Intelligent Price Forecasting System for Spice Traders with Machine Learning

Kiran Basavannappagowda[*1], J B Simha[2], Praveen M P[3], Gundlupet Sadananda Murthy[4]

[1,2] REVA Academy for Corporate Excellence -RACE, REVA University Rukmini Knowledge Park, Kattigenahalli,Yelahanka, Bengaluru, India- 560 064, India
[3] Director – Operations,Numentrix Consulting LLP , #332, VI Main,I Stage,KHB Colony,Basaveshwaranagar, Bangalore, India – 560079, India
[4]Director- samparkbindhu, J P Nagar 9th phase Bengaluru, India- 560 062, India.

[*]Kiran.AI07@race.reva.edu.in
[2]jb.simha@reva.edu.in
[3]praveen@numentrix.com
[4]g.sadananda.murthy@samparkbindhu.in

Abstract. Spices are essential agricultural products that hold considerable economic importance and fulfill various roles in culinary, medicinal, and industrial fields. Black pepper, a spice traded worldwide, experiences frequent price changes due to seasonal variations, inconsistent quality, and disruptions in the supply chain. This research introduces a forecasting model aimed at predicting black pepper prices in local markets of Karnataka. Conventional techniques such as Simple Moving Average (SMA), Auto-Regressive Integrated Moving Average (ARIMA), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) often yield subpar results when faced with irregular data conditions. To overcome this challenge, a Multiple Linear Regression (MLR) model with lag features was developed, utilizing domain-specific feature engineering. The data processing pipeline included steps for managing missing values, outliers, normalization, and identifying temporal patterns. A knowledge-driven nearest neighbor analysis was employed to improve forecasting accuracy. Among all the models assessed, the MLR model recorded the lowest Mean Absolute Percentage Error (MAPE) of 0.22%. The proposed work also features a user interface designed to aid traders in making informed decisions and allows for a more in-depth analysis of black pepper trading trends.

Keywords: Black Pepper, Commodity Price Forecasting, Deep Learning, Time Series Analysis, MLR, Lag feature, SMA, ARIMA, LSTM, GRU, Machine Learning, Price Volatility, Market Prediction.

© The Author(s) 2025
A. Jain and S. Gupta (eds.), *Proceedings of the International Conference on Policies, Processes and Practices for Transforming Underdeveloped Economies into Developed Economies (PPP-UD 2025)*, Advances in Economics, Business and Management Research 357,
https://doi.org/10.2991/978-94-6463-894-3_27

1 Introduction

Black pepper is among the most extensively traded spices worldwide, appreciated for its culinary and health advantages. It is mainly grown in tropical areas, with India, Vietnam, Indonesia, and Brazil being the top producers. In India, the states of Kerala, Karnataka, and Tamil Nadu play significant roles in production. The market experiences high volatility, affected by supply chain challenges, seasonal changes, and variations in quality creating uncertainty and underscoring the necessity for precise price forecasting to support decision-making and reduce risk.

Previous research [1] has utilized models such as ARIMA and ARCH to predict black pepper prices in Karnataka. While these models are effective for analyzing short-term trends and volatility, they face difficulties with non-linear trends and abrupt shocks, which limits their applicability in dynamic agricultural markets.

This proposed work to develop a strong forecasting model for black pepper prices, specifically designed for the SamparkBindhu [17] platform a digital trading center for farmers and traders. The model will produce forecasts for a two-week period based on historical price data. Additionally, it will tackle critical challenges such as inconsistent data, missing values, and the balance between model transparency and accuracy, thereby assisting traders, analysts, and platform managers in making more informed decisions.

Accurate price forecasting is crucial for traders to efficiently manage their procurement and logistics operations. Nevertheless, the existing manual forecasting methods are labor-intensive and prone to human error, which often results in delayed responses and inaccurate forecasts. Traditional models exacerbate the problem by yielding unreliable outcomes when faced with noisy, inconsistent, or incomplete dataset situations that frequently occur in agricultural market settings. The absence of automation not only raises operational costs but also hinders the decision-making process, requiring more human resources and diminishing overall market efficiency. These challenges underscore the pressing need for a more dependable, automated, and data-driven forecasting solution.

The main goal of this research is to create a dependable forecasting model for black pepper prices that satisfies practical performance standards and improves decision-making abilities for traders. The specific aims are as follows: Develop a price forecasting model for black pepper targeting a MAPE of under 5%. Integrate, prepare, and execute a thorough data modelling pipeline that includes data collection, preprocessing, and feature engineering. Design and implement a user-friendly interface to facilitate timely and informed decision-making for traders on the SamparkBindhu platform.

The rest of the manuscript is structured as follows. Section II reviews related literature on forecasting methods. Section III outlines the methodology, including data preprocessing and model selection. Section IV covers implementation and testing, while Section V presents results, emphasizing the performance of the MLR model with lag features. Section VI concludes the study and discusses future enhancements.

2 Literature Review

Traditional Econometric Models in Commodity Price Forecasting: Traditional econometric models such as ARIMA, AutoRegressive Conditional Heteroskedasticity (ARCH), and Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) have historically played a crucial role in forecasting commodity prices. Research, including [1], illustrates their effectiveness in capturing short-term fluctuations in black pepper prices. Nevertheless, these models struggle to account for non-linearities and external shocks [2], [3], [4]. Additionally, even when factors like weather are taken into account [5], their failure to demonstrate causality or adapt to real-time changes limits their overall effectiveness.

Deep Learning and Neural Networks for Price and Volatility Forecasting: The rise of AI models, particularly LSTM, signifies a major transformation in price forecasting methodologies. Studies [6], [7], and [8] highlight LSTM's capability in modelling non-linear relationships, surpassing traditional methods. However, issues such as small datasets, insufficient real-time adaptability, and the omission of macroeconomic factors remain challenges. Advancements like Convolutional Neural Networks (CNN) and Time-Delay Neural Network (TDNN) [9], [10], along with feature-rich LSTM models [11], show potential but continue to face difficulties regarding explainability and scalability.

Ensemble, Hybrid, and Machine Learning Approaches: Ensemble and hybrid modelling techniques are increasingly recognized for their ability to improve prediction accuracy in volatile market environments. Comparative analyses [12], [13] indicate that methods like Random Forest and XGBoost tend to outperform individual models, although they often lack contextual awareness. Creative integrations such as fuzzy-Support Vector Machine (SVM) [14] , fuzzified input data tuning for agriculture commodities price [15] and LSTM-Random Forest (RF) hybrids [16] attempt to fill some of these gaps, yet many still overlook the importance of real-time responsiveness and broader economic indicators. This trend underscores a growing focus on merging complex model frameworks with extensive datasets.

Integration of External Factors and Reinforcement Learning: Many current models fail to adequately incorporate essential external factors such as climate, policies, and market sentiment [6], [7], [9], [12]. This lack of integration undermines their robustness in volatile markets. Although Reinforcement Learning (RL) has significant potential, it is still not widely applied in adaptive forecasting frameworks [8], [9], [11], [13]. Future research suggests that merging RL with deep learning could create resilient, real-time systems that effectively respond to changing external signals.

The literature reviewed indicates a methodological shift from traditional econometric models to contemporary AI-driven forecasting. While traditional tools effectively handle linear and short-term trends, they struggle with non-linearities and market disruptions. Deep learning models exhibit potential but are frequently constrained by data deficiencies and absent contextual features. A targeted strategy employing GRU-based hybrids with lag and moving average inputs is crucial for producing accurate and interpretable forecasts of black pepper prices.

3 Methodology

This section describes the technical resources, tools, and methodologies utilized in the creation of the black pepper price forecasting system. It elaborates on the implementation environment, data sources, preprocessing methods, modeling framework, and deployment strategy, all of which are in accordance with the Team Data Science Process (TDSP) [18] lifecycle.

The implementation environment employed Python version 3.11.11, incorporating essential libraries such as pandas for data manipulation, numpy for numerical calculations, scikit-learn for training and assessing models, matplotlib for data visualization, and streamlit for developing the user interface. All experiments were performed on a CPU-based on-premise configuration.

This proposed work follows the TDSP methodology a structured, collaborative framework that outlines a well-defined lifecycle for executing data science projects as defined in figure (see Fig. 1).

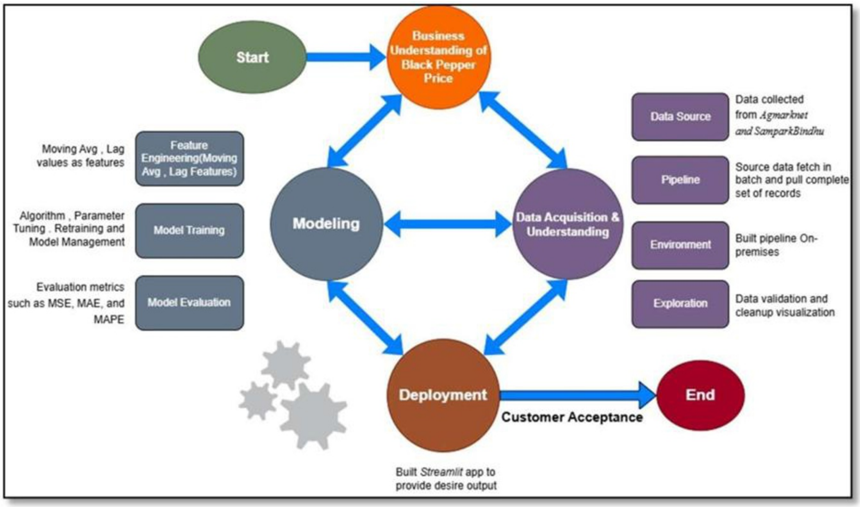

Fig. 1. Proposed Methodology

The main goal is to create a price forecasting model for black pepper to support traders and stakeholders on the *SamparkBindhu* platform. This model is designed to produce two-week price predictions based on historical pricing information. Stakeholders encompass local traders, market analysts, and administrators of *SamparkBindhu*.

Data collected from a variety of trustworthy sources, including governmental platforms like Agmarknet [19] and real-time market data from *SamparkBindhu* traders. The dataset spans over six years and includes attributes such as Date, Maximum Price, Minimum Price, and Modal Price. A thorough Exploratory Data Analysis (EDA) was performed to uncover price trends, analyze distribution patterns, and identify key features for modeling purposes. Visual representations (see Fig. 2) were employed to investigate seasonal behaviors and market fluctuations, thus creating a solid foundation for data-driven decision-making.

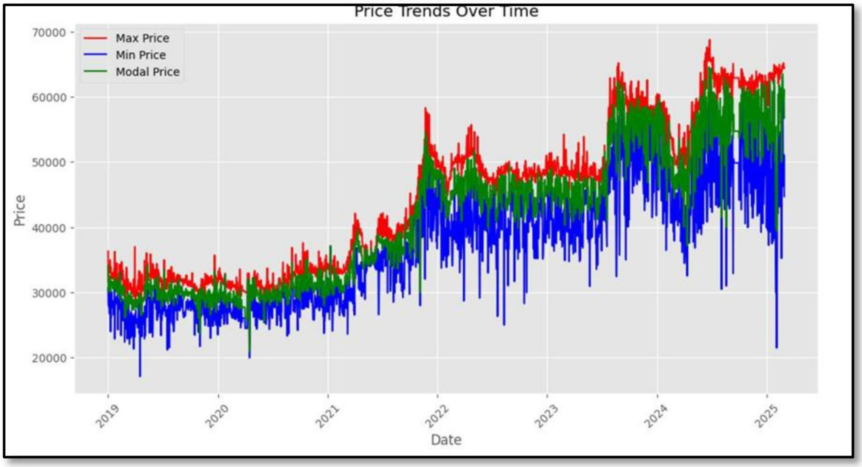

Fig 2. Price Trends Over Time (2019 – Feb 2025)

To maintain data integrity and consistency, missing values were filled in using statistical methods and prices from adjacent districts, while outliers were eliminated through the Interquartile Range (IQR) technique. Daily maximum prices were compiled across districts, with backfilling implemented as needed. Feature engineering involved creating weekly lag variables to reflect historical dependencies and computing moving averages to mitigate short-term variations, thereby improving the model's capacity to recognize temporal patterns.

Feature engineering played a crucial role in capturing the dynamics of the time series. Weekly lag variables were created to illustrate the influence of previous prices, and moving averages were computed to reduce short-term volatility, thus enhancing the model's ability to comprehend temporal relationships. The Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) plots prior to feature engineering, indicating significant correlations extending to Lag values (see Fig. 3). ACF following the incorporation of lag features, revealing a significant drop in autocorrelation, which suggests a successful representation of temporal relationships after choosing Lag 4.

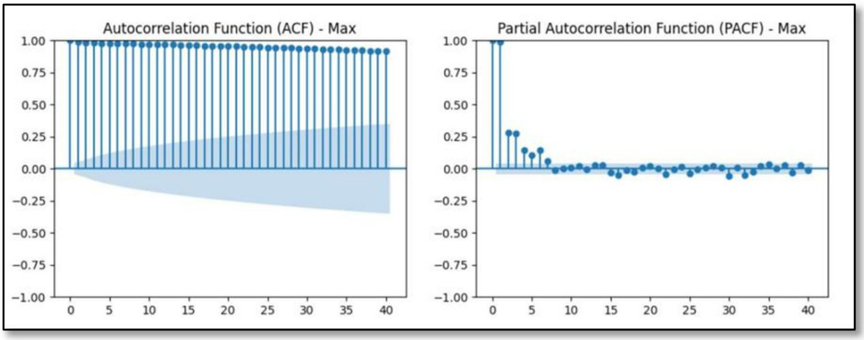

Fig 3. Autocorrelation check on dataset and diminishing partial autocorrelations

To determine the most effective forecasting method, an evaluation was conducted on both traditional and contemporary models. Statistical techniques such as SMA and ARIMA served as baseline comparisons, while machine learning approaches like MLR with lag features demonstrated enhanced accuracy. Additionally, deep learning methods, including LSTM and GRU, were assessed using smoothed and lagged data. Among these, the MLR model incorporating four lag features delivered the highest performance, achieving a MAPE of 0.22%. In contrast, deep learning models exhibited subpar performance, likely attributable to limited training data and inadequate tuning. Various model training dataset actual vs predicted values (see Fig.4).

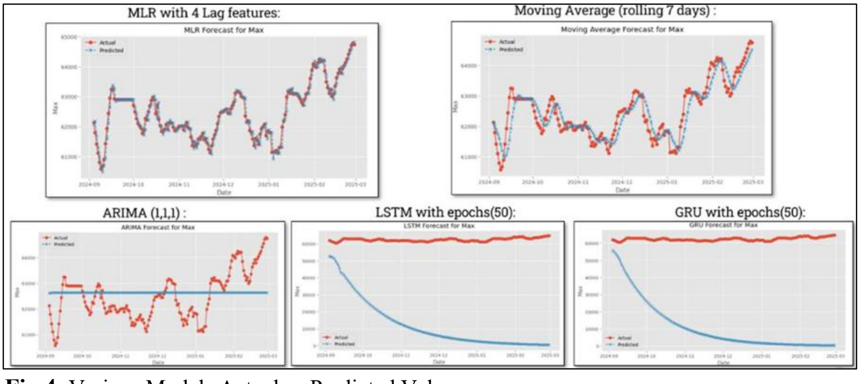

Fig 4. Various Models Actual vs Predicted Values

The architecture of the system adheres to a modular design to ensure maintainability and scalability. It commences with data ingestion from structured formats, followed by a preprocessing phase aimed at cleaning and feature engineering. In the modeling layer, forecasting models are trained and assessed. Subsequently, the forecasting module produces 15-day price predictions utilizing the chosen model. A visualization layer then graphically presents the results, including comparisons between actual and predicted values, along with essential performance metrics.

To facilitate practical use, the forecasting system was deployed as a Streamlit-based web application. Users can upload historical data files and receive interactive visualizations and predictive insights. The interface supports user-defined configurations such as district-level forecasting and custom date ranges, enabling adoption by spice traders on the *SamparkBindhu* platform.

4 Implementation And Testing

The processes of evaluation and verification conducted to assess the forecasting models developed for predicting black pepper prices. These models underwent thorough validation utilizing historical data, and various performance metrics were calculated to evaluate the accuracy and dependability of each method.

Among the models evaluated, the MLR model with Lag4 produces the most favorable outcomes, attaining the lowest figures across all three critical metrics: Mean Absolute Error (MAE) (137.55), Mean Squared Error (MSE) (35,607.62), and MAPE (0.2207%). These findings reflect a substantial degree of precision and slight divergence from actual market prices. In contrast, traditional models such as Moving Average and ARIMA exhibited elevated error rates, whereas deep learning models like LSTM and GRU performed considerably worse as mentioned in Table 1.

Table 1. Various Models Performance Metrics for Max Price

Model	MAE	MSE	MAPE %
Moving_Avg_Max	308.49	1,64,320.20	0.4935
MLR_Max_Lag1	152.78	42,295.71	0.2449
MLR_Max_Lag2	138.28	35,847.38	0.2218
MLR_Max_Lag3	137.88	35,624.03	0.2212
MLR_Max_Lag4	137.55	35,607.62	0.2207
MLR_Max_Lag5	138.14	35,846.73	0.2216
MLR_Max_Lag6	138.51	36,197.86	0.2222
ARIMA_Max	767.69	8,43,272.40	1.1672
LSTM_Max	50,765.08	2,77,86,71,000	81.126
GRU_Max	55,548.04	3,23,34,29,000	88.808

The daily forecasts for February 2025 were assessed in comparison to actual values by utilizing both absolute and percentage error metrics for maximum and minimum prices. The findings indicate that the anticipated maximum prices consistently exhibited low error margins, thereby affirming the reliability and accuracy of the MLR model in reflecting market trends. Conversely, the predictions for minimum prices displayed a greater degree of variability, implying that the model is more adept at forecasting maximum prices than minimum ones. These insights further validate the model's appropriateness for short-term forecasting, especially in tracking upper-bound price fluctuations.

In the course of the model validation process, historical data extending to January 2025 was utilized to train the forecasting model, and predictions were produced for the following period February 2025. This methodology aids in assessing the model's capacity to generalize beyond the training data and to replicate real-world forecasting.

Table 2. February 2025 Model Forecasting Performance

Model	MAE	MSE	MAPE %
MLR (4 Lags)	675.2691	8,50,737.21	1.0639

The validation outcomes indicate that the MLR model, which incorporates 4 lag features, sustained a robust predictive capability even when applied to unseen data. Although there was a minor rise in error metrics relative to the training performance, it still attained a MAPE of merely 1.06%, signifying a considerable degree of forecasting precision as shown in Table 2. This validates the model's relevance and resilience for short-term market predictions concerning black pepper prices.

To evaluate the practical relevance of the forecasting model, validation was conducted utilizing out-of-sample data from February 2025 a timeframe that was not part of the model training. The trained MLR model, which incorporates lag features, was employed to produce daily price forecasts, which were subsequently compared to the actual market prices recorded during the month.

The graph depicts a comparison between the actual maximum prices and the predicted values (see Fig. 5). The solid red line signifies the actual price variations, whereas the dashed blue line represents the model's forecasted prices. The vertical dashed line indicates the commencement of the forecast period (February 1, 2025).

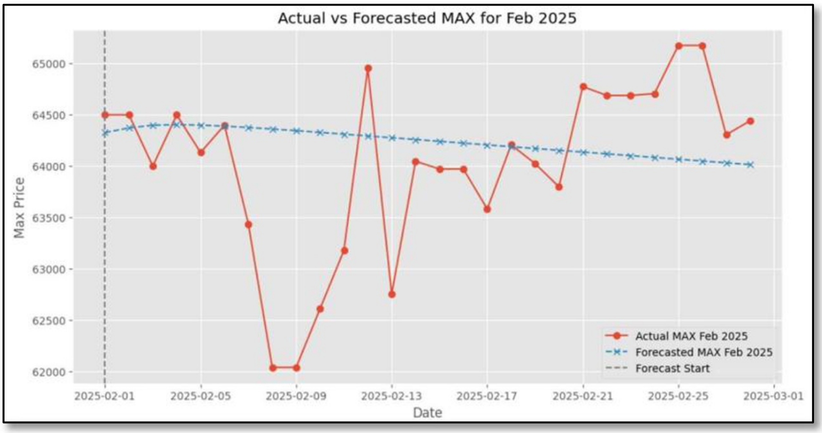

Fig 5. Actual vs Forecasting trends for Feb 2025 Month

5 Analysis And Results

This section outlines the results obtained from the forecasting models that were developed. Among the different models assessed, the MLR incorporating lag features proved to be the most precise, exhibiting low error rates across various metrics including MAE, MSE, and MAPE. These models are intended to assist in strategic planning for farmers, traders, and investors involved in the pepper market.

The predictive efficacy of the chosen best-performing model throughout the 6-month testing phase. It compares the actual smoothed maximum prices with their predicted equivalents. The close correspondence between the two curves highlights the model's strength and accuracy in reflecting real-world price variations (see Fig. 6)

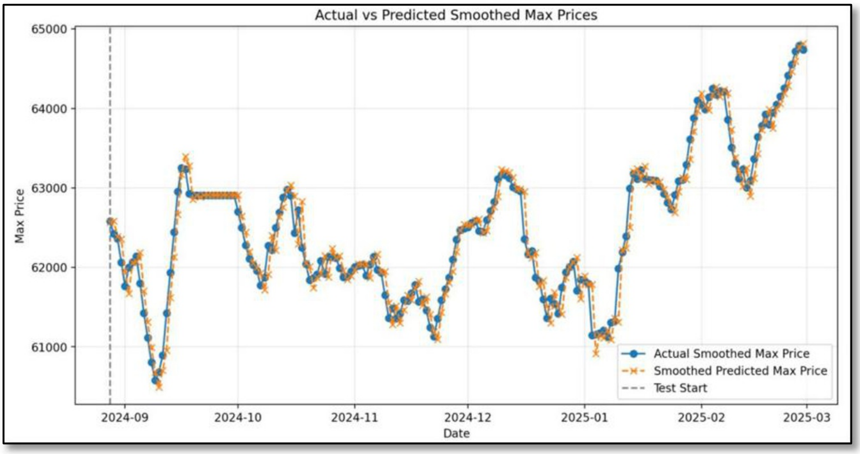

Fig 6. Actual vs Predicted Smoothed Max Price Values for Test Data Using Best Model

6 Conclusion and Future scope

This research introduces a comprehensive price forecasting system for black pepper, which has been developed utilizing a MLR model enhanced with lag-based features. The model demonstrates exceptional predictive accuracy, evidenced by an impressively low MAPE of 0.22%. Its straightforwardness, interpretability, and efficiency render it highly suitable for practical implementation on platforms such as SamparkBindhu.

In the future, various improvements are anticipated to expand the model's range and flexibility. Upcoming iterations of the system are designed to integrate external data sources, including meteorological factors such as precipitation, temperature, and humidity, along with supply chain metrics like import-export volumes and inventory levels. The incorporation of real-time data via API connections will facilitate ongoing updates and dynamic forecasting capabilities. Furthermore, the model can be adapted to provide price predictions for additional agricultural products, encompassing spices, grains, and pulses, thereby establishing it as a scalable solution for wider market applications.

7 References

[1] H. B. Mallikarjuna, A. Paul, A. Paul, A. S. Noel, and M. Sudheendra, “Forecasting of Black Pepper Price in Karnataka State: An Application of ARIMA and ARCH Models,” Int J Curr Microbiol Appl Sci, vol. 8, no. 01, pp. 1486–1496, Jan. 2019, doi: 10.20546/ijcmas.2019.801.159.
[2] V. Hosamani and T. C. M, “IJRTI1810018 International Journal for Research Trends and Innovation (www.ijrti.org) 94 PRICE FORECASTING & INTEGRATION OF DOMESTIC MARKETS OF SPICES IN INDIA-AN ECONOMETRIC ANALYSIS,” © 2018 IJRTI |, vol. 3, no. 10, 2018, [Online]. Available: www.ijrti.org
[3] A. Jain, S. Marvaniya, S. Godbole, and V. Munigala, “A Framework for Crop Price Forecasting in Emerging Economies by Analyzing the Quality of Time-series Data,” Sep. 2020, [Online]. Available: http://arxiv.org/abs/2009.04171
[4] S. S. S. A Kuruvila and K. Manojkumar, “Price behaviour of black pepper in Indian and international markets: a comparative analysis,” Journal of Spices and Aromatic Crops, pp. 27–33, Jan. 1970, doi: 10.25081/josac.2019.v28.i1.5741.
[5] B. Adebanjo Adeniyi, F. Ogbeide Okogun, O. Opiribo, A. Adeniyi α, F. Ogbeide Okogun σ, and O. Opiribo ρ, “Assessing the Price Relationship and Weather Impact on Selected Pairs of Closely Related Commodities,” 2021.
[6] S. Suman, P. Kaushik, S. S. N. Challapalli, B. P. Lohani, P. Kushwaha, and A. D. Gupta, “Commodity Price Prediction for making informed Decisions while trading using Long Short-Term Memory (LSTM) Algorithm,” in Proceedings of 5th International Conference on Contemporary Computing and Informatics, IC3I 2022, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 406–411. doi: 10.1109/IC3I56241.2022.10072626.
[7] P. B. Deepa and J. Daisy, “Deep Learning Based Prediction of Commodity Prices Using LSTM,” in Proceedings of the 4th International Conference on Smart Electronics and Communication, ICOSEC 2023, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 1708–1715. doi: 10.1109/ICOSEC58147.2023.10276011.
[8] M. Yusup, S. Y. Joko Prasetyo, and T. Wellem, “Evaluation of Prediction Accuracy in ARIMA and LSTM Algorithms for Agricultural Commodity Prices,” in 2024 3rd International Conference on Creative Communication and Innovative Technology, ICCIT 2024, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICCIT62134.2024.10701123.

[9] N. B. Van Le, Y. S. Seo, and J. H. Huh, “AgTech: Volatility Prediction for Agricultural Commodity Exchange Trading Applied Deep Learning,” IEEE Access, 2024, doi: 10.1109/ACCESS.2024.3479868.
[10] G. K. Jha and K. Sinha, “Time-delay neural networks for time series prediction: An application to the monthly wholesale price of oilseeds in India,” Neural Comput Appl, vol. 24, no. 3–4, pp. 563–571, Mar. 2014, doi: 10.1007/s00521-012-1264-z.
[11] S. Song, X. Pan, M. Guo, Q. Lang, and X. Liu, “Futures Price Forecasting Based on the Feature Fusion LSTM Model Using Long-Term Price Patterns,” in Chinese Control Conference, CCC, IEEE Computer Society, Jul. 2021, pp. 8193–8198. doi: 10.23919/CCC52363.2021.9549943.
[12] H. Ben Ameur, S. Boubaker, Z. Ftiti, W. Louhichi, and K. Tissaoui, “Forecasting commodity prices: empirical evidence using deep learning tools,” Ann Oper Res, vol. 339, no. 1–2, pp. 349–367, Aug. 2024, doi: 10.1007/s10479-022-05076-6.
[13] S. Vighnesh Dhanraj and M. M. Wakwadi, “Prediction the best price for crops and spices using Machine Learning,” International Journal of Scientific Research and Engineering Development, 2023, [Online]. Available: www.ijsred.com
[14] Y. Zhang and S. Na, “A novel agricultural commodity price forecasting model based on fuzzy information granulation and MEA-SVM model,” Math Probl Eng, vol. 2018, 2018, doi: 10.1155/2018/2540681.
[15] G. Hegde, V. R. Hulipalled, and J. B. Simha, “Fuzzified input data tuning for agriculture commodities price prediction,” IAES International Journal of Artificial Intelligence, vol. 13, no. 3, pp. 2725–2735, Sep. 2024, doi: 10.11591/ijai.v13.i3.pp2725-2735.
[16] L. Madaan, A. Sharma, P. Khandelwal, S. Goel, P. Singla, and A. Seth, “Price forecasting & anomaly detection for agricultural commodities in India,” in COMPASS 2019 - Proceedings of the 2019 Conference on Computing and Sustainable Societies, Association for Computing Machinery, Inc, Jul. 2019, pp. 52–64. doi: 10.1145/3314344.3332488.
[17] [Sadananda Murthy, “Samparkbindhu,” website. Accessed: Apr. 13, 2025. [Online]. Available: https://samparkbindhu.in/
[18] Sagar Lad, “TDSP : Team Data Science Process,” Website. Accessed: Apr. 05, 2025. [Online]. Available: https://sagu94271.medium.com/tdsp-team-data-science-process-cf08f5b0d1dd
[19] India Govt, “AGMarket Govt Website,” Website. Accessed: Apr. 05, 2025. [Online]. Available: https://agmarknet.gov.in/

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.