Journal of Agribusiness in Developing and Emerging Economies

Received 10 May 2026
Revised 23 June 2026
Accepted 27 July 2026

Temporal feature engineering for agricultural commodity price forecasting in Nigeria: evaluating machine learning, deep learning and time-series approaches

Damilare Ogunjobi
Department of Computer Engineering, Federal University of Technology Akure, Akure, Nigeria
Rania Shibl
Faculty of Science and Engineering, Southern Cross University, Gold Coast, Australia, and
Shahrzad Saremi
University of the Sunshine Coast – Moreton Bay Campus, City of Moreton Bay, Australia

Abstract

Purpose – This study investigates the effectiveness of machine learning, deep learning and traditional time-series approaches for predicting agricultural commodity prices in Nigeria, focussing on improving forecasting reliability through temporal feature engineering and time-aware evaluation using the WFP Nigeria dataset.

Design/methodology/approach – The study uses 60,566 observations across Nigerian commodity markets (2002–2025). Linear Regression, Random Forest, ARIMA, and LSTM models were evaluated using lag and rolling-average temporal features, with walk-forward validation preserving temporal integrity for realistic assessment.

Findings – Machine learning models outperformed traditional statistical approaches. Random Forest ($R^2 = 0.497$) exceeded Linear Regression ($R^2 = 0.465$), while ARIMA yielded a negative R^2 (-0.259). LSTM achieved its strongest performance for maize ($R^2 = 0.640$) and sorghum ($R^2 = 0.468$), while several highly volatile commodities produced negative R^2 values despite extended training and Early Stopping. These results highlight the role of commodity-specific characteristics in price forecasting, with walk-forward validation revealing considerable temporal performance variation.

Research limitations/implications – The study relies primarily on historical price data and does not incorporate exogenous variables such as weather, inflation, transportation costs or policy interventions.

Originality/value – This study contributes to the agricultural forecasting literature by integrating temporal feature engineering, chronological validation, walk-forward evaluation, and comparative modelling within a unified framework using a Nigerian commodity price dataset, providing commodity-level evidence on the varying effectiveness of machine learning, deep learning and traditional statistical approaches under realistic forecasting conditions.

Keywords Agricultural price prediction, Machine learning, Time-series forecasting, Random forest, LSTM, Temporal feature engineering, Nigeria

Paper type Research article

Funding statement: The authors received no specific funding for this research, authorship, or publication of this article.

Conflict of interest statement: The authors declare that there is no conflict of interest regarding the publication of this paper.

Journal of Agribusiness in Developing and Emerging Economies
Emerald Publishing Limited
e-ISSN: 2044-0847
p-ISSN: 2044-0839
DOI 10.1108/JADEE-05-2026-0385

Plain language summary
Agricultural food prices in Nigeria often change unpredictably due to factors such as seasonal production, transportation challenges, inflation, and market instability. These price fluctuations affect farmers, traders, consumers, and policymakers by making planning and decision-making more difficult. This study explored how artificial intelligence and data-driven forecasting techniques can help predict agricultural commodity prices more accurately. Using food price data from the World Food Programme (WFP) covering the period from 2002 to 2025, different forecasting models were tested, including machine learning, deep learning, and traditional statistical approaches. The results showed that machine learning models generally outperformed traditional statistical methods in predicting price movements. However, model performance varied depending on the type of commodity and the level of market volatility. The study also found that using historical price patterns and proper time-based validation methods significantly improved forecasting reliability. The findings of this research can support farmers, traders, and policymakers by improving market planning, reducing uncertainty, and contributing to better food security and agricultural decision-making in Nigeria and similar developing economies.

1. Introduction

Agricultural commodity price volatility remains a persistent and multifaceted challenge in Nigeria, where food systems are highly sensitive to seasonal variability, structural inefficiencies, and macroeconomic instability (Obayelu *et al.*, 2024). Price fluctuations directly influence household food access, farmer income stability, and national food security outcomes (Gilbert and Morgan, 2010; Obayelu *et al.*, 2024). In regions where a large proportion of the population relies on agriculture for both subsistence and income, unpredictable price movements can exacerbate poverty, distort market incentives, and undermine policy interventions aimed at stabilizing food systems. In Nigeria, agricultural markets are influenced by a complex interaction of local and external factors. Seasonal harvest cycles, transportation constraints, storage limitations, and market fragmentation contribute to price variability at the local level, while macroeconomic factors such as inflation, exchange rate fluctuations, and trade policies further shape market dynamics (Aker, 2010). These conditions lead to heterogeneous price behaviour across commodities and regions, making accurate forecasting particularly challenging.

Reliable forecasting of agricultural commodity prices is essential for improving decision-making across the value chain (Sun *et al.*, 2023). Farmers can optimize production and harvesting strategies, traders can enhance inventory planning, and policymakers can design targeted interventions to stabilize markets and improve food security. However, the dynamic, non-linear, and time-dependent nature of agricultural price data presents significant challenges for predictive modelling. A key limitation in existing studies is the inadequate treatment of temporal dependencies and improper evaluation of time-series data. Many approaches fail to preserve the chronological structure of the data, leading to unrealistic performance estimates (Makridakis *et al.*, 2018; Sun *et al.*, 2023). Additionally, temporal feature engineering is often limited, reducing the ability of models to capture meaningful patterns in price movements.

This study addresses these challenges by developing a time-aware modelling framework for agricultural price prediction using the World Food Programme (WFP) Nigeria dataset. The proposed approach incorporates multiple temporal features, including lag variables and rolling averages, and employs a chronological data-splitting strategy to prevent information leakage. Furthermore, walk-forward validation is implemented to provide a more robust assessment of model performance over time. The study evaluates multiple modelling approaches, including machine learning, traditional time-series methods, and deep learning techniques, within a unified framework. By comparing these approaches and analyzing their performance across different commodities, the study provides insights into the strengths and limitations of each method in real-world agricultural markets. This research contributes to the development of more reliable and context-aware forecasting models by emphasizing proper temporal handling, robust validation, and feature-driven modelling strategies. The findings are particularly relevant for stakeholders seeking to develop data-driven decision-support systems in emerging agricultural markets.

Despite growing interest in agricultural price forecasting, several important gaps remain in the literature. First, relatively few studies focussing on Nigerian agricultural markets have systematically examined the contribution of temporal feature engineering techniques, such as multiple lag variables and rolling statistics, to forecasting performance. Second, many existing studies rely on random train-test splits or single evaluation periods, which can lead to information leakage and overly optimistic performance estimates. Third, there is limited evidence comparing traditional statistical models, machine learning approaches, and deep learning methods within a consistent time-aware forecasting framework using Nigerian commodity price data.

This study addresses these gaps by developing a time-aware forecasting framework that integrates temporal feature engineering, chronological data splitting, and walk-forward validation using a large-scale Nigerian agricultural commodity price dataset. Unlike many previous studies, the research explicitly evaluates the contribution of engineered temporal features while examining the behaviour of traditional statistical, machine learning, and deep learning approaches under realistic forecasting conditions. The findings provide new evidence on the importance of temporal feature design and commodity-specific forecasting performance in Nigerian agricultural markets.

2. Literature review

Agricultural price forecasting has long been recognized as a critical component of economic planning, food security management, and market stabilization, particularly in developing economies where agricultural systems play a central role in livelihoods, such as Nigeria. The ability to accurately predict commodity prices enables stakeholders, including farmers, traders, and policymakers, to make informed decisions that optimize production, improve supply chain efficiency, and mitigate the adverse effects of market volatility. In Nigeria, where agricultural markets are often characterized by structural inefficiencies, limited infrastructure, and fragmented supply chains (Aker, 2010; Obayelu *et al.*, 2024), the importance of reliable price forecasting becomes even more pronounced. Over the past decades, a wide range of methodologies have been proposed for agricultural price prediction, broadly categorized into traditional statistical approaches, machine learning techniques, and, more recently, deep learning models. Each of these approaches offers unique advantages and limitations, particularly in relation to the complex and dynamic nature of agricultural markets.

2.1 Traditional time-series forecasting models

Early studies in agricultural price forecasting predominantly relied on statistical and econometric models, with the Autoregressive Integrated Moving Average (ARIMA) model being one of the most widely used approaches. ARIMA models are designed to capture temporal dependencies within time-series data by modelling relationships between past observations and future values (Box *et al.*, 2015). Their mathematical foundation and interpretability have made them a standard tool in economic forecasting. However, the effectiveness of ARIMA models is contingent upon several assumptions, including linearity, stationarity, and the absence of structural breaks. In real-world agricultural markets, these assumptions are often violated. Commodity prices are influenced by a wide range of factors, including weather variability, transportation constraints, policy interventions, and global market trends. These factors introduce non-linearities, abrupt changes, and irregular patterns that are difficult to capture using purely statistical models.

Empirical studies have highlighted the limitations of ARIMA in modelling complex agricultural price dynamics. Makridakis *et al.* (2018) emphasized that traditional statistical models often struggle in environments characterized by high uncertainty and non-linear interactions. Similarly, Wang *et al.* (2020) noted that while ARIMA performs well in controlled settings, its predictive accuracy declines when applied to volatile real-world datasets. These limitations have motivated the exploration of alternative approaches that can better capture the complexity of agricultural markets.

2.2 Emergence of machine learning approaches

The advancement of computational power and data availability has facilitated the adoption of machine learning techniques for forecasting tasks. Unlike traditional statistical models, machine learning algorithms do not rely on strict assumptions about data distribution or linearity. Instead, they learn patterns directly from data, making them well-suited for modelling complex, non-linear relationships. Linear Regression remains a commonly used baseline model due to its simplicity and interpretability. However, its effectiveness is limited in scenarios where relationships between variables are non-linear or influenced by multiple interacting factors. As a result, more sophisticated models such as Random Forest have gained popularity in forecasting applications.

Random Forest, introduced by Breiman (2001), is an ensemble learning method that constructs multiple decision trees and aggregates their predictions to improve accuracy and generalization. Its ability to handle high-dimensional data, capture non-linear interactions, and resist overfitting makes it particularly suitable for agricultural datasets, which often contain a mix of categorical and numerical features.

Several studies have demonstrated the superiority of machine learning models over traditional statistical approaches in agricultural price prediction. For example, Chen *et al.* (2021) developed an automated agricultural price prediction system using machine learning techniques and reported improved performance compared to conventional models. Similarly, Sharma *et al.* (2020) highlighted the growing role of machine learning in precision agriculture, emphasizing its ability to enhance decision-making through data-driven insights. Despite these advancements, the effectiveness of machine learning models is highly dependent on feature representation. Many studies rely on static features such as commodity type, geographic location, and market identifiers, which do not adequately capture the temporal dynamics inherent in price data. As a result, predictive performance may remain limited even when using advanced algorithms.

2.3 Temporal feature engineering in machine learning

One of the most critical aspects of time-series forecasting is the incorporation of temporal dependencies. Agricultural prices are inherently time-dependent, with current values influenced by historical trends, seasonal cycles, and market behaviour. Feature engineering techniques such as lag variables, rolling averages, and trend indicators are commonly used to capture these temporal relationships. Lag features represent previous values of the target variable and are essential for modelling short-term dependencies (James *et al.*, 2013; Sun *et al.*, 2023). Rolling averages, on the other hand, provide a smooth representation of historical trends, reducing the impact of short-term fluctuations and noise. These features enable machine learning models to approximate time-series behaviour without explicitly modelling temporal sequences.

However, many existing studies do not fully exploit the potential of temporal feature engineering. In some cases, only a single lag feature is used, which may be insufficient for capturing complex patterns such as seasonality and multi-step dependencies. Furthermore, inadequate feature design can limit the ability of models to generalize across different time periods and market conditions. Recent research suggests that combining multiple temporal features, including lag variables, rolling statistics, and trend-based indicators, can significantly enhance predictive performance in time-series forecasting applications (Sun *et al.*, 2023; Tran *et al.*, 2023). For example, Tran *et al.* (2023) reported that the effectiveness of machine learning models in agricultural price forecasting is strongly influenced by the quality of temporal feature representation, while Zhao *et al.* (2024) found that properly engineered lag and trend variables substantially improved predictive accuracy across multiple agricultural market datasets. By incorporating multiple lag values, rolling statistics, and derived trend indicators, models can capture richer representations of time-dependent behaviour (Sun *et al.*, 2023). This approach bridges the gap between traditional time-series modelling and machine learning, enabling more accurate and robust predictions.

2.4 Deep learning for time-series forecasting

In recent years, deep learning models have emerged as powerful tools for time-series analysis, particularly in applications involving sequential data. Among these, Long Short-Term Memory (LSTM) networks have gained widespread attention due to their ability to capture long-term dependencies and complex temporal relationships (Hochreiter and Schmidhuber, 1997). LSTM networks are a type of recurrent neural network (RNN) designed to address the limitations of traditional RNNs, such as vanishing and exploding gradients. By incorporating memory cells and gating mechanisms, LSTM models can retain relevant information over extended time periods, making them well-suited for forecasting tasks. In the applications of LSTM in agricultural price prediction have shown promising results, Gu *et al.* (2022) demonstrated that attention-based LSTM models can effectively capture complex temporal patterns in commodity price data. Similarly, other studies have reported improved performance when using deep learning models compared to traditional approaches, particularly in datasets with strong sequential dependencies.

Recent advances in agricultural price forecasting have also explored hybrid and ensemble approaches that combine the strengths of statistical, machine learning, and deep learning techniques. Examples include ARIMA-LSTM hybrid models, XGBoost-based forecasting frameworks, and transformer-based architectures such as the Temporal Fusion Transformer (TFT), which have demonstrated improved performance in complex forecasting environments (Ray *et al.*, 2023; Zhao *et al.*, 2024). These approaches can capture both linear and non-linear temporal relationships while incorporating multiple sources of information. However, they typically require greater computational resources, more extensive hyperparameter tuning, and larger datasets. Consequently, the present study focuses on establishing a robust baseline comparison using widely adopted machine learning, deep learning, and traditional time-series approaches, while the investigation of advanced hybrid and transformer-based methods is identified as an important direction for future research.

2.5 Challenges and limitations in existing research

Despite the significant progress in agricultural price forecasting, several methodological challenges persist in the literature. One major issue is the improper handling of time-series data during model evaluation. Many studies employ random train-test splits, which violate the chronological structure of the data and introduce information leakage (Makridakis *et al.*, 2018; Sun *et al.*, 2023). This leads to inflated performance metrics that do not reflect real-world forecasting conditions.

Another limitation is the reliance on single train-test evaluations, which may not capture the variability of model performance across different time periods. In time-series forecasting, patterns and relationships can change over time due to evolving market conditions. As a result, evaluating models on a single split may provide an incomplete and potentially misleading assessment of their reliability.

Furthermore, there is a lack of comprehensive comparative studies that evaluate multiple modelling approaches within a consistent framework. Many studies focus on a single model or a narrow set of techniques, making it difficult to draw general conclusions about their relative effectiveness.

Additionally, comparatively fewer agricultural price forecasting studies have been conducted in African markets than in developed economies, limiting the availability of context-specific evidence for countries such as Nigeria (Sun *et al.*, 2023; Tran *et al.*, 2023). Most existing studies are conducted in developed economies or controlled environments, where market conditions differ significantly from those in emerging economies. This limits the applicability of their findings to Nigeria, where market dynamics are influenced by unique structural and economic factors.

Recent reviews further emphasize that differences in evaluation methodology contribute substantially to inconsistencies in reported forecasting performance. Sun *et al.* (2023) observed that studies employing random train-test splits often report higher predictive accuracy than studies using chronologically consistent validation strategies, highlighting the importance of time-aware evaluation in agricultural forecasting research.

2.6 Summary of literature

The reviewed literature demonstrates significant progress in the application of statistical, machine learning, and deep learning approaches to agricultural price forecasting. Traditional models such as ARIMA provide a strong theoretical foundation but are limited in their ability to capture non-linear and volatile price dynamics. Machine learning models offer improved flexibility and predictive performance but often depend heavily on feature design and may fail to fully incorporate temporal dependencies. Deep learning approaches, particularly LSTM networks, show promise in modelling sequential data but are sensitive to data characteristics and may require large datasets for effective training. Despite these advancements, several challenges persist, including inadequate handling of temporal structures, inconsistent evaluation methodologies, and limited research focused on emerging economies. These issues highlight the need for more robust, context-aware, and methodologically sound approaches to agricultural price forecasting.

A cross-study examination of recent literature reveals a consistent trend toward the use of increasingly sophisticated forecasting frameworks. Studies conducted in developing and emerging economies generally report that machine learning and deep learning models outperform traditional statistical approaches when agricultural prices exhibit strong non-linear behaviour and volatility (Sun *et al.*, 2023; Tran *et al.*, 2023; Zhao *et al.*, 2024). However, the magnitude of performance improvement varies considerably across commodities, datasets, and evaluation methodologies. Recent evidence also suggests that forecasting accuracy is strongly influenced by temporal feature design, validation strategy, and market-specific characteristics rather than model complexity alone. These findings highlight the need for context-aware evaluations and reinforce the importance of robust temporal modelling frameworks for agricultural price forecasting.

3. Methodology

This section presents a comprehensive description of the data sources, preprocessing techniques, feature engineering procedures, modelling approaches, and evaluation strategies employed in this study. The methodology is carefully designed to ensure reproducibility, robustness, and adherence to best practices in time-series modelling, particularly in addressing common issues such as data leakage, improper validation, and inadequate feature representation.

3.1 Dataset description

The dataset used in this study was obtained from the WFP, which provides publicly accessible data on food prices across multiple countries. The Nigerian subset of the dataset contains historical records of agricultural commodity prices across various markets and administrative regions.

The dataset consists of 60,566 observations and includes both categorical and numerical variables. Key attributes include.

(1) Commodity type

(2) Administrative region (admin1)

(3) Market location

(4) Observation date

(5) Price and USD price

The price variable represents the market price of a given commodity at a specific location and time and serves as the primary target variable for prediction. The dataset spans multiple years and covers a diverse set of commodities, providing a rich temporal and spatial representation of agricultural price behaviour. Table 1 provides a summary of the key features used in the dataset. Figure 1 illustrates the variation of agricultural commodity prices over time, highlighting temporal fluctuations and seasonal patterns.

Table 1. Dataset feature description

Feature	Description
Commodity	Type of agricultural product
Admin1	State or administrative region
Market	Market location
Date	Observation timestamp
Price	Commodity price (local currency)
Usdprice	Price in USD
Year	Extracted year
Month	Extracted month

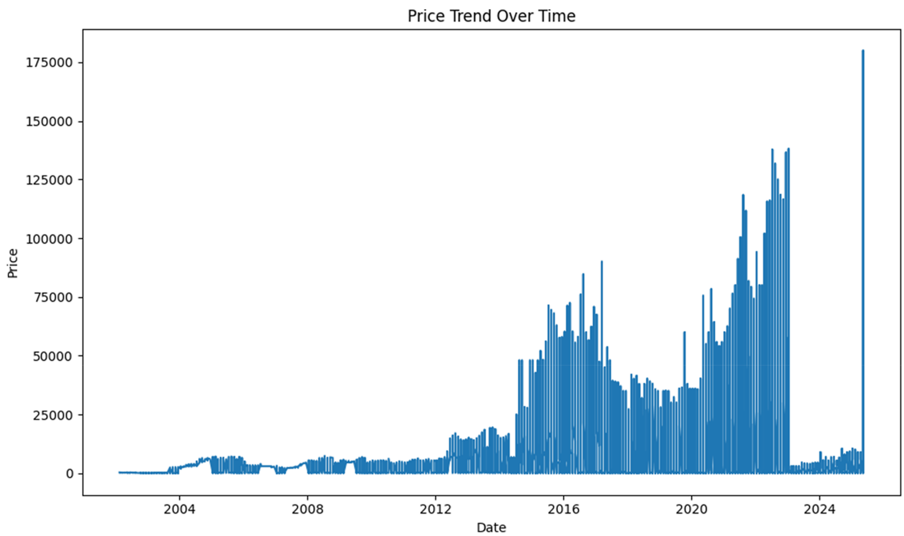

Figure 1. Temporal Trend of Agricultural Commodity Prices Over Time. (Source: Authors' own work)

The dataset covers fourteen administrative regions across Nigeria, representing diverse geographical, economic, and market conditions. These regions include both northern and southern states and capture heterogeneous agricultural market behaviours across the country. Figure 2 illustrates the spatial distribution of the administrative regions included in the study.

The geographical distribution of the selected regions enhances the diversity of the dataset by incorporating markets with varying levels of agricultural activity, price volatility, infrastructure development, and regional trade dynamics. This improves the representativeness of the dataset for evaluating agricultural commodity price forecasting models within Nigeria.

3.2 Data preprocessing

Data preprocessing was conducted to ensure data quality and prepare the dataset for modelling. Initially, the dataset was inspected to identify inconsistencies, missing values, and incorrect data types. The price and usdprice columns were converted to numeric format, with non-numeric entries coerced into missing values. The date column was converted into a datetime format to enable temporal feature extraction. Records with missing values in critical fields (price, usdprice, and date) were removed, as these variables are essential for both modelling and temporal analysis. The removal of missing values resulted in minimal data loss, preserving the integrity of the dataset. Following this, the dataset was sorted chronologically to maintain

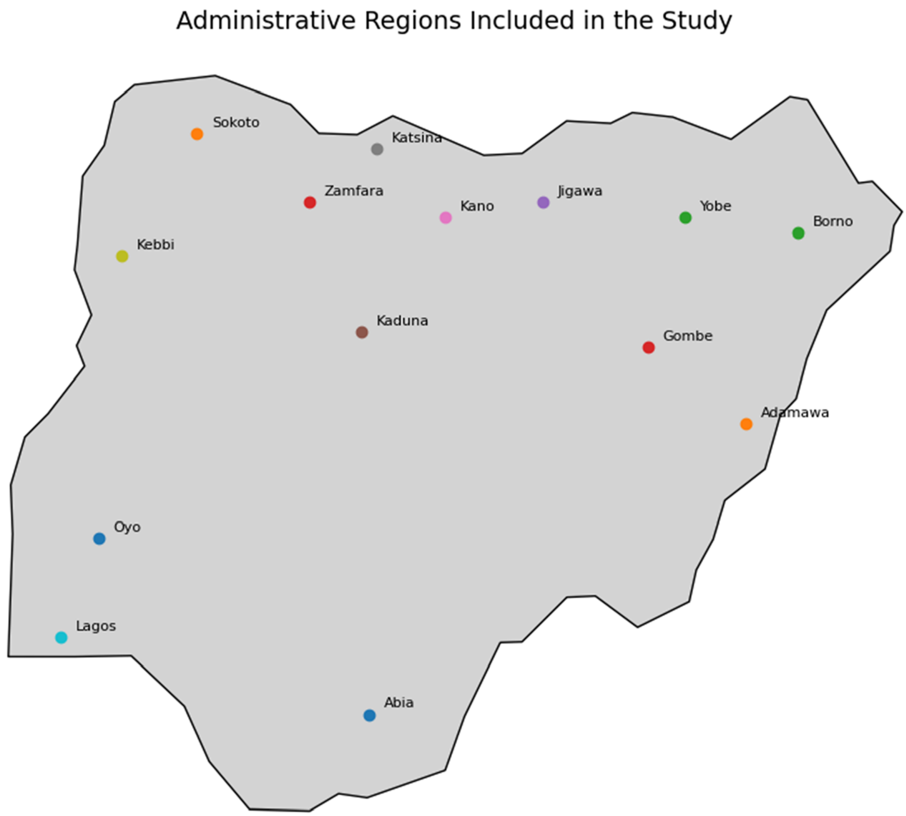

Figure 2. Administrative regions included in the WFP Nigeria food price dataset. (Source: Authors' visualization based on WFP Nigeria food price dataset)

temporal consistency. This step is crucial for ensuring that feature engineering operations, such as lag computation, are performed correctly and that models are trained on past data only. In addition, the target variable was constructed to enable forecasting. The target was defined as the future price of each commodity by shifting the price column one observation ahead within each commodity–administrative-region (admin1) group after chronological sorting. Thus, the forecasting horizon corresponds to the next available observation within a given commodity-admin1 series rather than a fixed calendar interval. This grouped shifting procedure ensures that future values are predicted only from historical observations belonging to the same commodity-region series, thereby preventing information leakage across independent series. This formulation allows the model to predict the next time step's price based on historical and current information while preserving temporal ordering.

Table 2 summarizes key statistics of the dataset after preprocessing. The dataset includes multiple commodities traded across different markets and administrative regions in Nigeria, providing a comprehensive representation of spatial and temporal price dynamics.

3.3 Temporal feature engineering

Feature engineering plays a critical role in improving model performance, particularly in time-series forecasting tasks where temporal dependencies are fundamental. To capture these dependencies, multiple temporal features were engineered. The selection of lag_1, lag_2, lag_3, and rolling_mean_3 was guided by exploratory temporal analysis and the objective of capturing short-term price persistence while preserving sufficient observations for model training. Although longer seasonal lags may be informative in some forecasting applications, the selected lag structure provided a balance between temporal representation and data availability across commodity-region series with varying record lengths.

3.3.1 Lag features. Three lag features (lag_1, lag_2, lag_3) were created using grouped shift operations based on commodity and administrative region. These features represent historical price values and enable the model to learn short-term temporal dependencies. To capture temporal dependencies in price movements, lag features were generated using grouped operations based on commodity and administrative region. All lag features were generated using grouped shift operations within commodity-admin1 series. This ensured that historical information from one commodity or administrative region was not transferred to another series during feature construction.

3.3.2 Rolling mean feature. A rolling mean feature was computed using a three-period window applied to past observations. This feature provides a smooth representation of recent price trends and helps reduce the impact of noise and short-term fluctuations.

3.3.3 Price change feature. A price change feature was calculated as the difference between the current price and the previous price. This feature captures short-term trends and directional movement in prices. All temporal features were computed using strictly past data to prevent information leakage. Records with missing values resulting from these operations were removed. Table 3 summarizes the engineered temporal features.

Table 2. Dataset summary statistics

Statistic	Value
Total Records	60,566
Time Range	January 2002–May 2025
Number of Commodities	42
Number of Markets	60
Number of Administrative Regions	14
Currency Unit	Nigerian Naira (₦)
Missing Values Removed	Minimal

Table 3. Engineered features

Feature	Description
lag_1	Price at the previous time step
lag_2	Price two time steps before
lag_3	Price three time steps before
rolling_mean_3	Average of the last 3 prices
price_change	Difference between the current and previous price

3.4 Feature selection and encoding

The final feature set includes temporal, categorical, and engineered features. Temporal features (year and month) capture seasonal patterns, while categorical features (commodity and admin1) represent spatial and commodity-specific variations. Categorical variables were encoded using one-hot encoding, converting them into binary indicator variables. This ensures compatibility with machine learning algorithms while avoiding the introduction of ordinal relationships. One-hot encoding increased the dimensionality of the dataset by creating binary indicator variables for commodity and administrative-region categories. Although this approach introduces sparsity, it was considered appropriate because the resulting feature space remained manageable relative to the dataset size (60,566 observations). Alternative encoding approaches, such as target encoding, frequency encoding, and learnt embedding representations, may further reduce dimensionality and improve computational efficiency. However, one-hot encoding was selected because of its simplicity, interpretability, and widespread use in machine learning applications involving categorical variables. Table 4 summarizes the final feature set used for the model training.

3.5 Time-aware data splitting

To preserve the temporal integrity of the dataset, a chronological train-test split was employed. The dataset was divided using a fixed date threshold. Observations before the threshold were used for training, while later observations were reserved for testing. This approach ensures that the model is evaluated on unseen future data, reflecting real-world forecasting scenarios. To preserve the temporal structure of the dataset and avoid information leakage, a time-based splitting strategy was employed.

3.6 Modelling approaches

It is important to note that the models were evaluated at different analytical scopes. Linear Regression and Random Forest were trained and evaluated using the full multi-commodity dataset, encompassing all commodities and administrative regions. In contrast, ARIMA was implemented on a representative univariate commodity-region series (Maize in Katsina), while LSTM models were evaluated on selected commodities (Maize and Rice) to examine performance under varying levels of price volatility. Consequently, performance metrics should be interpreted within the context of each model's evaluation scope and are not intended as direct one-to-one comparisons across all modelling approaches.

Table 4. Final feature set

Feature type	Features
Temporal	Year, month
Categorical	Commodity, Admin1
Engineered	lag_1, lag_2, lag_3, rolling_mean_3, price_change

The selected models were chosen to represent four widely used forecasting paradigms commonly employed in agricultural price prediction research. Linear Regression was included as a simple and interpretable baseline model capable of identifying linear relationships between explanatory variables and commodity prices. Random Forest was selected as a representative ensemble machine learning approach due to its ability to capture non-linear relationships and complex feature interactions. ARIMA was included as a classical statistical time-series model that has been extensively applied in agricultural forecasting studies. Finally, LSTM was selected as a deep learning architecture specifically designed to capture temporal dependencies and sequential patterns in time-series data. Together, these models provide a broad representation of traditional statistical, machine learning, and deep learning forecasting approaches.

3.6.1 Linear regression. Linear Regression was used as a baseline model due to its simplicity and interpretability. It assumes a linear relationship between input features and the target variable. Linear Regression models the relationship between input features and the target variable as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \in$$

where.

(1) y is the predicted price

(2) x_i are input features

(3) β_i are model coefficients

(4) $\in$ is the error term

3.6.2 Random forest. Random Forest is capable of capturing non-linear relationships and complex feature interactions due to its ensemble structure of decision trees (Breiman, 2001). The model was configured with 100 trees and trained on the engineered feature set. The Random Forest model was trained using the engineered features to capture non-linear relationships in the data. The number of trees was selected based on preliminary experimentation and common practice in ensemble learning literature, providing a balance between predictive performance and computational efficiency. The configuration and training parameters used for the Random Forest model are summarized in Table 5.

3.6.3 ARIMA model. ARIMA relies on assumptions of linearity and stationarity, which are often violated in real-world agricultural datasets (Box *et al.*, 2015; Wang *et al.*, 2020). An ARIMA model was implemented as a baseline time-series model. The model was trained on a univariate price series (Maize in Katsina) and used to forecast future values. The ARIMA model is defined as ARIMA (p, d, *q*), where:

$$\varnothing(B)\,(1 - B)^d y_t = \theta\,(B) \in_t$$

where.

Table 5. Random forest table parameters

Parameter	Value
Number of Trees	100
Criterion	Mean Squared Error
Max Depth	Default
Random State	42

(1) B is the backshift operator

(2) p = autoregressive order

(3) d = differencing order

(4) q = moving average order

In this study, the model was specified as:

$$\text{ARIMA}\,(1,1,1)$$

Prior to model specification, stationarity diagnostics were conducted on the Maize–Katsina price series using both the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests. The ADF test rejected the presence of a unit root (ADF statistic = −3.89, $p = 0.002$), while the KPSS test failed to reject the null hypothesis of stationarity (KPSS statistic = 0.346, $p > 0.10$). Complementing these tests, the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots (Figures 3 and 4 shown below) provided visual confirmation of temporal dependence within the series: the ACF exhibits gradual decay across successive lags, while the PACF shows significant spikes confined to the first few lags, together supporting a low-order autoregressive specification. Based on these diagnostics, an ARIMA(1,1,1) model was selected as a baseline statistical forecasting approach; first-order differencing was retained to improve model robustness and maintain consistency with common ARIMA practice, even where diagnostics suggested relative stationarity. For the machine learning models, lag_1, lag_2, lag_3, and a three-period rolling mean were used to capture short-term temporal dependencies while limiting feature sparsity and data loss associated with longer lag windows.

3.6.4 LSTM model and architecture. LSTM models are particularly effective in capturing long-term temporal dependencies in sequential data (Hochreiter and Schmidhuber, 1997; Gu *et al.*, 2022). Accordingly, an LSTM was implemented using a single layer of 50 units followed

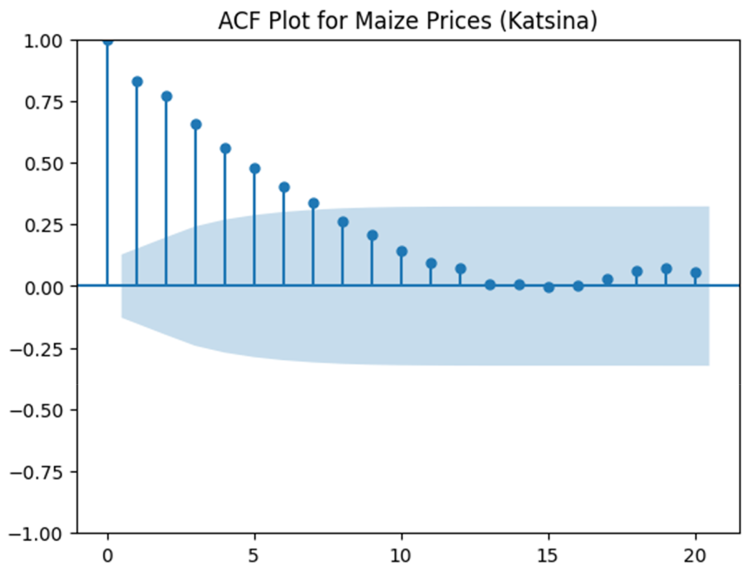

Figure 3. Autocorrelation function (ACF) Plot for maize prices in Katsina. The gradual decay in autocorrelation across lags indicates the presence of temporal dependence and supports the inclusion of autoregressive components in the ARIMA model. (Source: Authors' own work)

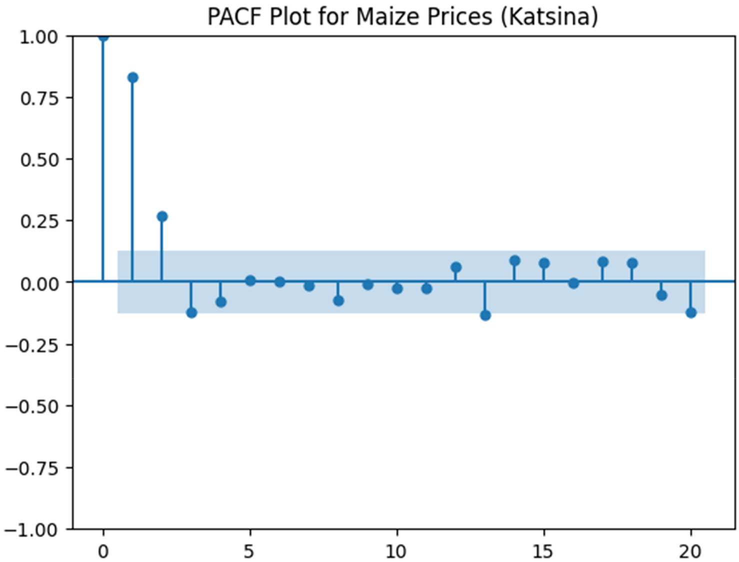

Figure 4. Partial Autocorrelation Function (PACF) Plot for maize prices in Katsina. Significant spikes at the first few lags suggest that a low-order autoregressive structure is appropriate for baseline time-series modelling. (Source: Authors' own work)

by a dense output layer for price prediction. The model was trained with the Adam optimizer (learning rate = 0.001) and Mean Squared Error (MSE) as the loss function, with a batch size of 16 and a maximum of 100 epochs. To improve convergence and prevent overfitting, Early Stopping was applied using validation loss as the monitoring metric, with a patience of 10 epochs and restoration of the best model weights. Table 6 summarizes the full architecture and training configuration.

3.7 Evaluation metrics

Model performance was evaluated using three widely used regression metrics.

3.7.1 Mean absolute error (MAE). Mean Absolute Error (MAE) measures the average absolute difference between predicted and actual values and provides an intuitive measure of forecasting accuracy in the same unit as the target variable.

Table 6. LSTM model architecture and training parameters

Layer/Parameter	Value
LSTM Layer	50 Units
Activation Function	ReLU
Dense Output Layer	1 Unit
Optimizer	Adam
Learning Rate	0.001
Loss Function	Mean Squared Error
Maximum Epochs	100
Batch Size	16
Early Stopping	Patience = 10
Restore Best Weights	True

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \dot{y}_i\right|$$

where.

(1) y_i = actual value

(2) $\dot{y}_i$ = predicted value

(3) n = number of observations

3.7.2 Mean squared error (MSE). Mean Squared Error (MSE) measures the average squared difference between predicted and actual values. Larger errors receive greater penalties because the differences are squared.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i-1}^{1}\left(y_i - \dot{y}_i\right)^2$$

where.

(1) y_i = actual value

(2) $\dot{y}_i$ = predicted value

(3) n = number of observations

3.7.3 Coefficient of determination (R^2*).* The coefficient of determination (R^2) measures the proportion of variance in the target variable explained by the model.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \dot{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y_i}\right)^2}$$

where.

(1) y_i = actual value

(2) $\dot{y}_i$ = predicted value

(3) $\overline{y}$ = mean of observed values

Higher R^2 values indicate better predictive performance.

These metrics provide a comprehensive evaluation of both the accuracy and robustness of the models.

3.8 Walk-forward validation of the random forest model

To ensure robust evaluation, walk-forward validation was implemented for the Random Forest model, which served as the primary machine-learning model for temporal performance assessment. The model was trained on progressively expanding datasets and tested on subsequent time periods. This process was repeated across ten sequential folds to assess predictive stability under changing market conditions.

4. Results and discussion

This section presents a comprehensive evaluation of the models developed for agricultural price prediction and provides an in-depth interpretation of the results. The analysis integrates exploratory data analysis, model comparison, temporal feature evaluation, and validation results to assess the effectiveness and reliability of the proposed approach.

4.1 Exploratory analysis of price behaviour

Before model development, exploratory data analysis was conducted to understand the structure, variability, and temporal characteristics of the dataset. Agricultural commodity prices exhibit clear fluctuations over time, reflecting seasonal patterns, market conditions, and external economic influences. To visualize (see Figure 5) the temporal dynamics of the dataset, a representative time series (Maize in Katsina) was analyzed.

The figure represents a subset of the available data due to missing observations in later years for this specific commodity-region pair. It demonstrates that agricultural prices are highly dynamic and exhibit non-stationary behaviour. The presence of upward and downward trends, as well as irregular fluctuations, indicates that price movements are influenced by both short-term and long-term factors. This observation highlights the necessity of incorporating temporal dependencies into predictive models.

The distribution of prices is positively skewed, with a concentration of lower values and a long tail of higher values. As shown in Figure 6, this indicates the presence of extreme values and potential outliers, which can significantly influence error metrics such as Mean Squared Error (MSE). The skewed nature of the data further reinforces the need for robust modelling approaches capable of handling variability.

4.2 Comparative model performance

To evaluate the effectiveness of different modelling approaches, Linear Regression, Random Forest, ARIMA, and LSTM models were implemented and compared using standard regression metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and the coefficient of determination (R^2). As depicted in Table 7, these metrics provide complementary perspectives on model performance, capturing both the magnitude of prediction errors and the proportion of variance explained by each model.

Before interpreting the results, it is important to acknowledge that the models were evaluated at different levels of aggregation. Linear Regression and Random Forest were trained and tested using the full multi-commodity dataset, while ARIMA and LSTM were applied to selected commodity-level series. Accordingly, the objective of this comparison is

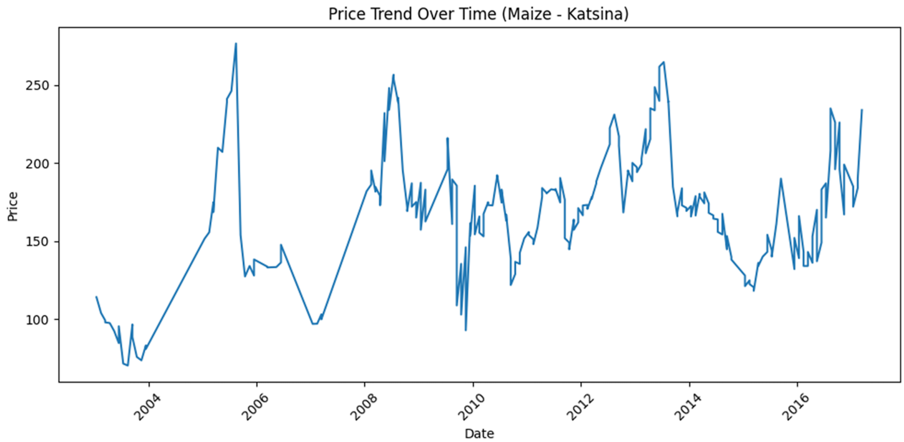

Figure 5. Temporal trend of maize prices in Katsina, illustrating seasonal fluctuations and long-term variability. (Source: Authors' own work)

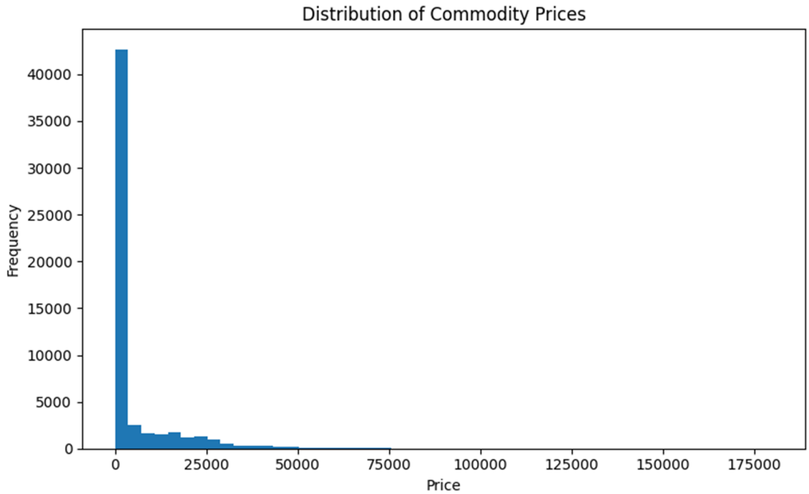

Figure 6. Distribution of agricultural commodity prices showing skewness and presence of extreme values. (Source: Authors' own work)

Table 7. Model performance across different evaluation scopes

Model	MAE	MSE	R^2
Linear Regression	3687.64	75,236,811.98	0.465
Random Forest	3474.73	70,883,256.96	0.497
ARIMA	26.34	1296.29	−0.259
LSTM (Maize)	15.36	364.62	0.640
LSTM (Rice)	11,893.29	153,813,836.48	−0.039

Note(s): Linear Regression and Random Forest were evaluated on the full multi-commodity dataset, whereas ARIMA and LSTM models were evaluated on selected commodity-level series. Accordingly, metric values should be interpreted within their respective evaluation contexts and are not directly comparable across all models

not to establish a strict one-to-one benchmark across identical datasets, but rather to illustrate how different modelling paradigms perform within representative agricultural forecasting scenarios. The reported metrics should therefore be interpreted within the scope of each model's evaluation setting.

The Linear Regression model achieved an R^2 value of 0.465, indicating moderate predictive capability. This suggests that nearly half of the variability in commodity prices can be explained through linear relationships between the selected features and the target variable. While this demonstrates that the chosen features contain useful predictive information, it also highlights the inherent limitations of linear models in capturing complex market behaviour. Agricultural price dynamics are influenced by non-linear interactions between multiple factors, including temporal dependencies, seasonal patterns, and external shocks. As a result, the Linear Regression model tends to oversimplify these relationships, leading to systematic errors, particularly during periods of rapid price fluctuations or structural changes in the market.

The Random Forest model demonstrated improved performance, achieving an R^2 of 0.497 and lower error metrics compared to Linear Regression. This improvement can be attributed to the model's ability to capture non-linear relationships and higher-order interactions between features through its ensemble of decision trees. By aggregating multiple decision paths, Random Forest is able to model more complex patterns in the data without relying on strict assumptions about linearity or data distribution. However, the relatively modest performance improvement suggests that increasing model complexity alone does not guarantee substantial gains in predictive accuracy. This indicates that the effectiveness of machine learning models in this context is strongly dependent on the quality and representativeness of the input features.

Furthermore, the Random Forest model exhibits a tendency to smooth predictions, particularly in the presence of extreme values. This behaviour arises from the averaging mechanism inherent in ensemble models, which can lead to underestimation of sharp price spikes and overestimation of lower values. Given the wide range and skewed distribution of commodity prices observed in the dataset, this limitation becomes particularly significant. It suggests that while Random Forest is effective in capturing general trends, it may struggle to accurately model rare but impactful events, such as sudden market disruptions or supply shocks.

In contrast, the ARIMA model performed poorly, yielding a negative R^2 value. This indicates that the model fails to capture meaningful patterns in the data and performs worse than a naive baseline predictor. The poor performance of ARIMA highlights the limitations of traditional statistical approaches when applied to complex, real-world datasets. ARIMA relies on assumptions of linearity, stationarity, and consistent temporal structure, which are rarely satisfied in agricultural price data. The presence of structural breaks, irregular fluctuations, and external influences violates these assumptions, leading to inaccurate forecasts. Additionally, the univariate nature of ARIMA prevents it from incorporating additional explanatory variables, further limiting its ability to model multifaceted price dynamics.

An important observation is the apparent contradiction in the ARIMA model results. While the model achieved relatively low MAE (26.34) and MSE (1296.29), it produced a negative R^2 value (-0.259). This paradox arises due to the scale and variance structure of the univariate time series used for ARIMA modelling. Specifically, the maize price series used for ARIMA exhibits a relatively narrow value range compared to the aggregated dataset used in machine learning models. Mean Absolute Error (MAE) and Mean Squared Error (MSE) measure absolute prediction differences and are therefore sensitive to the scale of the data. Since the ARIMA model was applied to a single commodity with lower variance, these metrics appear small. However, the coefficient of determination (R^2) evaluates how well the model explains the variance relative to a baseline model. In this case, the ARIMA model fails to capture the underlying variance structure of the time series, resulting in a negative R^2 value. This indicates that despite producing numerically small errors, the ARIMA model does not provide meaningful predictive power beyond a naive baseline. This finding highlights the limitation of relying solely on error-based metrics and reinforces the importance of using variance-based evaluation measures in time-series forecasting.

The LSTM model exhibited its strongest performance when applied to maize, achieving an R^2 value of 0.640 and substantially lower prediction errors than the other commodity-level evaluations. This result demonstrates the ability of deep learning models to capture sequential dependencies and long-term temporal patterns more effectively than traditional and machine learning approaches when relatively stable temporal structures are present. By maintaining internal memory states, LSTM networks can learn complex temporal relationships that are not explicitly encoded through manual feature engineering. This allows the model to adapt to gradual trends and recurring patterns in the data.

However, the performance of the LSTM model declined sharply when applied to a more volatile commodity (rice), resulting in a negative R^2 value. This contrast highlights a critical limitation of deep learning approaches: their sensitivity to data characteristics and variability. In highly volatile environments, where price movements are irregular and influenced by external factors not captured in the dataset, the model may struggle to identify consistent patterns. Additionally, deep learning models typically require large volumes of stable and representative data for effective training. In cases where the data is noisy or highly variable, the model may overfit to short-term fluctuations or fail to generalize to unseen patterns.

Another important observation is the inconsistency between evaluation metrics across models. For example, the ARIMA model exhibits relatively low MAE and MSE values despite having a negative R^2 score. This suggests that while the absolute prediction errors may appear small, the model fails to capture the overall variance and structure of the data. This reinforces the importance of using multiple evaluation metrics when assessing model performance, as reliance on a single metric may lead to misleading conclusions. The comparative analysis generally reveals that no single modelling approach consistently outperforms others across all scenarios. Instead, model performance is highly context-dependent and influenced by factors such as data distribution, temporal structure, and volatility. Machine learning models provide a balance between flexibility and interpretability, while deep learning models offer superior performance in stable environments but may struggle under high variability. Traditional statistical models, while theoretically robust, are less suitable for complex and non-linear datasets. These findings underscore the importance of adopting a context-aware modelling strategy that considers both the characteristics of the data and the limitations of each modelling approach. Rather than relying on a single model, a hybrid or adaptive framework that combines multiple techniques may provide a more robust solution for agricultural price forecasting in dynamic and uncertain environments. The findings are broadly consistent with previous agricultural forecasting studies that report stronger performance for machine learning and deep learning approaches relative to traditional statistical models when price behaviour is characterized by non-linearity and volatility (Sun *et al.*, 2023; Zhao *et al.*, 2024). However, the relatively modest improvement of Random Forest over Linear Regression in the present study suggests that feature quality and temporal representation may be as important as model complexity. This observation supports findings from Tran *et al.* (2023), who noted that forecasting performance is strongly influenced by data preparation, temporal feature design, and evaluation methodology.

4.3 Impact of temporal feature engineering

A key objective of this study was to evaluate the impact of temporal feature engineering on model performance. Initial experiments conducted using basic features, such as commodity type and location, resulted in relatively poor predictive performance, particularly for the Random Forest model. This indicates that static features alone are insufficient for capturing the dynamic nature of agricultural price movements. The introduction of temporal features, including multiple lag variables and rolling averages, led to a noticeable improvement in model performance. These features enable the model to incorporate historical price information, allowing it to learn short-term dependencies and underlying trends in the data. As a result, the model becomes more capable of capturing temporal patterns that would otherwise be overlooked. This improvement highlights the inherently time-dependent nature of agricultural prices, which are strongly influenced by historical patterns and temporal dependencies (Sun *et al.*, 2023). Price movements are not independent observations but are strongly influenced by previous values, seasonal cycles, and market behaviour. By incorporating temporal features, the model is better aligned with the data's structure, resulting in more accurate predictions. However, while temporal feature engineering enhances performance, it does not fully resolve the challenges associated with highly volatile price series. In cases where price fluctuations are irregular or influenced by external factors not captured in the dataset, the model's ability to generalize remains limited. This suggests that while temporal features are essential, they must be complemented with additional contextual variables to achieve higher predictive accuracy.

To quantify the contribution of engineered predictors, feature importance analysis was performed using the Random Forest model. The results, presented in Table 8 and Figure 7,

Table 8. Random forest feature importance rankings

Feature	Importance
rolling_mean_3	0.488
price_change	0.118
lag_1	0.117
lag_2	0.062
lag_3	0.052
year	0.042
month	0.039

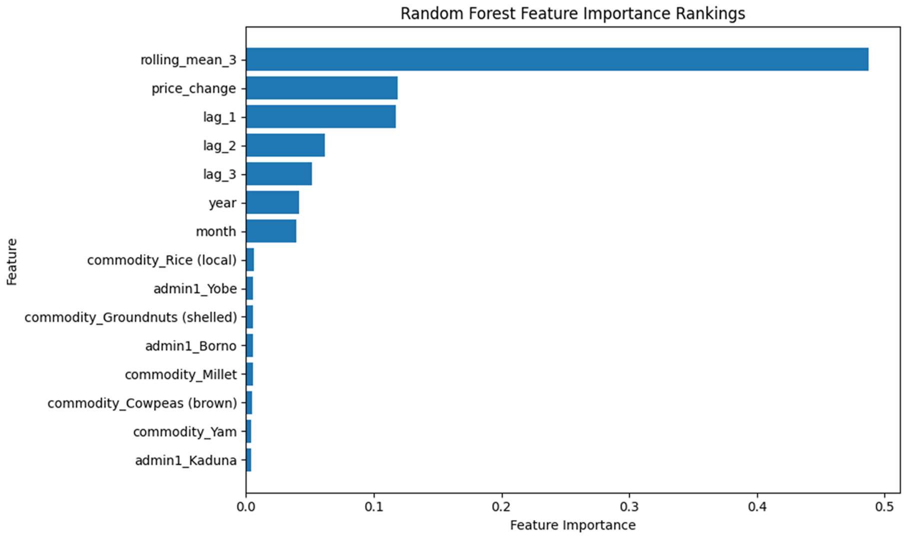

Figure 7. Random forest feature importance rankings showing the dominant contribution of temporal features to agricultural price prediction. (Source: Authors' own work)

indicate that temporal features were the dominant drivers of predictive performance. The rolling_mean_3 feature achieved the highest importance score (0.488), followed by price_change (0.118), lag_1 (0.117), lag_2 (0.062), and lag_3 (0.052). Collectively, these temporal features accounted for more than 80% of the total feature importance, substantially exceeding the contribution of categorical variables such as commodity and administrative region.

While one-hot encoding increased the number of input variables through the expansion of commodity and administrative-region categories, the feature-importance analysis indicates that temporal variables contributed substantially more to predictive performance than categorical indicators. This suggests that historical price dynamics were more informative than location-specific or commodity-specific identifiers within the modelling framework adopted in this study.

These findings provide strong empirical support for the central premise of this study that temporal feature engineering significantly improves agricultural price forecasting performance. The dominance of lag-based and rolling-window features demonstrates that historical price behaviour contains valuable predictive information and highlights the importance of incorporating temporal dependencies when modelling agricultural commodity markets.

4.4 Walk-forward validation

The walk-forward validation results reported in this section correspond exclusively to the Random Forest model. To assess the robustness and reliability of the models, walk-forward validation was implemented. Unlike a single train-test split, this approach evaluates model performance across multiple sequential time periods, providing a more realistic representation of real-world forecasting scenarios. The results (shown in Table 9) reveal significant variability in model performance across different time periods.

While some folds exhibit relatively high predictive accuracy, others show substantially lower or even negative R^2 values. This variability reflects the changing nature of agricultural markets, where price dynamics can shift due to seasonal effects, supply disruptions, and broader economic conditions. The average R^2 value of approximately 0.33 is notably lower than the value obtained from a single train-test split. This observation aligns with prior studies showing that single train-test evaluations can produce overly optimistic performance estimates in time-series forecasting (Makridakis *et al.*, 2018). In contrast, walk-forward validation provides a more robust and realistic assessment by accounting for fluctuations over time. These findings highlight an important limitation of predictive models in dynamic environments. Even when a model performs well during certain periods, its performance may degrade under different market conditions. This highlights the importance of using evaluation methods that reflect real-world deployment scenarios. The results generally demonstrate that while machine learning models can achieve reasonable predictive performance, their reliability is influenced by temporal variability. This reinforces the need for continuous model updating and the incorporation of additional features that capture external influences on price dynamics. Lower predictive performance was observed during certain evaluation periods, particularly those corresponding to major economic and market disruptions in Nigeria. For example, the weaker validation folds covering approximately 2018–2020 and 2022–2025 coincide with periods characterized by food inflation, border-trade restrictions, exchange-rate instability, insecurity-related supply disruptions, and fuel subsidy reforms. These events likely altered historical price relationships and reduced forecasting reliability, highlighting the sensitivity of agricultural price models to changing economic conditions. Each fold used an expanding training window and a fixed-length subsequent testing window, preserving temporal ordering throughout the evaluation process.

To improve interpretability, the walk-forward validation results are visualized in Figure 8. The figure illustrates the variation in R^2 scores across folds, highlighting the instability of model performance over time.

Table 9. Random forest walk-forward validation results

Fold	R^2
1	0.65
2	0.79
3	0.11
4	0.28
5	0.44
6	0.48
7	0.48
8	0.09
9	0.20
10	−0.24
Average	0.33

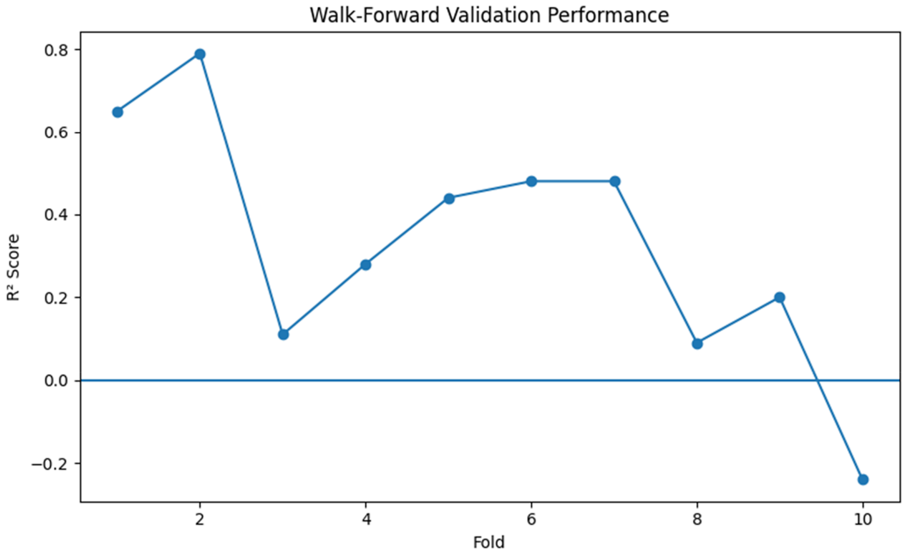

Figure 8. Walk-forward validation R^2 scores across sequential folds showing performance variability over time. (Source: Authors' own work)

4.5 Actual Vs predicted prices

To further evaluate model performance, a comparison between actual and predicted values was conducted. The comparison between actual and predicted values reveals that the Random Forest model can capture general trends in commodity prices but significantly underestimates extreme values and price spikes. While the model performs reasonably well for moderate price ranges, it struggles to accurately predict high-value observations. This limitation can be attributed to the high variability and wide range of prices present in the dataset, as well as the aggregation of multiple commodities with different price scales into a single model. The model exhibits reduced sensitivity to extreme observations, as evidenced by its consistent underestimation of high-value price spikes in the prediction results.

Furthermore, the results highlight the challenges of modelling heterogeneous datasets that combine multiple commodities with distinct price behaviours. This suggests that more specialized models, trained on individual commodities or incorporating additional contextual features, may be required to improve predictive performance. Figure 9 presents the comparison between actual and predicted commodity prices using the Random Forest model.

For clarity, extreme values were filtered to improve visualization. The results show that the model is able to capture the general trend of price movements, particularly in relatively stable periods. However, significant deviations are observed during periods of high volatility, where the model either overestimates or underestimates price values. This behaviour indicates that while the Random Forest model effectively learns underlying patterns in the data, it struggles to generalize in the presence of sharp fluctuations and extreme price variations. The inconsistencies observed in later periods suggest that the model is sensitive to changes in market conditions and may not fully capture the complexity of agricultural price dynamics.

4.6 Residual analysis of random forest model

To further evaluate model performance, residual analysis was conducted on the Random Forest model. Residuals represent the difference between actual and predicted values and

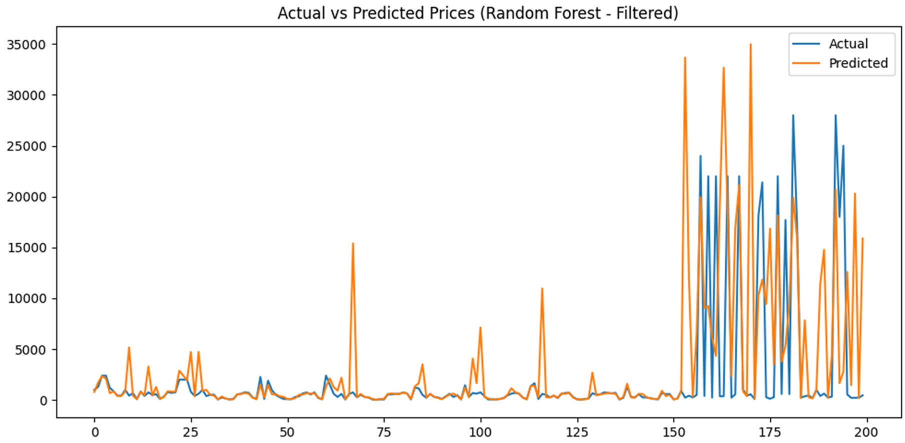

Figure 9. Actual Vs predicted prices. (Source: Authors' own work)

provide insights into systematic prediction errors. Figure 10 illustrates the residual distribution of the Random Forest model. Ideally, residuals should be randomly distributed around zero, indicating that the model captures all underlying patterns. However, the results reveal that residuals are not evenly distributed. Large positive and negative residuals are observed, particularly during periods of extreme price fluctuations. This pattern confirms that the model struggles to accurately predict high-value observations and price spikes. The presence of structured residual patterns suggests that certain temporal dynamics and external factors influencing price movements are not fully captured by the model. This supports earlier findings that ensemble models tend to smooth predictions and underrepresent extreme values. The residual analysis provides strong empirical evidence that the model's limitations are not random but systematic, particularly in volatile market conditions. This further reinforces the need for incorporating additional explanatory variables and developing commodity-specific models.

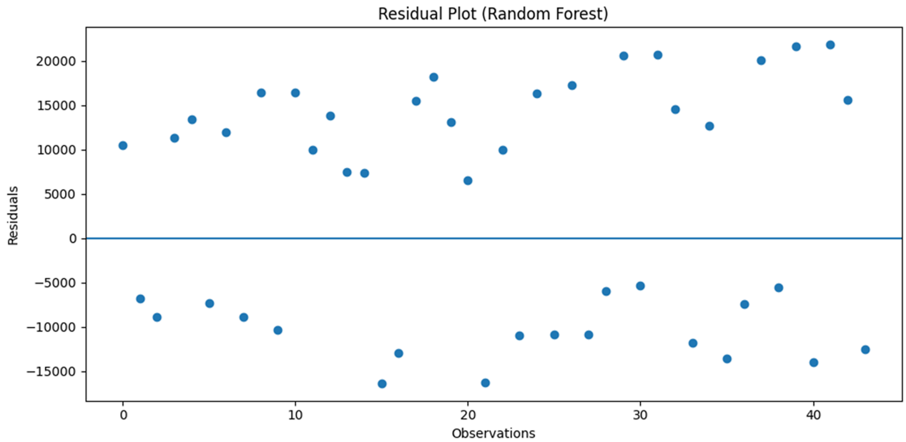

Figure 10. Residual plot for random forest model showing systematic prediction errors and large deviations during price spikes. (Source: Authors' own work)

4.7 Commodity-level model performance

To further investigate model performance variability, results were analyzed at the commodity level. Agricultural commodities exhibit different price dynamics due to variations in demand, supply conditions, and market structure. Table 10 presents model performance across selected commodities. The results reveal that predictive accuracy varies significantly depending on the commodity being analyzed. For relatively stable commodities such as maize, models achieve higher predictive performance. In contrast, highly volatile commodities such as rice exhibit significantly lower performance, particularly for deep learning models. This variation highlights the importance of commodity-specific modelling approaches. A single unified model may not adequately capture the heterogeneous behaviour of different commodities, leading to reduced predictive accuracy. These findings support the need for more specialized modelling frameworks tailored to individual commodity characteristics. To provide a broader assessment of model robustness, seven commodities representing different levels of price volatility and data availability were selected for additional LSTM evaluation, including relatively stable commodities (e.g. maize and sorghum) and highly volatile commodities (e.g. rice, yam, and cowpeas).

Commodities with fewer observations may exhibit greater estimation uncertainty, which could partially contribute to the lower predictive performance observed for some commodity-level models. The commodity-level LSTM results presented in Table 10 were obtained from the expanded evaluation conducted and therefore differ slightly from the initial commodity-level experiments reported in Table 7.

Commodity-level evaluation revealed substantial variability in forecasting performance across agricultural products. The LSTM model achieved its strongest performance for maize ($R^2 = 0.640$) and sorghum ($R^2 = 0.468$), indicating that commodities with relatively stable temporal patterns can be forecast effectively using sequence-based deep learning approaches. Groundnuts (shelled) also produced a positive R^2 value, although predictive performance remained moderate. In contrast, rice (imported), millet, yam, and cowpeas (white) yielded negative R^2 values despite extended training and the implementation of Early Stopping. These results indicate that highly volatile commodities present greater forecasting challenges and may be influenced by external factors not captured in the historical price series alone. The variation in predictive performance demonstrates that commodity-specific characteristics play a critical role in determining model effectiveness.

The observed variation across commodities can be partly explained by differences in volatility, sample size, and temporal stability. Maize and sorghum exhibited relatively stable price patterns and sufficient observations for model training, allowing the LSTM architecture to learn meaningful temporal relationships. In contrast, commodities such as rice (imported), millet, yam, and cowpeas (white) displayed greater price variability and were likely influenced by external market forces not represented in the dataset. Under such conditions, historical price information alone may be insufficient to capture future movements accurately.

Table 10. Commodity-level LSTM performance across agricultural commodities with different volatility profiles

Commodity	Observations	MAE	MSE	R^2
Maize	235	15.36	364.62	0.640
Sorghum	140	13.83	415.73	0.468
Groundnuts (shelled)	91	19,252.02	506,076,700	0.148
Rice (imported)	223	12,177.92	162,869,500	−0.100
Cowpeas (white)	125	21,915.47	512,876,000	−0.158
Yam	87	21,156.72	539,573,300	−0.214
Millet	318	8612.67	107,818,300	−0.228

Similar patterns have been reported in previous agricultural forecasting studies. Gu *et al.* (2022) and Ray *et al.* (2023) found that deep learning models generally perform well when commodity prices exhibit consistent temporal structures but may struggle under highly volatile market conditions. The present findings support this observation, demonstrating that LSTM performance varies considerably across commodities, and that forecasting accuracy is strongly influenced by commodity-specific market characteristics.

The findings further suggest that the success of deep learning approaches in agricultural price forecasting is highly context dependent. While LSTM models are capable of capturing temporal dependencies in relatively stable commodity markets, their effectiveness decreases when price dynamics become highly irregular or are influenced by structural market disruptions. Consequently, commodity-level evaluation is essential for understanding forecasting reliability and identifying scenarios where additional explanatory variables or hybrid modelling approaches may be required. To address concerns regarding potential undertraining, the Rice (Imported) LSTM model was retrained using a maximum of 100 epochs together with Early Stopping based on validation loss. The resulting training history is presented in Figure 11.

Figure 11 illustrates the training and validation loss curves for the Rice (Imported) LSTM model. Both loss functions decreased progressively throughout training, indicating stable optimization and successful convergence of the learning process. The implementation of extended training and Early Stopping resulted in a measurable improvement in predictive performance relative to the initial 20-epoch configuration, although forecasting accuracy remained lower than that observed for more stable commodities such as maize and sorghum. This finding suggests that the weaker performance is attributable not only to training duration but also to the higher volatility and structural complexity of the rice price series.

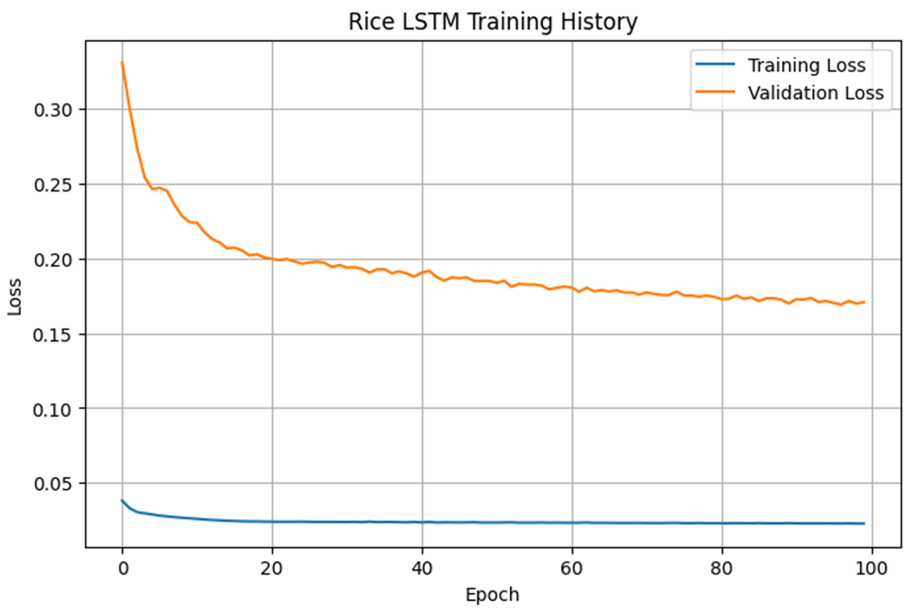

Figure 11. Training and validation loss curves for the rice (imported) lstm model showing stable convergence under extended training with early stopping. (Source: Authors' own work)

5. Conclusion

5.1 Main contribution

This study developed a time-aware forecasting framework for agricultural commodity price prediction in Nigeria using the World Food Programme (WFP) food price dataset covering the period 2002–2025. The framework integrated temporal feature engineering, chronological data splitting, and walk-forward validation to evaluate the performance of Linear Regression, Random Forest, ARIMA, and LSTM models. The results demonstrate that temporal feature engineering contributes substantially to forecasting performance and that model effectiveness varies considerably across commodities and market conditions. The study further highlights the importance of commodity-specific evaluation and realistic time-aware validation strategies for agricultural forecasting applications.

The results also suggest that forecasting systems should be deployed as decision-support tools rather than standalone decision-making mechanisms. Because agricultural prices are influenced by weather variability, transportation constraints, exchange-rate fluctuations, insecurity, and policy changes, predictive models should be combined with domain expertise and complementary market information. Such an integrated approach would improve the practical usefulness of forecasting systems for government agencies, development organizations, and agricultural market stakeholders.

An important finding of this study is the temporal variability revealed through walk-forward validation. While the Random Forest model achieved reasonable overall predictive performance, forecasting accuracy varied substantially across different testing periods. Particularly low performance was observed during the 2018–2019 and 2022–2025 evaluation windows, periods that coincided with major disruptions in Nigerian agricultural markets, including border-trade restrictions, inflationary pressures, currency depreciation, and the removal of fuel subsidies. These findings suggest that agricultural price forecasting models are sensitive to structural market changes and economic shocks that alter historical price relationships. Consequently, predictive systems should be periodically retrained and continuously updated to remain effective under changing market conditions.

5.2 Practical and policy significance

From a broader perspective, the findings highlight opportunities for strengthening agricultural market intelligence systems in Nigeria. Forecasting tools based on temporal feature engineering could support organizations such as the Federal Ministry of Agriculture and Food Security, the National Bureau of Statistics (NBS), and the National Agricultural Extension and Research Liaison Services (NAERLS) by providing early warning signals of abnormal price movements. Improved forecasting could assist policymakers in monitoring food inflation, planning targeted interventions, improving market transparency, and supporting evidence-based food security programs in major agricultural markets across the country. The practical applications of these forecasting models extend to multiple stakeholder groups. For farmers, forecast outputs could be integrated into agricultural extension services and mobile advisory platforms to provide advance information on potential price movements, supporting planting, harvesting, and storage decisions. For traders and market intermediaries, short-term price forecasts could improve inventory management, procurement planning, and market timing strategies. Market information systems operated by government agencies and development organizations could incorporate forecasting outputs into periodic price bulletins and early-warning dashboards to identify unusual price fluctuations before they become severe. In addition, policymakers could use forecast-based indicators to support food inflation monitoring, targeted subsidy programs, strategic grain reserve planning, and emergency response interventions during periods of market disruption. In the Nigerian context, such forecasting tools could complement existing market monitoring activities conducted by the Federal Ministry of Agriculture and Food Security, the National Bureau of Statistics (NBS), and NAERLS, thereby improving the timeliness and effectiveness of agricultural market interventions.

From a practical perspective, the findings suggest that the deployment of predictive models in real-world agricultural systems should be approached with caution. Decision-support tools based on these models must account for uncertainty, variability, and changing market conditions. For policymakers and market stakeholders, this highlights the importance of combining data-driven models with domain knowledge and external contextual information to improve decision-making.

5.3 Interpretation of walk-forward validation findings

The walk-forward validation results further demonstrate that forecasting performance is not constant over time. While the Random Forest model achieved strong predictive performance during several evaluation periods, weaker performance was observed in folds corresponding to periods of heightened economic uncertainty and market disruption. In Nigeria, these periods coincided with events such as border-trade restrictions, rising inflation, exchange-rate depreciation, insecurity-related disruptions to agricultural supply chains, and fuel subsidy reforms. These findings suggest that historical price patterns alone may be insufficient during periods of structural change, and that forecasting systems should be continuously updated to adapt to evolving market conditions.

5.4 Limitations

Although the proposed framework improved agricultural price forecasting performance, several limitations remain. The models rely primarily on historical price information and engineered temporal features, while potentially important explanatory variables such as weather conditions, transportation costs, inflation, exchange-rate fluctuations, market accessibility, and government policy interventions were not included. These factors can significantly influence agricultural price dynamics and may explain a portion of the unexplained variance observed in the results. Consequently, the reported forecasting performance should be interpreted within the context of the available data. In addition, the deep learning experiments were limited to selected commodity-level series due to computational considerations and data heterogeneity. Future studies may benefit from evaluating a broader range of commodities and advanced architectures such as transformer-based and hybrid forecasting models.

5.5 Future research directions

Future research should focus on incorporating exogenous variables such as climate data, macroeconomic indicators, and policy interventions, as well as exploring hybrid and adaptive modelling frameworks that can respond to changing market dynamics in real time. While machine learning, deep learning, and traditional time-series approaches each demonstrated strengths within their respective evaluation settings, model effectiveness varied considerably across commodities and market conditions. The results suggest that forecasting performance is highly dependent on data characteristics, temporal stability, and commodity-specific behaviour rather than the choice of modelling technique alone. Model effectiveness varies significantly depending on the stability, variability, and structural characteristics of the commodity being analyzed. An important implication of this study is that predictive performance in agricultural markets cannot be evaluated in isolation from data characteristics. Models that perform well under stable conditions may fail under volatile market dynamics, highlighting the need for context-aware modelling strategies. Furthermore, the results emphasized that methodological choices, particularly feature design and validation approach, play a critical role in determining model reliability. The use of walk-forward validation demonstrates that conventional single-split evaluation methods may provide overly optimistic performance estimates and fail to capture temporal instability. The findings generally demonstrate that effective agricultural commodity price forecasting depends not only on model selection but also on robust temporal feature engineering, realistic validation strategies, and an understanding of commodity-specific market dynamics. These insights provide a foundation for developing more reliable forecasting systems capable of supporting food security, market planning, and agricultural policy decision-making in Nigeria and other emerging economies.

Ethics statement

This study used publicly available secondary data obtained from the World Food Programme (WFP) Vulnerability Analysis and Mapping (VAM) Food Prices Database. The dataset contains aggregated agricultural commodity price information collected across markets and administrative regions in Nigeria and does not contain personal, confidential, or human participant data. Consequently, ethical approval was not required for this study.

Primary data source

World Food Programme (WFP) Vulnerability Analysis and Mapping (VAM) Food Prices Database.

Available at: World Food Programme.

For reproducibility, a mirrored version of the Nigeria food price dataset covering the period 2002–2025 was accessed through the Kaggle platform and used during the analysis.

Implementation availability

The source code and implementation details used in this study are publicly available at Github.

References

Aker, J.C. (2010), "Information from markets near and far: mobile phones and agricultural markets in Niger", *American Economic Journal: Applied Economics*, Vol. 2 No. 3, pp. 46-59, doi: 10.1257/app.2.3.46.

Box, G.E., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M. (2015), *Time Series Analysis: Forecasting and Control*, John Wiley & Sons, 108201.

Breiman, L. (2001), "Random forests", *Machine Learning*, Vol. 45 No. 1, pp. 5-32, doi: 10.1023/a:1010933404324.

Chen, Z., Goh, H.S., Sin, K.L., Lim, K., Chung, N.K.H. and Liew, X.Y. (2021), "Automated agriculture commodity price prediction system with machine learning techniques", arXiv preprint arXiv:2106.12747.

Gilbert, C.L. and Morgan, C.W. (2010), "Food price volatility", *Philosophical Transactions of the Royal Society B: Biological Sciences*, Vol. 365, 1554, pp. 3023-3034, doi: 10.1098/rstb.2010.0139.

Gu, Y., Dong, J., Yin, H., Zheng, R., Piao, X. and Yoo, S. (2022), "Forecasting agricultural commodity prices using dual-input attention LSTM", *Agriculture*, Vol. 12 No. 2, p. 256, doi: 10.3390/agriculture12020256.

Hochreiter, S. and Schmidhuber, J. (1997), "Long short-term memory", *Neural Computation*, Vol. 9 No. 8, pp. 1735-1780.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013), *An Introduction to Statistical Learning: With Applications in R*, Vol. 103, Springer, New York.

Makridakis, S., Spiliotis, E. and Assimakopoulos, V. (2018), "Statistical and Machine Learning forecasting methods: concerns and ways forward", *PLoS One*, Vol. 13 No. 3, e0194889, doi: 10.1371/journal.pone.0194889.

Obayelu, A.E., Edewor, S.E., Ogbe, A.O. and Oyedepo, E.O. (2024), "Assessment of agricultural trade flow and food security status: evidence from Nigeria", *Agriculturae Conspectus Scientificus*, Vol. 89 No. 2, pp. 175-186.

Ray, S., Lama, A., Mishra, P., Biswas, T., Das, S.S. and Gurung, B. (2023), "An ARIMA-LSTM model for predicting volatile agricultural price series with random forest technique", *Applied Soft Computing*, Vol. 149, 110939, doi: 10.1016/j.asoc.2023.110939.

Sharma, A., Jain, A., Gupta, P. and Chowdary, V. (2020), "Machine learning applications for precision agriculture: a comprehensive review", *IEEE Access*, Vol. 9, pp. 4843-4873, doi: 10.1109/access.2020.3048415.

Sun, F., Meng, X., Zhang, Y., Wang, Y., Jiang, H. and Liu, P. (2023), "Agricultural product price forecasting methods: a review", *Agriculture*, Vol. 13 No. 9, p. 1671, doi: 10.3390/agriculture13091671.

Tran, N.Q., Felipe, A., Ngoc, T.N., Huynh, T., Tran, Q., Tang, A. and Nguyen, T. (2023), "Predicting agricultural commodities prices with machine learning: a review of current research", arXiv preprint arXiv:2310.18646.

Wang, L., Feng, J., Sui, X., Chu, X. and Mu, W. (2020), "Agricultural product price forecasting methods: research advances and trend", *British Food Journal*, Vol. 122 No. 7, pp. 2121-2138, doi: 10.1108/bfj-09-2019-0683.

Zhao, Y., Liu, J. and Wang, H. (2024), "A comparative study of machine learning algorithms for agricultural market price prediction", *Expert Systems with Applications*, Vol. 238, 121912.

Corresponding author
Rania Shibl can be contacted at: rania.shibl@scu.edu.au