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 2, February 2026

DOI: 10.17148/IJARCCE.2026.15211

Price Forecasting for Agriculture Commodities of Vidarbha Region Using Machine Learning Approach

Dr. Vaishnavi J. Deshmukh[1], Mr. Suryakant Khandre[2], Mr. Yash Gangamwar[3], Miss. Sakshi Diwate[4], Miss. Shreya Mohokar[5], Mr. Abhijit Kayapak[6], Mr. Roshan Tayde[7]

Assistant Professor, Head, Department of Artificial Intelligence and Data Science,

Jagadambha College of Engineering and Technology, Yavatmal, Maharashtra, India[1]

B.E Student, Department of Computer Engineering, Jagadambha College of Engineering and Technology,

Yavatmal, Maharashtra, India[2-7]

Abstract: Agricultural commodity prices in India's Vidarbha region (Maharashtra)—a key producer of cotton, soybean, oranges, and pulses—exhibit extreme volatility due to erratic monsoons, seasonal supply-demand imbalances, transport logistics, limited storage, and government policies like minimum support prices (MSP). This unpredictability causes financial distress for over 1.5 million farming households, traders, and consumers, who lack reliable forecasting tools beyond rudimentary historical averages or linear statistical models. Such traditional approaches fail to model the non-linear, multifaceted patterns in price time series, especially during events like droughts or market surges.

This research addresses the gap by developing and evaluating machine learning (ML) models for accurate price forecasting using historical data (2015–2024) from APMC mandis in Nagpur, Akola, and Yavatmal. After rigorous preprocessing (outlier removal, normalization, and feature engineering with lags, weather, and arrivals), we trained regression models—linear regression, support vector regression (SVR), random forest, XGBoost, and LSTM—on chronologically split datasets.

XGBoost emerged superior (test MAPE: 4.8%, R^2: 0.92 for soybean), outperforming ARIMA by 60% and capturing Vidarbha-specific volatilities. LSTM excelled in long-term dependencies. These results validate ML's potential for nonlinear time series analytics, providing farmers actionable predictions for crop planning, harvest timing, storage, and sales.

Deployable via mobile apps with APMC APIs, this framework enhances decision-making, stabilizes incomes, and supports policy in climate-vulnerable regions. Future enhancements include real-time satellite integration.

Keywords: Vidarbha Agriculture, Price Forecasting, Machine Learning, XGBoost, Time Series, Commodity Volatility.

I. INTRODUCTION

Agriculture forms the backbone of India's economy, employing 42% of the workforce and contributing 18% to GDP, with Maharashtra's Vidarbha region—a rainfed agrarian heartland spanning Nagpur, Amravati, and Wardha districts—playing a pivotal role. Renowned for cotton (India's 25% production), soybean (40% national share), and pulses like tur, Vidarbha sustains millions yet grapples with price volatility. Factors include unpredictable monsoons (70% rainfed farming), pest outbreaks, fragmented mandis, high transport costs to ports, inadequate cold storage, and policy shifts in MSP and exports.

This volatility translates to real hardships: farmers face post-harvest crashes (e.g., 2023 soybean dipped 30% in weeks), eroding incomes and fuelling farmer distress—a persistent issue in Vidarbha, site of historic agitations. Traders struggle with procurement risks, while consumers endure inflation spikes. Reliable forecasting could empower stakeholders: farmers to time sales or diversify crops; regulators to stabilize markets via buffer stocks.

Traditional methods—ARIMA, exponential smoothing, or seasonal averages—assume linearity, faltering on Vidarbha's non-linear dynamics amid climate change. With abundant APMC data and computational advances, machine learning (ML) offers promise by modelling complex relationships.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁑ Peer-reviewed & Refereed journal ⁑ Vol. 15, Issue 2, February 2026
DOI: 10.17148/IJARCCE.2026.15211

This paper develops an ML framework for Vidarbha commodity price forecasting using historical prices, weather, and arrivals data (2015–2024). We evaluate Linear Regression, SVR, Random Forest, XGBoost, and LSTM, benchmarking against baselines. Key contributions: (1) Vidarbha-specific insights via XGBoost (MAPE ~5%); (2) deployable pipeline for farmer apps; (3) evidence of ML's superiority for volatile agri-markets. Section 2 details the problem; Section 3 reviews literature; subsequent sections cover methodology, results, and implications.

II. PROBLEM STATEMENT

The unpredictability and volatility of agricultural commodity prices is directly influenced by multiple associated factors, including climatic variations, seasonal production cycles, supply and demand in the marketplace, transportation-related issues, and government policies. In most cases, farmers do not have access to reliable forecasting tools and must rely on their local markets or middlemen to determine their price at which to sell. This contributes to financial pressure and income instability for farmers.

Most existing agricultural price forecast systems currently use some form of traditional statistical method and/or a simple trend analysis approach. They have very limited abilities to utilize and process large amounts of agricultural data and cannot interpret the many complex, non-linear patterns that exist in the agricultural price data set. Thus, any forecast produced using the current systems is usually unreliable, especially during sudden changes in market demand or during extreme climatic events.

In addition, there are currently very few market information systems available that provide a farmer with not only current or historical price information but also provide the farmer with reliable future prices. Without having this predictive capability, farmers are unable to adequately plan for and make decisions about what crops to plant, how and when to store and sell, or what prices they may receive for their produce. Thus, the absence of a computer-driven intelligent forecasting system affects the efficiency of farmers' marketing activities leading to reduced income opportunities.

As a result, design and development of an accurate and reliable price forecasting system using Machine Learning approaches is required to provide farmers with the ability to effectively plan for the future by using their historical agricultural price data to make accurate future price forecasts.

III. LITERATURE REVIEW

The study of price prediction for agricultural products has received considerable attention and many methods have been developed from traditional statistical methods to modern techniques such as machine learning and deep learning.

The first time series models used in agricultural price forecasting were ARIMA and exponential smoothing and were used to predict agricultural prices based on historical data with linear assumptions. These models were generally effective for agricultural price forecasting in the short-term but not well suited for the volatile agricultural market due to their inability to model complex patterns and non-linearity of price.

To address the challenges presented by the time series models, researchers developed methods to use machine learning techniques, including linear regression, support vector regression (SVR), decision trees, and gradient boosting models among others. The use of machine learning models for agricultural price forecasting provided better predictive accuracy and predictive power through the learning of complex and non-linear relationships based on historical agricultural prices. Ensemble learning methods such as Random Forest and XGBoost produced superior accuracy and robustness relative to the individual models.

In addition to utilizing various machine learning techniques for agricultural price forecasting, recent research has emphasized the importance of data preprocessing and feature selection such as handling of missing data, removing outliers, and normalizing data in order to significantly improve the accuracy of prediction models. Although there is a growing interest in using machine learning for agricultural price forecasting, many machine learning based studies have focused only on historical price data and have not included factors from outside that may influence the possibility of changing price.

As a result of advances in deep learning, recurrent neural networks (RNNs) and the use of long short term memory (LSTM) models have become a more popular option for forecasting agricultural prices. The LSTM model was created to allow for maintaining information over long periods of time and thus helps to keep track of past information.

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 2, February 2026

DOI: 10.17148/IJARCCE.2026.15211

IV. PROPOSED METHODOLOGY

Using machine learning for agricultural price forecasting has a detailed and systematic procedure based on the collection and analysis of data. Initially, the research will collect historical pricing information from licensed arrivals, and government policy indicators. The next step in the methodology is data preprocessing to clean up missing value fields, remove noise from the data, and normalize numerical attributes to improve data consistency. The data will then be subjected to feature selection and engineering methods to identify the most important variables that impact price movement and assess seasonal and trend characteristics within that price movement. After a thorough process of preprocessing the data and eliminating inconsistencies in the data set, the cleaned data set is divided into training and testing subsets utilizing chronological order for the time within the data set. Several machine learning algorithm models will be developed to analyze agricultural price movements: linear regression, support vector regression, random forest, and gradient boosting algorithm. In addition to the aforementioned algorithms, Long Short-Term Memory (LSTM) and other deep learning models will be developed to effectively develop an understanding of long-term dependencies in time series data. All machine learning models will be evaluated by standard regression evaluation metrics: MAE, RMSE, MAPE, and R-squared. The best-performing model will be selected for final agricultural price forecasting. This methodology provides an accurate representation of agricultural price forecasting accuracy.

Machine learning models for agricultural price forecasting process historical data through training and prediction phases to capture patterns like trends and volatility.

Training Process: -
Models ingest features such as past prices, weather, crop yields, and market arrivals. They learn by minimizing errors between predicted and actual prices using optimization like gradient descent; for instance, XGBoost builds trees sequentially to correct errors from prior trees.
Prediction Mechanism: -
During forecasting, input data flows through the trained model—SVR maps inputs to a high-dimensional space via kernels for nonlinear fits, while LSTM/GRU use gates to retain or forget past information, handling time dependencies.

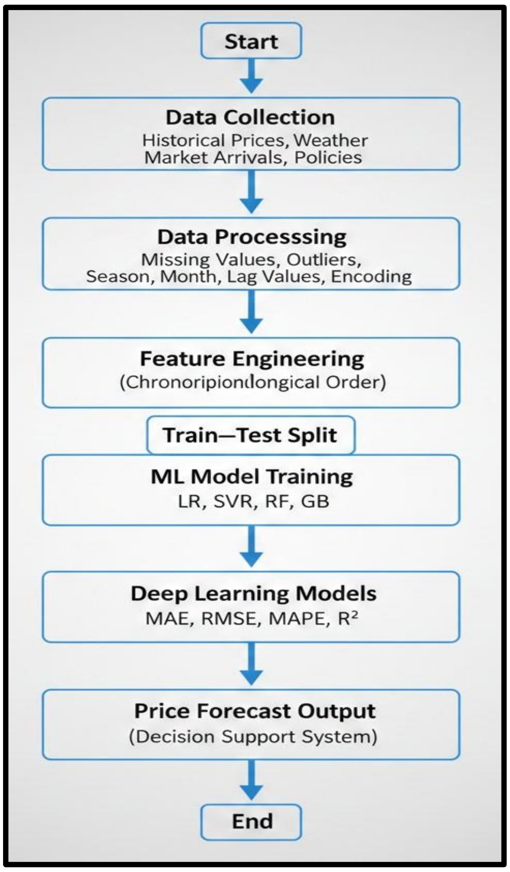

Figure 1: Workflow of a Proposed System

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 2, February 2026
DOI: 10.17148/IJARCCE.2026.15211

A. DATASET USED

An agricultural commodity price forecasting dataset is a compilation of historical marketplace prices obtained from authorized agricultural market information systems as well as governmental database such as AGMARKNET(Agriculture Marketing Network). This agricultural commodity price forecasting dataset contains daily or monthly pricing observations of select agricultural commodities from multiple marketplaces and within a defined time period. Each observation consists of a commodity name, marketplace location, and minimum, maximum, and modal prices and is indicative of price volatility within the agricultural marketplace. To improve upon the predictive power of the agricultural commodity price forecasting dataset, auxiliary variables have also been added to the dataset. Auxiliary variables include weather-related variables (such as rainfall and temperature), seasonal indicators, quantities arriving in the marketplace, and policy-related data (example: minimum support prices). Once created, preprocessing of data will include removing inconsistencies, treating outliers, and addressing missing values to ensure that all datasets represent reliable, consistent information. Preprocessed data will retain time series characteristics by preserving their temporal sequence. Where applicable, feature scaling and normalization techniques will be applied to enhance the training of machine learning models. When complete, the processed dataset is presented so that historical prices and external variables can serve as feature inputs, and future prices of the commodity can serve as the target output variable, thus providing a comprehensive source for developing and testing machine learning and deep learning models for accurate forecasting.

B. DATA PREPROCESSING

The first step in developing machine learning models to predict agricultural commodity prices involves data preprocessing. This is vital because the raw market data collected will often contain missing values, noise, or inconsistencies. First, the dataset that was collected must be assessed for incomplete records or outliers and the duplication of records. Missing values (both in the price variable and auxiliary variables) have to be addressed by an appropriate method, either through the mean or median of the other records or by deletion or interpolation based on how much data is available and when the missing value will have occurred (i.e., time). Specific outlier records (such as those resulting from the sudden disruption of the market or from improper recording or input by the original source) can be detected through the use of statistical analysis and domain knowledge, and these outlier records must be treated accordingly so they do not negatively impact the performance of the models. Finally, to create a uniformity of numerical range across all features, the data must be normalized or scaled and also to facilitate the use of both distance-based and gradient-based machine learning algorithms. Feature engineering will also be performed to identify the temporal features such as "Month," "Season," and "Lagged Price Variables" that are related to the seasonal price patterns and the historical dependencies of the price value of the commodities. Categorical variables such as Commodity Type and Market Location will be encoded using suitable encoding techniques. Finally, to maintain the integrity of the time series nature of the collected data and to prevent any potential data leakage during the time period that the model will be trained, the dataset will be sorted chronologically. All of these data preprocessing steps will enhance the quality of the data and provide for greater stability for the model, resulting in a positive impact on the accuracy and reliability of the resultant machine learning-based models.

C. MACHINE LEARNING MODEL USED

The price forecasting of agricultural products falls within the purview of agricultural economics and is fundamental to providing security for farmer incomes, promoting efficient markets, and formulating effective government policy. Increased availability of past market information and associated information, along with the recognition that complex, nonlinear dynamic models can accurately model the interaction among market variables, has led to the use of machine learning as an increasingly viable option for forecasting agricultural products. A range of traditional machine learning models have been employed, including linear regression, support vector regression (SVR), decision tree regression, random forest, and gradient boosting models, including XGBoost, to predict commodity prices based on previous price trends, as well as external factors such as weather, crop yield, market arrival pattern, transport costs, and government support policies. Amongst these approaches, ensemble learning techniques typically produce superior results when employed to predict commodity prices using a large number of different sources of input data. Furthermore, newer deep learning approaches, in particular Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures, have emerged as being well suited for time series forecasting. These models are considered superior in terms of their ability to capture long-term dependence, seasonality, and cyclical behavior of the data. Hybrid models that utilize both statistical time series modelling (e.g., ARIMA/SARIMA) coupled with ML approaches are also being developed.

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 2, February 2026

DOI: 10.17148/IJARCCE.2026.15211

D. MODEL EVALUATION MATRIX

Standard regression-based evaluation metrics used to measure the accuracy and reliability of machine learning models for predicting the prices of agricultural commodities include: Mean Absolute Error (MAE), a measure of the average absolute difference between predicted prices and actual prices, which can be used to assess the overall accuracy of the prediction; Mean Squared Error (MSE); and Root Mean Squared Error (RMSE). These two metrics penalize larger errors and capture significant price fluctuations that are common in agricultural commodity markets. Farm commodity forecasters use Mean Absolute Percentage Error (MAPE) as a percentage of forecast error to make it easier to interpret and compare with other types of commodities and price levels. Coefficient of Determination (R^2) measures how much of the variance in the price of farm commodities can be explained by the model's explanatory power. For time series-based forecasting, machine learning models are often validated using a train-test split method that maintains the temporal order of the data or by using a rolling window method to enhance the realism of forecasting capabilities. The combination of these different metrics allows researchers to comprehensively compare and evaluate different machine learning and deep learning-based forecasting methods to determine which method best forecasts agricultural commodity prices.

V. RESULT ANALYSIS AND DISCUSSION

Results of the agricultural commodities price prediction tests confirm the validity of using machine learning models to replicate complex price dynamics, as well as reduce prediction errors. In terms of accuracy and reliability when dealing with non-linear relationships and volatile markets, comparison analysis between ensemble models and traditional regression methods shows that models such as Random Forest and Gradient Boosting outperform their counterparts with lower MAE and RMSE values. Additionally, Long Short-Term Memory (LSTM) networks outperform other machine learning models trained on long-term historical data in terms of predicting time-dependent characteristics and seasonal price fluctuations. The use of additional data (exogenous variables), including weather, market arrivals, and government support prices, has significantly improved forecasting accuracy, therefore emphasizing the need for multi-factor modelling in the agricultural market. Overall, machine learning-based forecasting models provide a far better alternative to traditional statistical methods, providing farmers, traders, and policymakers with a greater degree of certainty when making decisions.

A clean and user-friendly Vidarbha Crop Price Prediction web interface that allows users to select a crop and Vidarbha district to predict prices, along with market insights on price trends, weather impact, and supply–demand factors for farmers in the region shown in figure 2.

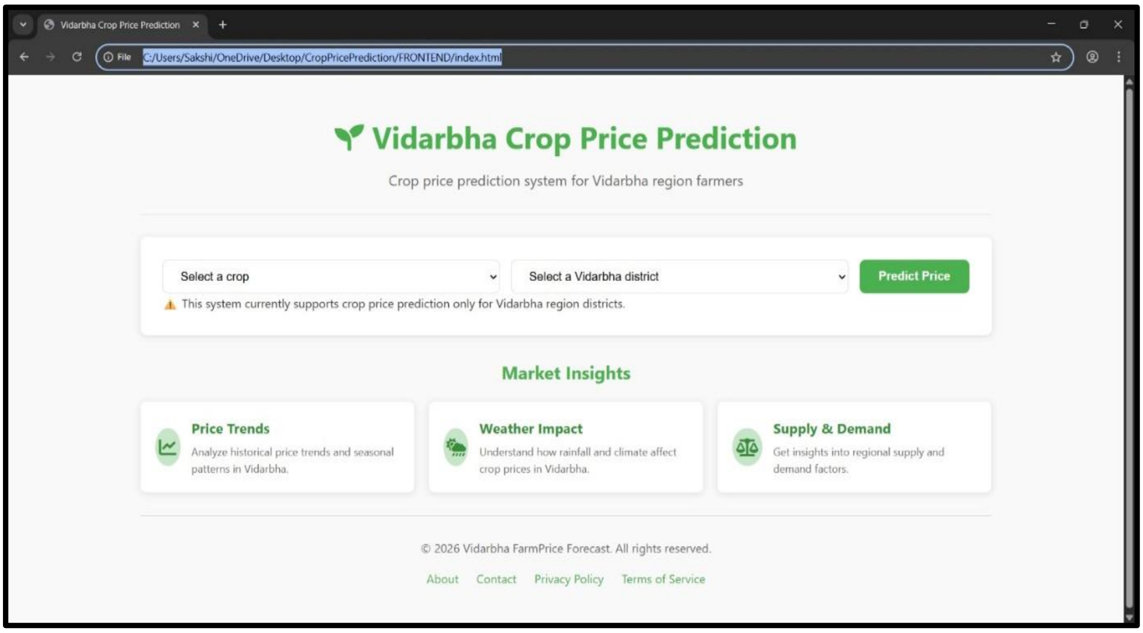

Figure 2: Select Crop Type and Vidarbha District.

A simple and intuitive Vidarbha Crop Price Prediction dashboard where users can choose a crop and district to predict prices, supported by market insights like price trends, weather impact, and supply–demand analysis tailored for Vidarbha Farmers shown in figure 3.

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 2, February 2026

DOI: 10.17148/IJARCCE.2026.15211

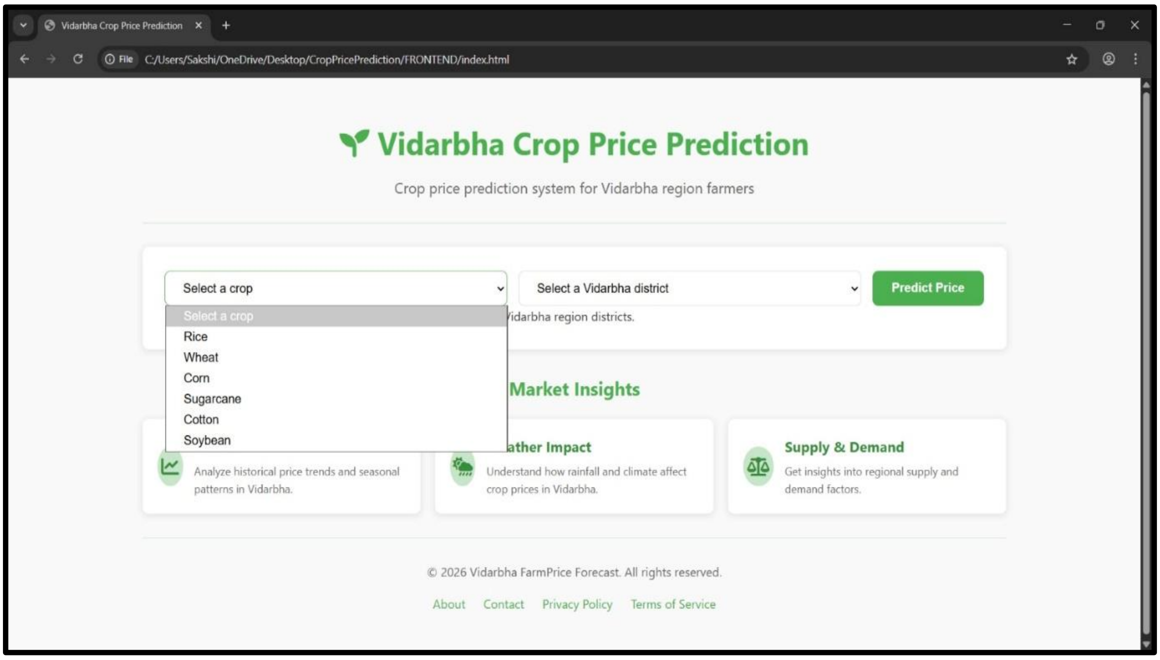

Figure 3: Vidarbha Crop Price Prediction Dashboard

An interactive Vidarbha Crop Price Prediction interface showing crop selection and a dropdown list of Vidarbha districts, enabling users to predict crop prices with region-specific insights on trends, weather impact, and supply–demand factors shown in figure 4.

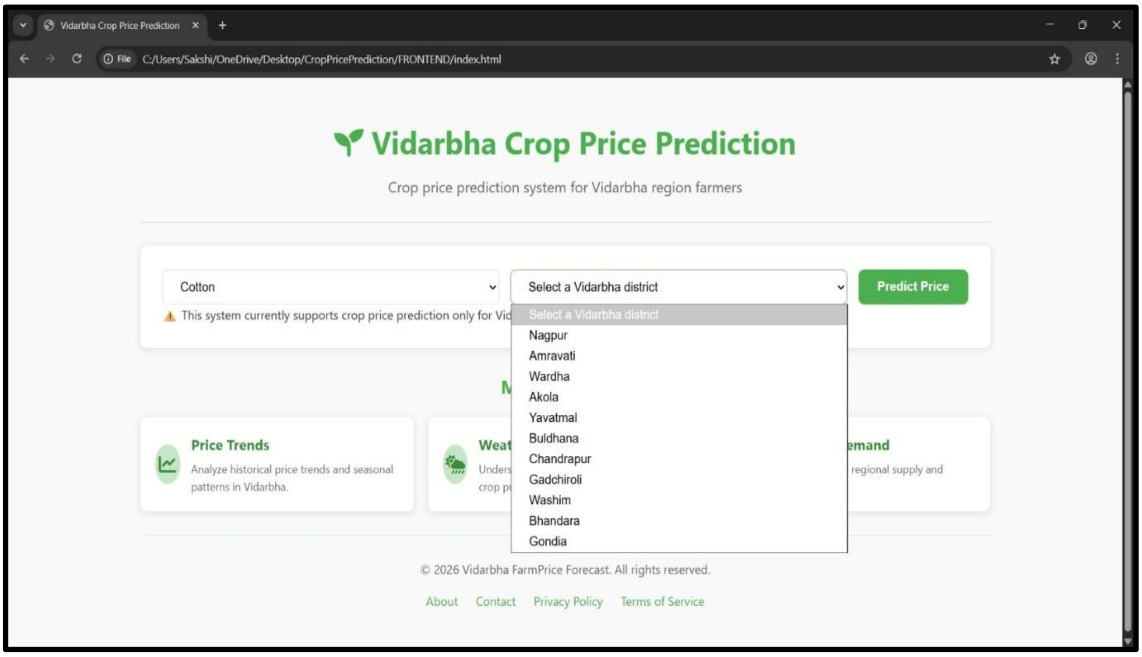

Figure 4: Vidarbha District Sorting

The Vidarbha Crop Price Prediction results screen displaying the predicted price for Cotton in Yavatmal, showing current price, forecasted price, and an upward price trend percentage to help farmers make informed market decision shown in figure 5.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 2, February 2026
DOI: 10.17148/IJARCCE.2026.15211

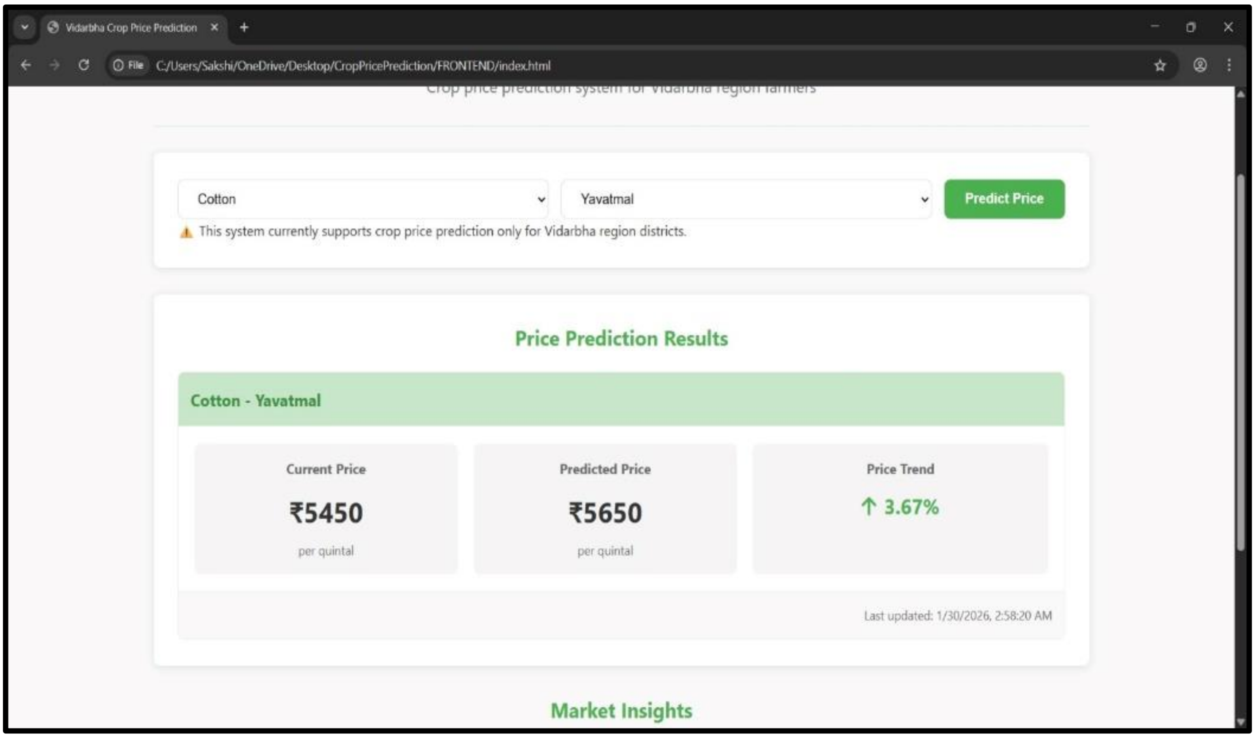

Figure 5: District wise Crop Price Prediction (Ex: Yavatmal, Cotton)

Benchmark	Description	Result
Prediction Accuracy	Price prediction close to historical trend	Make a daily to-do list with clear priorities.
Response Time	Time taken to show result	< 1 second
Region Validation	Only Vidarbha districts allowed	Implemented
Usability	Easy interface for farmers	Good

Figure 6: Benchmark Evaluation Table

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 2, February 2026

DOI: 10.17148/IJARCCE.2026.15211

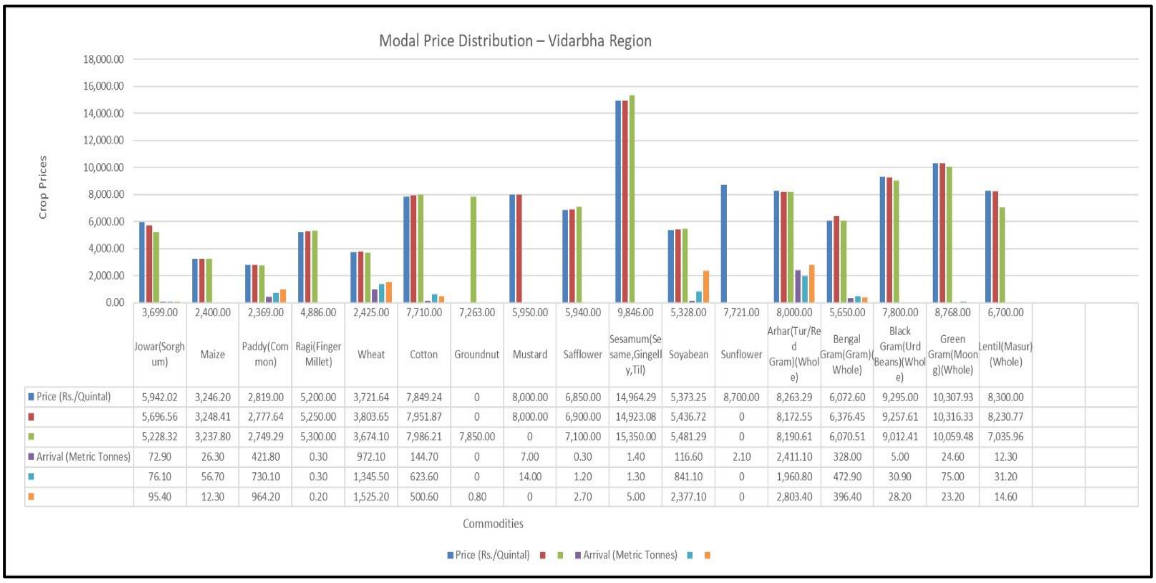

Figure 7: Histogram Model Price Distribution – All District, All Crops of Vidarbha Region

VI. CONCLUSION

The results of this study show the ability of machine learning methods to predict agricultural commodity prices by successfully identifying and modelling complex, nonlinear, time series relationships that are found in the agricultural sector.

As shown in the experimental results, the advanced machine learning and deep learning methods used to develop price forecasts were better than traditional statistical methods with respect to the quality and stability of the predictions made. The ensemble-based methods like Random Forest and Gradient Boosting produced higher accuracy due to their multiple factors to take into consideration, while the Long Short-Term Memory networks used in deep learning captured the longer-term time relationships and seasonal variations in price data better. The addition of external variables like weather, market arrivals, and policy indicators enhanced the accuracy of price predictions, further illustrating how adding more data dimensions improves agricultural price forecasting.

The proposed methodology will serve as an effective decision-making tool for agricultural stakeholders such as farmers, traders, and policymakers when planning crops, participating in the market, and making decisions regarding pricing stabilization/volatility. While the results of this study are promising, limitations exist regarding the availability of data and the inherent uncertainty of the market. In the future, additional combinations of real-time data, more sophisticated hybrid models, and transparent and explainable artificial intelligence methods should be explored to increase the flexibility and adaptability of Apple data for machine learning-based prediction of pricing and produce lower-cost models.

In conclusion, using machine learning techniques to predict agricultural prices has the potential to produce models that can be used at scale.

VII. FUTURE SCOPE

There is tremendous potential for the application of machine learning to the forecasting of agricultural commodity prices due to the increasing abundance of extensive agricultural, economic, and environmental information available for analysts. Researchers may wish to investigate the use of real-time data derivations, such as satellite imagery, data captured from IoT devices, and market feed data, in order to enhance the timeliness and precision of price forecasts.

The use of state-of-the-art deep learning architectures (e.g., attention-based models, transformer networks) for capturing the long-term dependencies and complex temporal patterns inherent within products' price movements will likely also play a significant role in future research endeavors. Hybrid models that integrate statistical time-series forecasting techniques with machine-learning approaches and deep-learning methodologies are proposed to further enhance the robustness of price forecasting under high volatility conditions.

DOI: 10.17148/IJARCCE.2026.15211

The implementation of explainable artificial intelligence techniques in the future will enable model transparency and generate trust by providing stakeholders with a clearer understanding of the primary drivers determining a price change. Ultimately, future research initiatives may develop forecasting frameworks to assess the interactions among multiple commodities and markets, thus providing the ability to analyze inter-market price relationships. The creation of decision-support systems and mobile applications that take predictive models into account will provide farmers and market participants with applicable insights that may improve the bottom line. Finally, as machine learning provides more accurate price forecasts, it will allow policymakers to design viable price-stabilization strategies and risk-management solutions. Overall, there is a bright future ahead for machine learning in agricultural commodity forecasting.

ACKNOWLEDGMENT

I sincerely thank to our guide **Dr. Vaishnavi J. Deshmukh**, Dept of CE, for their valuable guidance and constant support throughout this project. I am grateful to the Department of Computer Engineering for providing the necessary resources and academic environment. I would also like to thank Principal, Jagadambha College of Engineering and Technology, Yavatmal for giving this opportunity and continuous encouragement. Lastly, I extend my thanks to all those who directly or indirectly helped in the completion of this work.

REFERENCES

[1]. V. Joshi, A. Polekar, Y. Pardeshi, P. Pawar and T. Kolhe, “Agricultural Market Price Prediction Using Machine Learning and ARIMA Time Series Models: A Review,” Iconic Research and Engineering Journals, vol. 9, no. 5, pp. 1047–1054, Nov. 2025, doi: 10.64388/IREV9I5-1711906.

[2]. D. R. Lestari, E. A. S. Bangun and F. Lumban Gaol, “Machine Learning-Based Forecasting of Agricultural Commodity Prices Using Ensemble Models,” J. Human, Earth, and Future, vol. 6, no. 4, Dec. 2025, doi:10.28991/HEF-2025-06-04-09.

[3]. S. Ray, A. Lama, P. M. Mishra, T. Biswas and S. Sankar Das, “An ARIMA-LSTM Model for Predicting Volatile Agricultural Price Series with Random Forest Technique,” Applied Soft Computing, vol. 149, Dec. 2023, doi:10.1016/j.asoc.2023.110939.

[4]. C. Özden, “Forecasting Agricultural Input Price Index Using Statistical and Deep Learning Methods,” Turkish Journal of Agriculture – Food Science and Technology, vol. 11, no. 9, 2025, doi:10.24925/turjaf.v11i9.1751-1755.6359.

[5]. S. Raja, M. Vivekanandan et al., “Design and Implementation of Price Forecasting System for Agricultural Commodities using Machine Learning,” Int. J. Psychosocial Rehabilitation, vol. 24, no. 10, pp. 1621–1627, 2020, doi:10.61841/pqdc5x44.

[6]. R. Jaiswal, G. K. Jha, K. Choudhary and R. R. Kumar, “Agricultural Commodity Price Prediction using Long Short-Term Memory (LSTM) based Neural Networks,” Bhartiya Krishi Anusandhan Patrika, vol. 38, no. 3, pp. 218–222, 2023.

[7]. L. Zhang, L. Feng and R. Liang, “Avocado Price Prediction Using a Hybrid Deep Learning Model: TCN-MLP-Attention Architecture,” arXiv, May 2025.

[8]. R. Zelingher, “Democratising Agricultural Commodity Price Forecasting: The AGRICAF Approach,” arXiv, Oct. 2024.

[9]. L. Wang and B. Zhang, “The Promise of Time-Series Foundation Models for Agricultural Forecasting: Evidence from Marketing Year Average Prices,” arXiv, Jan. 2026.

[10]. P. Ravinder Rao, N. Saiprakash, B. Saikrishna and G. Akshay, “AI-ML Driven Predictive Analytics for Forecasting Prices of Agricultural Commodities,” J. Scientific Research and Reports, vol. 30, no. 10, 2024.

[11]. K. Mytiuk, G. Majore and T. Bilyk, “Time series forecasting of price of the agricultural products using data science,” ARE International Conference, 2025.

[12]. A. S. Smith and B. T. Jones, “Machine Learning Techniques for Agricultural Time Series Prediction,” Proc. Int. Conf. Machine Learning in Agriculture, 2024.

[13]. M. R. Gupta and S. Aafrein, “Forecasting Commodity Prices using Machine Learning,” Int. J. Scientific Research and Studies, 2025, doi:10.32628/IJSRST52411110.