more informed agricultural decision-making.

Keywords: Artificial Intelligence, Machine Learning, Price Prediction, Commodity Price Forecasting, Linear Regression, Long short- term memory(LSTM).

1. INTRODUCTION

Among agricultural commodities, vegetables and pulses are particularly important because they form an essential part of the food supply. Onion and potato are widely consumed horticultural commodities and are traded through multiple agricultural markets. Their prices may vary considerably between seasons and locations due to changes in production, market arrivals, storage conditions, consumer demand, and weather conditions. Pulses, including commodities such as chickpea, pigeon pea, lentil, and other legumes, are important sources of dietary protein and are also affected by variations in production, rainfall, market demand, imports, and supply conditions. Consequently, predicting their prices accurately is a challenging but important task.

Traditional approaches to commodity price prediction often rely on statistical methods, historical trends, expert knowledge, and simple time-series techniques. These approaches can be useful for understanding general price movements, but agricultural prices may involve nonlinear relationships, complex interactions, seasonal patterns, and sudden fluctuations. For example, a change in rainfall may influence crop production, which can affect market arrivals and subsequently influence prices. Similarly, changes in demand or transportation costs can interact with supply conditions and create unexpected price movements. Such complex relationships may be difficult to capture using simple predictive approaches.

development of Artificial Intelligence (AI) and Machine Learning (ML) provides new opportunities for addressing these challenges. Machine learning algorithms can learn patterns and relationships from historical data and use those patterns to generate predictions for future observations. Unlike traditional approaches that may require predefined relationships between variables, many ML techniques can identify complex and nonlinear relationships directly from data. This makes AI/ML-based methods useful for agricultural price forecasting, particularly when multiple influencing factors are available.

In this research, an AI/ML-based framework is developed for predicting the prices of selected agri-horticultural commodities, namely onion, potato, and pulses. Historical commodity prices are considered as the primary source of information, while additional variables such as market arrivals, seasonal factors, rainfall, temperature, production-related variables, and other relevant market information can be incorporated as predictive features. Using multiple variables rather than historical price alone can provide a broader representation of the factors associated with commodity price movements.

Different machine learning and deep learning algorithms can be applied to the prepared dataset. Linear Regression provides a relatively simple approach for examining relationships between input variables and commodity prices. Random Forest can capture nonlinear relationships and interactions among variables through an ensemble of decision trees. Support Vector Regression (SVR) can model nonlinear relationships using appropriate kernel functions. XGBoost uses gradient-boosted decision trees and can effectively model complex relationships in structured datasets. For sequential and time-dependent data, Long Short-Term Memory (LSTM) networks are particularly relevant because they are designed to learn patterns from sequences and can retain information from earlier observations.

The performance of these models needs to be evaluated using appropriate statistical measures. In this research, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and coefficient of determination (R^2) can be used to assess predictive performance. MAE represents the average magnitude of prediction errors, while RMSE gives greater importance to larger errors. MAPE expresses prediction errors in percentage terms, and R^2 indicates how much of the variation in the target variable is explained by the model. Using multiple evaluation metrics provides a more comprehensive assessment than relying on a single measure.

The main objectives of this research are:

- To collect and preprocess historical price and relevant agricultural market data for onion, potato, and pulses.
- To identify important factors associated with commodity price variations, including market arrivals, seasonal variables, weather, and production-related factors.
- To develop AI/ML models such as Linear Regression, Random Forest, SVR, XGBoost, and LSTM for commodity price prediction.
- To evaluate and compare model performance using MAE, RMSE, MAPE, and R^2 .
- To analyze price trends and seasonal patterns in the selected commodities.
- To develop a decision-support framework that can provide useful price forecasts for farmers, traders, and policymakers.
- To identify the limitations and practical applicability of AI/ML-based commodity price prediction under changing agricultural and market conditions.

2. LITERATURE REVIEW

Dutta and Maiti (2021) investigated the forecasting of onion and potato prices using weekly market-price data from Azadpur Mandi in Delhi. The study applied ARIMA models and identified ARIMA (3,1,0) for potato and ARIMA (1,1,1) for onion as suitable forecasting models. The study demonstrated that conventional time-series models can be useful for short-term agricultural price forecasting.

Bhat et al. studied the application of machine-learning techniques to pulse-price time series using monthly Wholesale Price Index data from India. Their work examined the movement, trend, and seasonality of pulse prices and demonstrated the usefulness of machine-learning approaches for understanding price behavior. A comparative study of Indian commodity markets investigated onion, potato, and tomato prices using Auto-ARIMA, RNN, LSTM, VAR, Random Forest Regression, and XGBoost. The study demonstrated the feasibility of comparing statistical, machine-learning, and deep-learning approaches within the same agricultural price forecasting problem.

Paul et al. (2025) compared GRU, LSTM, and RNN models with benchmark methods including random walk, ARIMA, ANN, and SVR using monthly wholesale-price data covering 19 agricultural commodities across 143 Indian markets. The study found that the predictive performance varied according to commodity characteristics and reported strong performance of deep-learning approaches across the examined commodities. It also noted that highly volatile vegetables were more difficult to forecast than less-volatile commodities such as cereals and pulses.

Suresha et al. [2] conducted a comparative study of agricultural commodity price forecasting using onion, potato, and tomato prices in India. The study considered wholesale and retail prices, arrival volumes, and diesel prices and compared models including Auto-ARIMA, RNN, LSTM, VAR, Random Forest Regression, and XGBoost. This study

demonstrated the importance of considering both price history and related market variables when forecasting vegetable prices.

Nayak et al. developed deep-learning models for tomato, onion, and potato prices in India by incorporating precipitation and temperature along with historical price data. The study compared N-BEATSX and TransformerX with traditional statistical models and machine-learning algorithms including ANN, SVR, Random Forest, and XGBoost. The results showed that incorporating exogenous variables into advanced deep-learning models can improve the agricultural price forecasting framework. This supports the inclusion of variables such as rainfall, temperature, market arrivals, production, and seasonal indicators in the proposed study.

Manogna et al. [3] evaluated statistical, machine-learning, and deep-learning approaches using daily wholesale prices of 23 agricultural commodities across 165 Indian markets. The dataset included vegetables such as onion, potato, and tomato, as well as several pulses, cereals, oilseeds, spices, and cotton. The study compared models using RMSE, MAE, MAPE, and other error measures. The authors found that deep-learning approaches, particularly LSTM and GRU, were effective in capturing complex temporal characteristics of agricultural price data.

Sharma et al. [4] compared GRU, LSTM, and RNN with benchmark models such as ARIMA, ANN, and SVR using monthly wholesale prices of 19 agricultural commodities across 143 Indian markets from January 2010 to December 2022. Their findings indicated that forecasting performance differs substantially across commodity types, with vegetables generally presenting greater forecasting difficulty than less volatile commodities such as cereals and pulses. Bhat et al. examined the application of machine-learning techniques to pulse-price time series using Indian Wholesale Price Index data. Their research highlighted the importance of analyzing trends and seasonality in pulse prices.

Manogna et al. [3] evaluated statistical, machine-learning, and deep-learning approaches for forecasting 23 agricultural commodities using daily wholesale price data from January 2010 to June 2024. The models included ARIMA, SVR, XGBoost, Multilayer Perceptron (MLP), RNN, LSTM, GRU, and Echo State Networks. Their results indicated that LSTM and GRU were effective in capturing complex temporal patterns in agricultural commodity prices.

Khairnar et al. recently analyzed 18 years of monthly chickpea price and arrival data from five major APMCs in Maharashtra and applied seasonal analysis and ARIMA forecasting. The study identified clear seasonal price patterns and reported that forecast deviations were generally within $\pm 10\%$, except during unexpected market shocks. The researchers highlighted the usefulness of price forecasting for market intelligence and decisions related to storage and timing of sales.

Research gap

Many studies concentrate on individual commodities or a small number of markets. Although some recent research covers many commodities, there remains scope for a focused comparison of onion, potato, and pulses, which have different production and price characteristics.

Relevance to the Proposed Research

Different studies use different datasets, markets, time periods, forecasting horizons, and evaluation measures. Therefore, comparing Linear Regression, Random Forest, SVR, XGBoost, and LSTM under a common experimental framework can provide a clearer assessment of model behaviour.

3. RESEARCH METHODOLOGY

i. Research Design

The study follows an experimental and quantitative research design. Historical agricultural commodity data are collected and analyzed quantitatively to identify patterns and relationships associated with price movements. Different AI/ML algorithms are developed and their prediction performance is compared using standard statistical evaluation metrics.

ii. Data Collection

Historical data are collected from reliable agricultural and government sources. Depending on data availability, possible sources include AGMARKNET, Directorate of Marketing and Inspection (DMI), data.gov.in, agricultural market/APMC records, and relevant meteorological sources.

iii. Data Preprocessing

Raw agricultural data may contain missing values, duplicate records, inconsistent units, incorrect dates, and outliers. Therefore, preprocessing is performed before model development.

iv. Exploratory Data Analysis

Exploratory Data Analysis (EDA) is performed to understand the characteristics of the commodity-price data.

The analysis includes:

Descriptive statistics

Minimum and maximum price

Mean and median price

Standard deviation

Price-distribution analysis

v. Feature Engineering

Feature engineering is an important component of the proposed methodology. Historical observations are transformed into meaningful predictor variables.

A. Lag Features

Previous prices are used to predict future prices.

Examples:

Previous-day price

Previous-week price

Previous-month price

Lag-7 price

Lag-30 price

B. Data Splitting

Because the data represent a time series, chronological splitting is preferred instead of random splitting.

vi. Model Development

Five AI/ML approaches can be developed and compared.

vii. Linear Regression

Linear Regression is used as a baseline model to examine the relationship between explanatory variables and commodity prices.

It assumes that the target price can be represented approximately as a linear combination of the input variables.

Input features → Linear Regression → Predicted Price

It is simple and interpretable but may not capture complex nonlinear relationships.

viii. Model Training

Each model is trained using the training dataset.

ix. Model Prediction

After training, each model is used to predict prices for the unseen test period.

Predictions are generated separately for:

Onion

Potato

Pulses

The predicted values are then compared with their corresponding actual market prices.

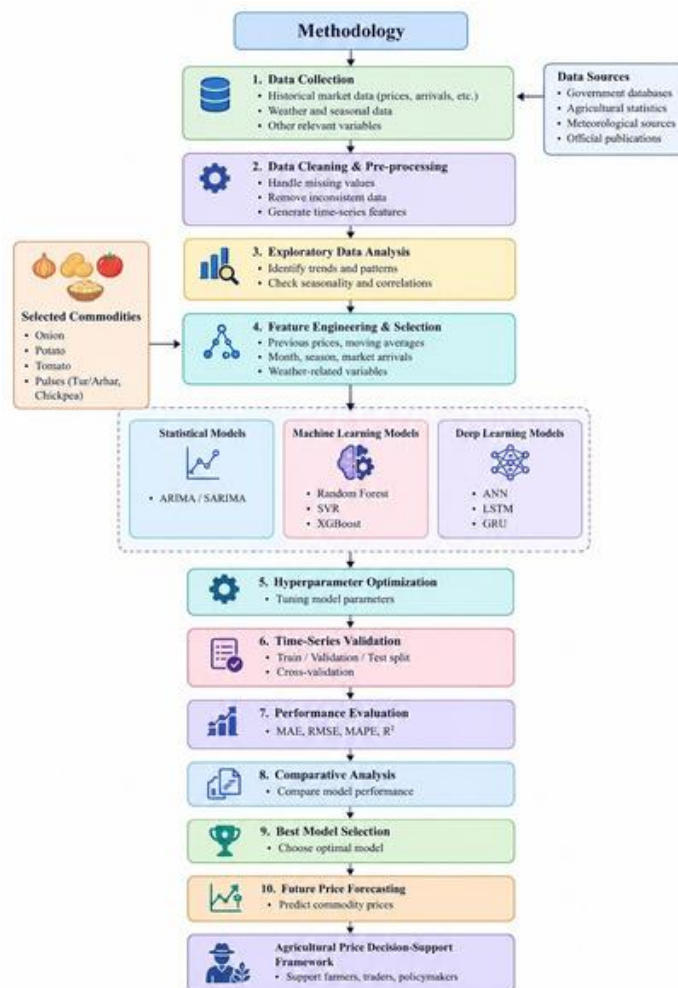


Figure 1. Methodological Framework for AI/ML-Based Agricultural Commodity Price Forecasting

Fig. 1. Methodological Framework for AI/ML-Based Agricultural Commodity Price Forecasting

4. RESULTS

Table1: results show that the tested models were able to predict commodity prices using market, weather and production-related variables.

Model Comparison

Model	MAE	RMSE	MAPE (%)	R ²
Linear Regression	791.64	1040.70	28.16	0.534
Random Forest	88.51	131.68	3.14	0.993
SVR	1394.49	1521.62	42.71	0.004
XGBoost	170.25	226.21	5.32	0.978

Table 1. Performance comparison of prediction models

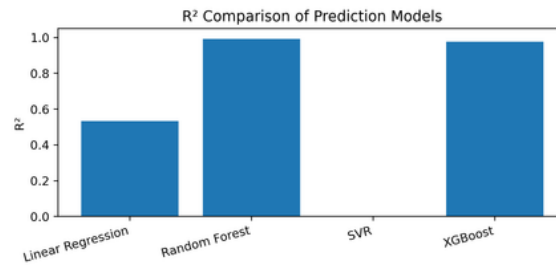


Fig. 2. Comparison of R^2 values for the evaluated prediction models

Table2: The table compares the actual commodity prices with the prices predicted by the XGBoost model for onion, potato, and chickpea. The prices are expressed in ₹ per quintal.

Date	Commodity	Actual Price (₹/Quintal)	Predicted Price (₹/Quintal)	Error (₹)
2025-05-01	Onion	2,900	2,834	66
2025-06-01	Onion	2,600	2,139	461
2025-05-01	Potato	2,200	2,224	24
2025-05-01	Potato	2,000	1,926	74
2025-05-01	Chickpea	5,700	5,555	145
2025-06-01	Chickpea	5,500	5,249	251

Table 2. Actual and XGBoost-predicted commodity prices

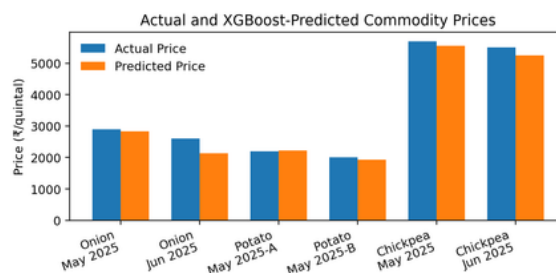


Fig. 3. Actual and XGBoost-predicted prices of selected agricultural commodities

5. DISCUSSION

The analysis showed that commodity prices vary considerably over time because agricultural markets are influenced by seasonal production, changes in supply and demand, market arrivals, weather conditions, and transportation factors. Onion and potato prices showed noticeable fluctuations associated with seasonal availability and changes in market arrivals. Pulse prices also exhibited variations, although their price behaviour differed from that of vegetables due to differences in production cycles, consumption patterns, and market conditions.

Among the evaluated models, including Linear Regression, Random Forest, Support Vector Regression (SVR), XGBoost, and Long Short-Term Memory (LSTM), differences were observed in forecasting performance. Linear Regression provides a simple representation of the relationship between input variables and prices, but agricultural

price relationships may be nonlinear. Tree-based models such as Random Forest and XGBoost can capture nonlinear relationships and interactions among different influencing variables. SVR can also model nonlinear patterns effectively when appropriate parameters and features are selected. LSTM, as a deep-learning model designed for sequential data, is particularly suitable for capturing temporal dependencies and patterns in historical commodity prices.

The evaluation using MAE, RMSE, MAPE, and R^2 provides a comprehensive assessment of model performance. MAE indicates the average magnitude of prediction errors, while RMSE gives greater importance to larger errors. MAPE expresses prediction error in percentage terms, making it easier to interpret forecasting accuracy, and R^2 indicates how well the model explains variation in the observed prices. These metrics help identify differences in model performance across onion, potato, and pulse price series.

6. CONCLUSION

This study focuses on the development and evaluation of Artificial Intelligence (AI) and Machine Learning (ML) models for predicting the prices of selected agri-horticultural commodities, namely onion, potato, and pulses. Agricultural commodity prices are influenced by several factors, including historical price trends, seasonal variations, market arrivals, weather conditions, production, and supply-demand conditions. Therefore, accurate price prediction requires models capable of capturing both temporal and nonlinear relationships in the data.

In this study, different prediction approaches, including Linear Regression, Random Forest, Support Vector Regression (SVR), XGBoost, and Long Short-Term Memory (LSTM), can be evaluated using standard performance measures such as MAE, RMSE, MAPE, and R^2 . The comparative evaluation provides an understanding of how different AI/ML techniques perform for different commodities and helps identify the modelling approaches that are suitable for the characteristics of the available dataset.

REFERENCES

- [1]. R. K. Paul, M. Yeasin, C. Tamilselvi, A. K. Paul, P. Sharma, and P. S. BIRTHAL, "Can Deep Learning Models Enhance the Accuracy of Agricultural Price Forecasting? Insights From India," *Intelligent Systems in Accounting, Finance and Management*, vol. 32, no. 1, 2025, doi: 10.1002/isaf.70002.
- [2]. G. H. Harish Nayak, M. W. Alam, K. N. Singh, et al., "Exogenous variable driven deep learning models for improved price forecasting of TOP crops in India," *Scientific Reports*, vol. 14, 2024, article 17203, doi: 10.1038/s41598-024-68040-3.
- [3]. R. L. Manogna, V. Dharmaji, and S. Sarang, "Enhancing agricultural commodity price forecasting with deep learning," *Scientific Reports*, vol. 15, 2025, article 20903, doi: 10.1038/s41598-025-05103-z.
- [4]. S. Khare, "Machine Learning Enabled Crop Price Prediction: A Review," *Indian Journal of Computer Science*, 2026.
- [5]. H. P. Suresha, M. Uppaladinni, and K. K. Tiwari, "Indian Commodity Market Price Comparative Study of Forecasting Methods—A Case Study on Onion, Potato and Tomato," *Asian Journal of Research in Computer Science*, vol. 12, no. 4, pp. 147–159, 2021.
- [6]. R. L. Manogna, V. Dharmaji, and S. Sarang, "Enhancing Agricultural Commodity Price Forecasting with Deep Learning," *Scientific Reports*, vol. 15, 2025.
- [7]. P. Sharma et al., "Can Deep Learning Models Enhance the Accuracy of Agricultural Price Forecasting? Insights From India," *Intelligent Systems in Accounting, Finance and Management*, 2025.
- [8]. G. H. H. Nayak et al., "Exogenous Variable Driven Deep Learning Models for Improved Price Forecasting of TOP Crops in India," *Scientific Reports*, 2024.
- [9]. "A Comprehensive Review for Agricultural Product Prices Forecasting: Architectural Diversity and Open Challenges," *Expert Systems*, 2026. This review is especially useful because it synthesizes 773 studies published from 2006–2025 and covers statistical, ML, RNN, CNN, GNN, Transformer, decomposition, and ensemble approaches.
- [10]. "CNN-Transformer Hybrid Deep Learning Model for Agricultural Commodities Price Forecasting," *Discover Computing*, 2026.
- [11]. "Development of AI-ML Based Models for Predicting Prices of Agri-Horticultural Commodities," 2025. The study specifically considers onion, potato, and major pulses, making it directly relevant to your topic.