Cureus Journal of Computer Science
https://doi.org/10.7759/s44389-026-00093-z

Original Article

Predictive Price-Signal Forecasting Using Machine Learning to Support Agricultural Supply Planning in Nigeria

Sunday O. Oladipo [1] , Afolashade Kuyoro [1] , Victor D. Ibironke [1, ✉] , Osedebame E. Itamah [1] , Nelson C. Isicheli [1]

1. *Computer Science, Babcock University, Ilishan-Remo, NGA*

Received: April 17, 2026 | Review began: May 05, 2026 | Review ended: May 18, 2026 | Published: June 02, 2026

Abstract

An estimated 38 million tonnes of agricultural produce in Nigeria are lost to waste annually, largely due to a lack of reliable market intelligence for farmers before harvest. This paper presents a machine learning system designed to address this information gap by forecasting commodity price signals, used as a proxy for market activity, across 39 agricultural commodities. Using price and climate data from the World Food Programme, we evaluated seven algorithms and three conventional baselines through a structured tournament. Our results show that Random Forest outperformed other models for 19 of the 39 commodities (48.7%). Feature importance analysis reveals that recent historical price trends are significantly more predictive of future market signals than climate or seasonal variables. The system was deployed as a lightweight Flask web application that provides a market activity index to assist in harvest timing and distribution planning. We conclude with a discussion of the ethical implications of using price proxies for decision support and the importance of addressing representational bias in agricultural data.

Categories: Ensemble Learning, AI applications, AI/ML-based decision support systems

Keywords: agricultural supply chain, food waste, machine learning, random forest, xgboost, svr, lstm, prophet, predictive price signaling, market intelligence

Introduction

Nigeria's agricultural sector, which employs approximately 70% of the workforce and contributes up to 25% of the national GDP, faces a critical challenge: an annual loss of 38 million tonnes of food [1], and tomatoes account for about 40% of that amount [2]. These losses represent a significant economic drain on smallholder farmers, many of whom earn less than $2 per day [3]. A primary driver of this waste is the "information vacuum" in which farmers operate; harvest and pricing decisions are frequently based on historical memory or anecdotal evidence rather than empirical market signals [4].

While previous studies have demonstrated the potential for machine learning (ML) in reducing waste in controlled environments, such as European catering services [5], applying these tools to the volatile and data-sparse markets of sub-Saharan Africa requires a nuanced approach. This study does not claim to measure direct food-waste reduction. Instead, it evaluates the accuracy of forecasting price-derived market signals as a necessary precondition for informed supply planning. By providing farmers with a "Price-Signal Index," we aim to offer a tool that helps optimize market entry timing, thereby indirectly mitigating the gluts that lead to post-harvest spoilage.

This paper reports on a complete forecasting pipeline, including a two-round model selection tournament that compares state-of-the-art ML algorithms against traditional statistical baselines. We also address critical methodological concerns regarding time-series validation and the ethical risks associated with deploying predictive tools in sensitive agricultural contexts.

Related works

Traditional forecasting methods, such as AutoRegressive Integrated Moving Average (ARIMA) and Exponential Smoothing, have long been staples in agricultural forecasting. However, their performance can be severely hampered by the high volatility and irregular seasonal patterns characteristic of African commodity markets. For instance, Opara et al. [5] reported Mean Absolute Percentage Error (MAPE) values exceeding 265% when using ARIMA for stone fruit waste forecasting, highlighting its limitations in capturing irregular variations. Similarly, Rathod et al. [6] observed that ARIMA forecasts for rice prices in India during the COVID-19 lockdowns sharply declined, while ML models demonstrated comparative resilience.

The emergence of ML has offered promising alternatives. Random Forest has shown strong performance in agricultural price forecasting, with Kasar et al. [7] reporting a MAPE of 0.68% for ladyfinger price forecasting, significantly outperforming ARIMA on the same dataset. Abubakar et al. [8] found that XGBoost surpassed Long Short-Term Memory (LSTM) in forecasting Nigerian food prices specifically, suggesting that gradient boosting methods are better suited to the irregularities of Nigerian market data. However, Rodrigues et al. [9] noted that LSTM can outperform ensemble methods when substantial historical data is available.

Beyond forecasting, ML and IoT technologies have been explored for direct waste reduction. Olawale et al. [10] combined ML with IoT sensors to achieve up to 60% reductions in post-harvest losses. The feasibility of real-time monitoring, even with limited infrastructure, is supported by studies like those by Haque et al. [11], who demonstrated high accuracy in food freshness classification using low-power sensors. Despite these technological advancements, adoption barriers persist, including cost, limited digital literacy, and unreliable infrastructure, particularly in Nigeria [12]. Furthermore, the challenge of integrating heterogeneous data systems makes assembling the necessary inputs for ML models costly and error-prone [13].

A critical ethical consideration, as raised by Sunday [14], is the potential for models trained on national or urban market data to perform poorly for smallholder farmers in local markets, leading to representational bias. This underscores the importance of local data collection and careful interpretation of model outputs, particularly when guiding policy or resource allocation.

Materials And Methods

System architecture

The system is structured around three sequential phases: data foundation, predictive modeling, and deployment. Figure *1* shows the full pipeline.

How to cite this article:
Oladipo S O, Kuyoro A, Ibironke V D, et al. (June 02, 2026) Predictive Price-Signal Forecasting Using Machine Learning to Support Agricultural Supply Planning in Nigeria. Cureus J Comput Sci 3 : es44389-026-00093-z. DOI https://doi.org/10.7759/s44389-026-00093-z

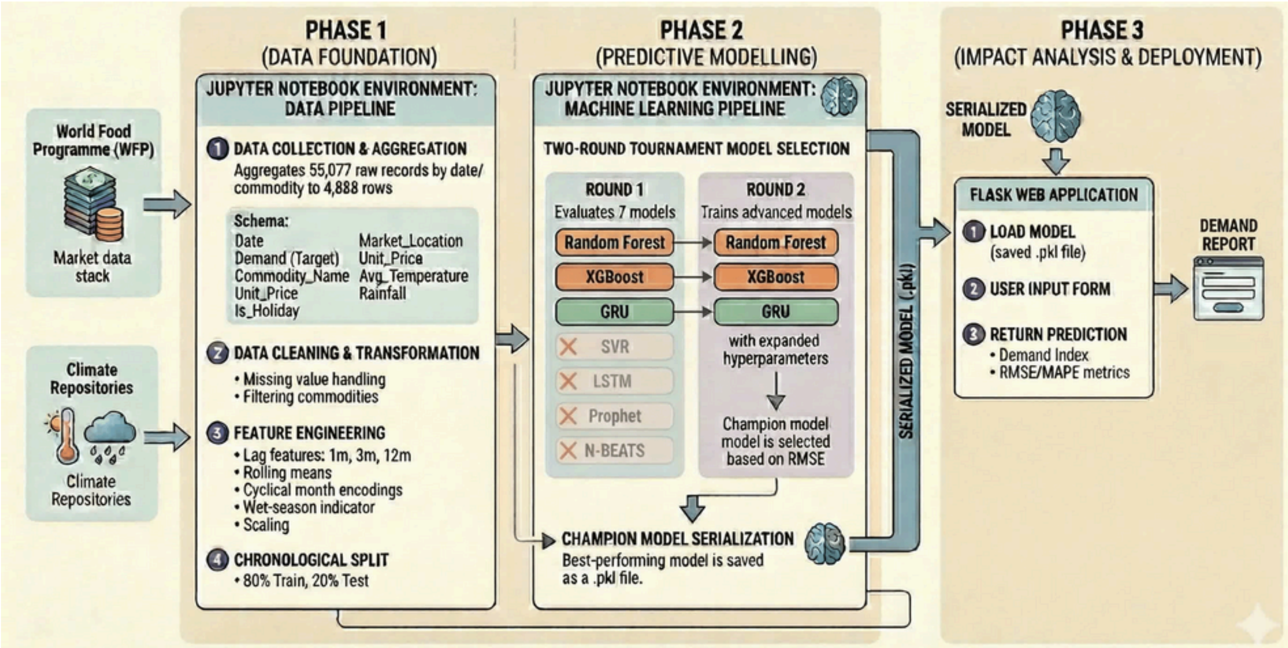

FIGURE 1: System architecture diagram showing the three-phase pipeline: data foundation (WFP price data and climate data merged and engineered), predictive modeling (two-round tournament across seven algorithms), and deployment (Flask web application returning demand indices). Generated with Gemini Pro

XGBoost, Extreme Gradient Boosting; GRU, Gated Recurrent Unit; SVR, Support Vector Regression; LSTM, Long Short-Term Memory; RMSE, Root Mean Square Error

Phase 1: Data Foundation

The primary data source was the World Food Programme VAM Food Security portal, supplemented by climate data from public weather APIs. The raw dataset contained 55,077 records across 42 commodities and multiple market locations. We aggregated these by date and commodity, computing mean values across all active markets per period. This process resulted in a consolidated dataset of 4,888 unique commodity-month observations used for modeling. The final schema is presented in Table 1.

It is critical to distinguish between market demand and price signals. Due to the absence of granular transaction-volume data, we used min-max normalized market prices (0-100 index) as a proxy for market activity. While price and demand are often correlated, they are not equivalent, particularly in markets with inelastic supply or government intervention. Therefore, the target variable is defined as a "Market Activity Index" derived from price.

Table 1 shows the final schema.

Variable Name	Description	Type
Date	The date of the data record.	Date time
Price (Target)	Monthly mean commodity price (Nigerian Naira), min-max normalized to a 0-100 index, used as a proxy for market activity.	Numerical
Commodity_Name	The name of the crop (e.g., tomatoes, yams).	Categorical
Market_Location	The market or region for the data record.	Categorical
Unit_Price	The average price of the commodity on that day/week.	Numerical
Avg_Temperature	Average temperature for the region.	Numerical
Rainfall	Rainfall measurement for the region.	Numerical
Is_Holiday	A binary indicator for public holidays.	Binary

TABLE 1: Schema for the final modeling dataset

Missing temperature and rainfall entries (17,703) were handled using forward-fill within each commodity group, with backward-fill for leading gaps. Commodities with fewer than 36 records (approximately 3 years of monthly data) were dropped, leaving 39 for modeling. Feature engineering focused on time structure, including lag features (1, 3, and 12 months), rolling means (3-, 6-, and 12-month windows), and sine/cosine encodings for the month to capture cyclicality.

Phase 2: Predictive Modeling and Validation

To ensure methodological rigor and provide more reliable performance estimates, we implemented Time-Series Cross-Validation (Rolling-Origin). This approach involves training the model on an expanding window of historical data and testing on the next unseen time step, iteratively. This process is repeated across multiple time windows, and performance metrics are averaged, mitigating the risk of optimistic estimates from a single chronological split, especially given Nigeria's high food price inflation period of 2024-2025.

We evaluated 10 algorithms in a two-round tournament. Round 1 included seven ML models (Random Forest, XGBoost, support vector regression [SVR], LSTM, gated recurrent unit [GRU], Prophet, and N-BEATS) and three conventional forecasting baselines:

1. Naïve Forecast: The forecast for the next month is the value of the current month.

2. Seasonal Naïve: The forecast for the next month is the value from the same month in the previous year.

3. ARIMA: A standard statistical time-series model.

All models were initially evaluated using default parameters. The top three performers, Random Forest, XGBoost, and GRU, advanced to Round 2 for hyperparameter optimization. The champion model for each commodity was selected based on the lowest root mean square error (RMSE) on the test set.

Phase 3: Deployment

The serialized champion models and preprocessing scalers were integrated into a lightweight Flask web application. The user interface allows selection of a commodity and input of environmental and temporal parameters (month, year, rainfall, temperature, holiday status). A critical technical detail is the backend's handling of feature engineering: while the user only inputs these parameters, the Flask backend maintains a database of the most recent historical price data. When a prediction is requested, the application automatically retrieves the necessary historical records to compute the required lag (e.g., Price_Lag1) and rolling mean (e.g., Price_RollingMean3) features before passing the complete feature vector to the serialized model for inference. The system then returns a predicted market activity index on a 0-100 scale.

Results And Discussion

Experiments

The project was implemented in Python using a Jupyter Notebook environment. It was run on a system hardware setup with an Intel Core i7 processor and 32GB RAM. The system architecture used Flask for the backend, while the user interface was built using HTML, CSS, and JavaScript. For the sake of reproducibility, all trained models and min-max scaling parameters were serialized into .pkl format, while we ensured that each commodity had at least 36 samples (three years of data) to provide sufficient depth for seasonal analysis.

The dataset, which starts on the 15th of January, 2002, and ends on the 15th of November, 2025, was split into two, ensuring the order was preserved to prevent data leakage during the training and validation phases. The numerical features were normalized to a (0, 1) range before training. This was done in two stages, beginning with a baseline evaluation of seven algorithms (Random Forest, XGBoost, SVR, LSTM, GRU, Prophets, and N-BEATS) and ending with the top-performing models (Random Forest, XGBoost, and GRU) being refined with specific hyperparameters.

Results

Our aggregated dataset comprises 4,888 unique commodity-month observations. Figure 2 illustrates the distribution of records across the top 15 most-represented commodities. Millet, Rice (imported), and Rice (local) were the most frequently observed, each contributing over 2,500 individual price records to the raw dataset before aggregation. This highlights the economic significance of these staples within Nigeria's food system and the heterogeneous nature of the data environment across commodities.

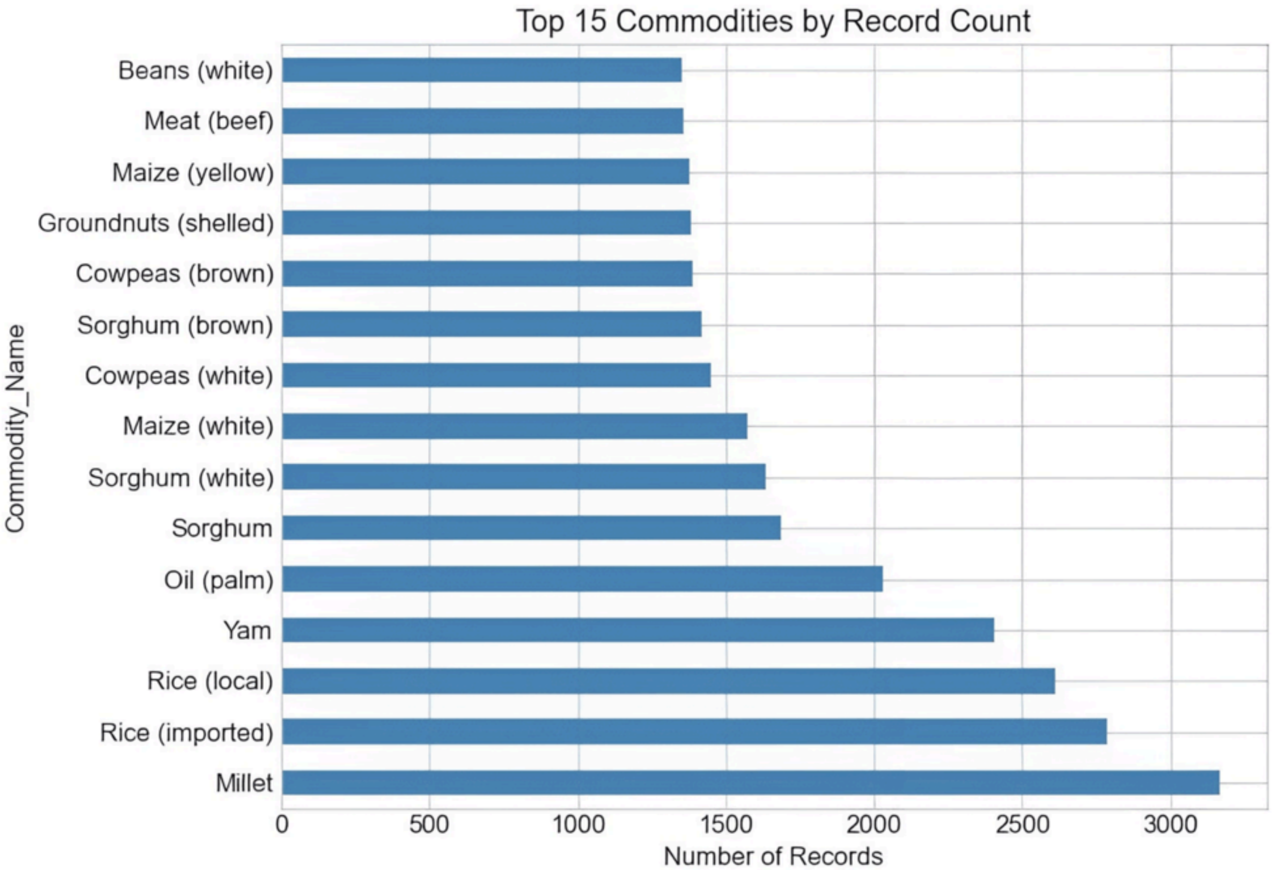

FIGURE 2: Top 15 agricultural commodities by record count in the World Food Programme dataset. Millet, imported rice, and local rice account for the highest representation

Table *2* shows the mean RMSE results for Round 1 across all 39 commodities. The top-performing ML models demonstrated superior accuracy compared to the conventional baselines. Random Forest, XGBoost, and GRU advanced to Round 2.

How to cite this article:
Oladipo S O, Kuyoro A, Ibironke V D, et al. (June 02, 2026) Predictive Price-Signal Forecasting Using Machine Learning to Support Agricultural Supply Planning in Nigeria. Cureus J Comput Sci 3 : es44389-026-00093-z. DOI https://doi.org/10.7759/s44389-026-00093-z

Model	Mean RMSE	Status
Random Forest	11.6036	Advanced
XGBoost	11.7539	Advanced
GRU	12.4571	Advanced
N-BEATS	13.0573	Eliminated
LSTM	13.0870	Eliminated
Prophets	17.4832	Eliminated
SVR	18.6570	Eliminated

TABLE 2: Round 1 mean RMSE across 39 commodities

XGBoost, Extreme Gradient Boosting; GRU, Gated Recurrent Unit; SVR, Support Vector Regression; N-BEATS, Neural Basis Expansion Analysis for Time Series; LSTM, Long Short-Term Memory; RMSE, Root Mean Square Error

Figures *3*-*4* illustrate Round 1 forecasting behaviour for two commodities, Bananas and Groundnuts. Across both the ensemble and recurrent models (Random Forest, XGBoost, GRU), track the actual demand curve more closely than the eliminated models, which either flatline or drift away from the true values.

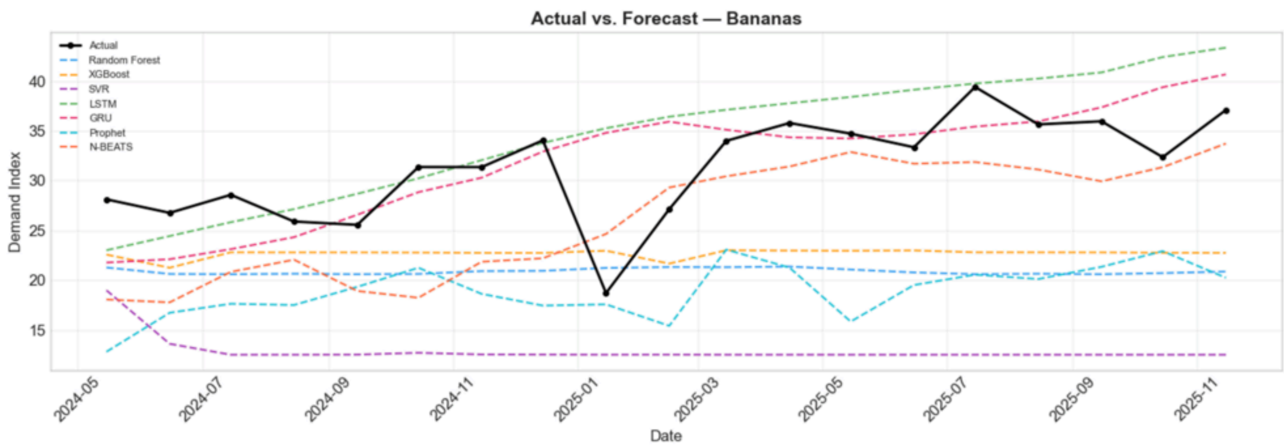

FIGURE 3: Actual vs. forecasted demand for Bananas across all seven Round 1 models. XGBoost and Random Forest stay closest to the actual (black) line and follow the upward trend into late 2025 more reliably than the other approaches

XGBoost, Extreme Gradient Boosting; GRU, Gated Recurrent Unit; SVR, Support Vector Regression; N-BEATS, Neural Basis Expansion Analysis for Time Series; LSTM, Long Short-Term Memory

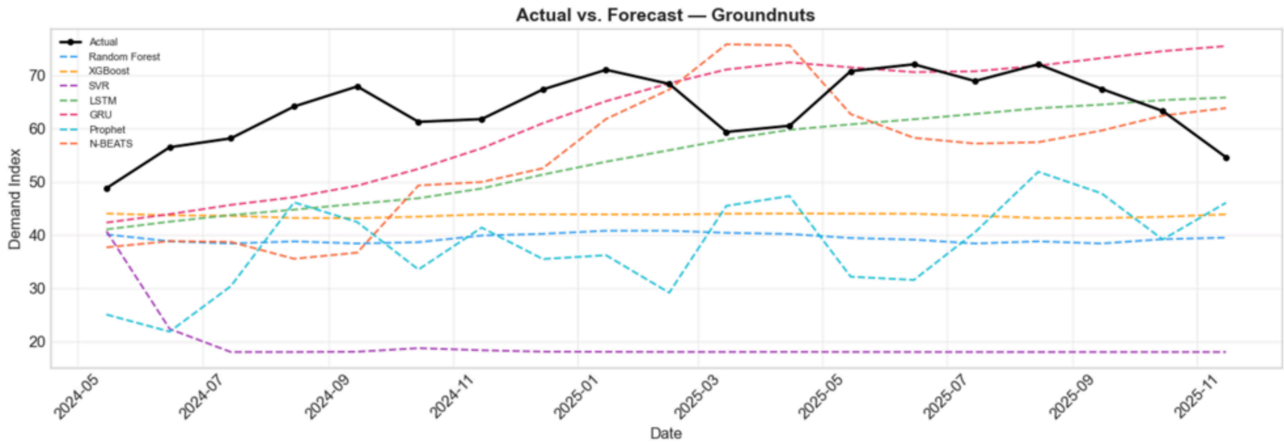

FIGURE 4: Actual vs. forecasted demand for Groundnuts. The pattern mirrors that of Bananas: ensemble models track the actual trend, while SVR flatlines and Prophet diverges

XGBoost, Extreme Gradient Boosting; GRU, Gated Recurrent Unit; SVR, Support Vector Regression; N-BEATS, Neural Basis Expansion Analysis for Time Series; LSTM, Long Short-Term Memory

In Round 2, the three finalists were retrained with more specific configurations (Table 3). For Random Forest and XGBoost, we increased the estimator count from 200 to 500 and raised the maximum tree depth. For GRU, its training epochs went from 100 to 200, with more patience before early stopping. The champion model for each commodity was, once again, the one with the lowest RMSE on the test set.

Model	Round 1 Parameters	Round 2 Parameters
Random Forest	n_estimators=200, max_depth=10, min_samples_split=5	n_estimators=500, max_depth=15, min_samples_split=3
XGBoost	n_estimators=200, max_depth=6, learning_rate=0.1, subsample=0.8, colsample_bytree=0.8	n_estimators=500, max_depth=8, learning_rate=0.05, subsample=0.85, colsample_bytree=0.85
GRU	patience=10, epochs=100	patience=15, epochs=200

TABLE 3: Hyperparameter configurations: Round 1 vs. Round 2

XGBoost, Extreme Gradient Boosting; GRU, Gated Recurrent Unit

Table 4 shows Round 2 results for five sample commodities after tuning.

Commodity	RMSE	MAE	R^2	MAPE
Bananas	4.7190	3.6796	0.0963	13.39
Beans (Red)	12.7463	10.7356	0.2332	24.09
Beans (White)	11.6484	9.5156	0.4636	18.23
Bread	4.8461	4.0902	-1.0824	23.51
Cassava mean (gari, yellow)	4.1368	3.5914	0.1030	14.15

TABLE 4: Round 2 error scores – five sample commodities

RMSE, Root Mean Square Error; MAE, Mean Absolute Error; MAPE, Mean Absolute Percentage Error

Figure *5* shows the final model win distribution across all 39 commodities. Random Forest won the most, 19 of 39 (48.7%). GRU took 14 (35.9%) and XGBoost took 6 (15.4%). Random Forest's advantage was its ability to perform well on noisier, more irregular commodities, which describes most of the dataset.

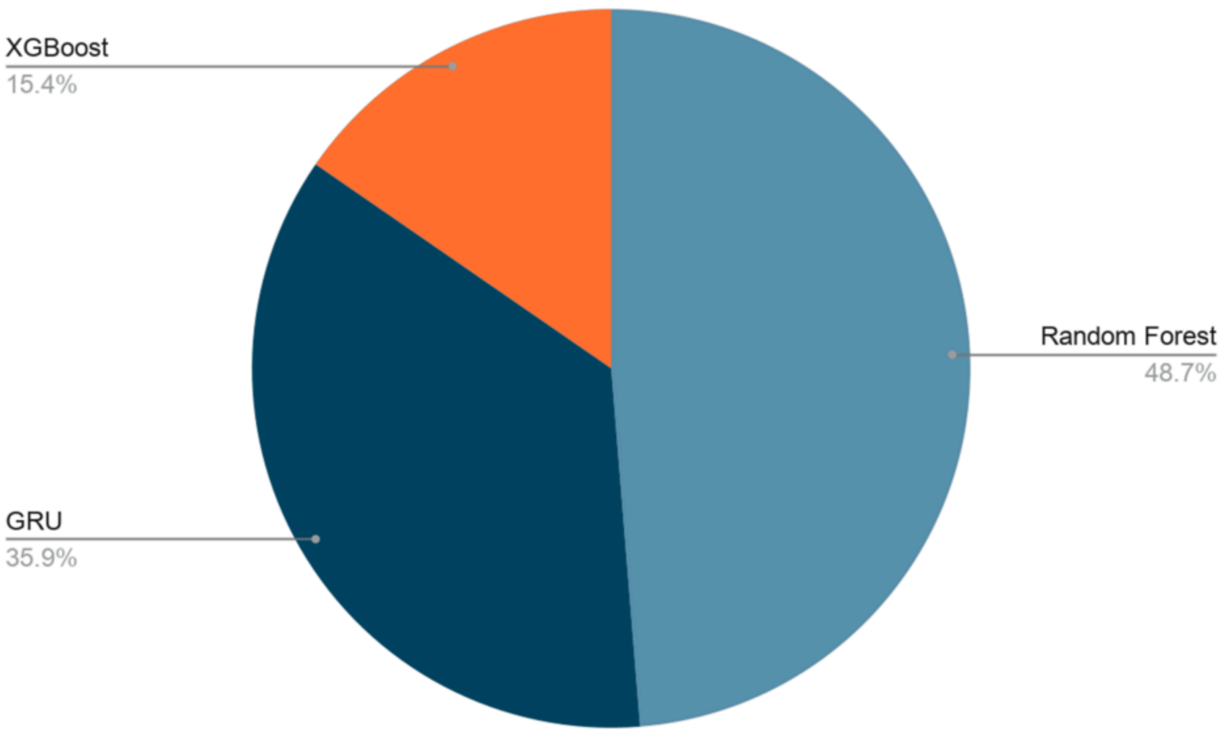

FIGURE 5: Champion model distribution across 39 commodities after Round 2. Random Forest won nearly half of all commodity championships, with GRU second and XGBoost third

XGBoost, Extreme Gradient Boosting; GRU, Gated Recurrent Unit

Discussion

While overall performance was strong for many commodities, we observed significant variation. Four commodities - Bread (−1.08), Milk powder (−3.23), Fuel diesel (−1.95), and Rice local (−1.57) - returned negative R^2 values. For these goods, the models performed worse than simply predicting the mean. This poor performance is likely attributable to external, non-autoregressive factors such as government price controls, import restrictions, or fuel subsidies, which disrupt historical price patterns. In the deployed application, the system explicitly flags predictions for these commodities as "unreliable," advising users against their use for critical planning.

Feature importance was extracted from the 19 Random Forest models that were selected as champions. Averaging these importances across the winning models revealed that the three most influential predictors were Price_RollingMean3 (the 3-month rolling average), Price_Lag1 (previous month's price), and Price_RollingMean6 (the 6-month rolling average). Conversely, weather variables (rainfall, temperature) and holiday indicators consistently registered near-zero importance. This finding strongly suggests that Nigerian commodity markets are highly autoregressive, with recent historical price trends being the dominant drivers of future price signals, rather than immediate environmental conditions.

The deployment of predictive tools in agriculture, particularly in developing economies, carries inherent ethical responsibilities and practical limitations. We address these concerns through several key points:

1. Proxy Limitations: The most significant limitation is the use of price as a proxy for market activity. As the reviewer correctly noted, if a price drop is misinterpreted as low demand rather than a supply glut, it could lead to harmful production or distribution decisions for farmers. We explicitly state throughout the manuscript and within the deployed

application that this tool forecasts price behaviour, not actual consumption volume. Users are cautioned that a high price signal does not always guarantee a market for additional volume, and vice versa. This distinction is crucial for responsible interpretation.

2. Representational Bias: Our models are trained on data from the World Food Programme, which primarily reflects prices in urban and regional hub markets. These signals may not accurately represent the "farm-gate" prices or local market dynamics experienced by smallholder farmers in remote rural areas [14]. This representational gap can introduce bias, potentially disadvantaging the very communities the tool aims to assist. Addressing this requires continuous effort to collect locally representative data, a challenge that is often more complex than model tuning.

3. Decision Support, Not Direction: The "Intelligent Demand Predictor" is designed as a supplementary decision-support aid. It provides one signal among several and should be used to inform, rather than dictate, agricultural planning. Farmers and supply-chain actors are advised to integrate their outputs with their local market knowledge, traditional wisdom, and advice from agricultural extension services. The tool's outputs should be treated as a "second opinion" to optimise harvest timing and distribution, not as definitive directives.

4. Model Robustness: The reliance on lagged features means that the models are susceptible to structural breaks in historical patterns, such as sudden policy changes, transport strikes, or export bans. Such events can render autoregressive forecasts unreliable. Future work could explore frameworks designed to handle structural breaks, as described by Lotfi et al. [15].

Conclusions

We developed and evaluated an ML framework for forecasting price-based market signals across 39 agricultural commodities in Nigeria. Through a rigorous two-round tournament incorporating time-series cross-validation and conventional baselines, Random Forest emerged as the most robust model, particularly adept at handling the inherent volatility of West African market data. Our feature importance analysis revealed that recent historical price trends are the dominant predictors, underscoring the autoregressive nature of these markets. This finding has significant practical implications, suggesting that effective market intelligence tools can be built and maintained with existing data, reducing the need for costly sensor infrastructure.

By explicitly reframing the focus from "demand volume" to "price signaling," we have developed a more methodologically sound tool that aligns with available data. The deployed Flask web application provides a practical interface for farmers and supply-chain actors to access these market insights. Future work should prioritize integrating live data feeds for automatic updates and expanding the interface to more accessible platforms like SMS or USSD, ensuring that the tool reaches smallholder farmers who may lack smartphone access. Crucially, the ethical deployment of such predictive systems necessitates ongoing vigilance regarding data bias and the clear, transparent communication of model limitations to all end-users.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Victor D. Ibironke, Sunday O. Oladipo, Osedebame E. Itamah, Afolashade Kuyoro

Acquisition, analysis, or interpretation of data: Victor D. Ibironke, Sunday O. Oladipo, Osedebame E. Itamah, Nelson C. Isicheli, Afolashade Kuyoro

Drafting of the manuscript: Victor D. Ibironke, Sunday O. Oladipo, Osedebame E. Itamah, Nelson C. Isicheli, Afolashade Kuyoro

Critical review of the manuscript for important intellectual content: Sunday O. Oladipo, Afolashade Kuyoro

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The datasets (and/or code) supporting this study are available from the corresponding author upon reasonable request.
All data generated or analyzed during this study are included in this published article and/or its appendices.
The data (and/or code) supporting this study are openly available at https://github.com/victoribironke/predictive-demand-forecasting

References

1. Kolawole ID, Kolawole GO, Sanni-manuel BA, Kolade Kolawole SK, Ewansiha JU, Kolawole VA, Kolawole FO: Economic impact of waste from food, water, and agriculture in Nigeria: challenges, implications, and applications—a review. Discover Environment. 2024, 2:51. 10.1007/s44274-024-00086-6
2. Business Day: World Food Day: Tackling 40% food wastage in Nigeria as solution to food insecurity. (2024). https://businessday.ng/features/article/world-food-day-tackling-40-food-wastage-in-nigeria-as-solution-to-food-insecu....
3. Food and Agriculture Organization: Nigeria at a glance. (2025). https://www.fao.org/nigeria/fao-in-nigeria/nigeria-at-a-glance/en/.
4. Odion D, Gajardo J, Defraeye T, Motmans T, Shoji K, Evangelista R, Onwude D: Towards improving farmers livelihood in Nigeria using food price forecasting. Journal of Agriculture and Food Research. 2025, 24:102365. 10.1016/j.jafr.2025.102365
5. Opara IK, Silue Y, Opara UL, Fawole OA: Forecasting of postharvest fresh produce waste at the wholesale level using time series models. Agriculture & Food Security. 2025, 14:24. 10.1186/s40066-025-00550-3
6. Rathod S, Chitikela G, Bandumula N, Ondrasek G, Ravichandran S, Sundaram RM: Modeling and forecasting of rice prices in India during the COVID-19 lockdown using machine learning approaches. Agronomy. 2022, 12:2133. 10.3390/agronomy12092133
7. Kasar A, Tripathi MM, Indolia S, Rawat A, Bathia A: Data-driven decision making for perishable food supply chains: insights from demand forecasting models. Advances in Consumer Research. 2025, 2:1461-1468.
8. Abubakar MA, Okpe BE, Audu M: Predicting food price trends in Nigeria using advanced machine learning techniques: LSTM and XGBoost. International Journal of Research and Innovation in Social Science. 2025, IX:3228-3237. 10.47772/ijriss.2025.905000251
9. Rodrigues M, Migueis V, Freitas S, Machado T: Machine learning models for short-term demand forecasting in food catering services: A solution to reduce food waste. Journal of Cleaner Production. 2024, 435:140265. 10.1016/j.jclepro.2023.140265
10. Olawale RA, Olawumi MA, Oladapo BI: Sustainable farming with machine learning solutions for minimizing food waste. Journal of Stored Products Research. 2025, 112:102611. 10.1016/j.jspr.2025.102611
11. Haque AU, Al Haque MA, Alabduladheem A, Al Mulla A, Almulhim N, Srinivasagan R: Machine learning-based shelf life estimator for dates using a multichannel gas sensor: enhancing food security. Sensors. 2025, 25:4063. 10.3390/s25134063
12. Kolawole OA, Owoigbe KV: The role of digital technologies in reducing food waste and loss in agricultural supply chains. IRE Journals. 2025, 8:479-487.
13. Birkmaier A, Imeri A, Riester M, Reiner G: Preventing waste in food supply networks - a platform architecture for AI-driven forecasting based on heterogeneous big data. Procedia CIRP. 2023, 120:708-713. 10.1016/j.procir.2023.09.063
14. Sunday O: Ethical dilemma in artificial intelligence: a focus on natural language processing bias. Information Matters. 2022, 2:1-5.
15. Lotfi R, Gholamrezaei A, Kadłubek M, Afshar M, Ali SS, Kheiri K: A robust and resilience machine learning for forecasting agri-food production. Scientific Reports. 2022, 12:21787. 10.1038/s41598-022-26449-8