

Predicting Philippine Food Prices with Machine Learning: A Random Forest and Neural Network Analysis of 153,000 WFP Data Points

Exzil Calanza

Independent Researcher, Iloilo City, Philippines

Independent research preprint. IEEE-style formatting only; this work has not been accepted, published, or peer reviewed by IEEE.

ABSTRACT This article examines Philippine food price prediction using World Food Programme monitoring data and machine-learning models. The original study is a Random Forest & Neural Network Analysis of 153,000 WFP Data Points; the analyzed dataset contains 153,404 price observations spanning January 2000 through December 2023 across 73 commodities and 17 administrative regions. A Random Forest regressor trained on 2000–2019 data and evaluated on a held-out 2020–2023 test set achieves an R^2 of 0.91, with Mean Absolute Error of ₱12.34 and RMSE of ₱18.76. Lagged prices contribute 42% of Random Forest feature importance. A separate PyTorch feedforward-network implementation was smoke-tested on Iris classification at 90.00% validation and 93.33% test accuracy before the price-regression implementation was built with MSELoss; those Iris scores do not validate the food-price regression model. The analysis also documents regional price differences, historical shocks affecting rice and pork, reproducibility details, and food-security applications. Updates deployed on March 18, 2026 add a 2-layer LSTM, exogenous features, climate scenario analysis, a 4-detector early-warning system, ensemble stacking, and 174 automated tests. The resulting system is presented as a basis for anticipatory intervention while retaining stated limitations concerning sparse early regional data and measurement inconsistencies.

INDEX TERMS — data science, food price prediction, machine learning, neural network, Philippines, PyTorch, random forest, WFP data.

I. INTRODUCTION

The Philippines faces a uniquely precarious food security position. As a nation of over 115 million people across 7,641 islands [9] (#ref-9), the archipelagic geography creates supply chain vulnerabilities that continental nations never encounter. The country is a net food importer—spending approximately \$3.5 billion annually on rice imports alone [7] (#ref-7). When India restricted rice exports in 2023 (affecting roughly 40% of global supply), the Philippines absorbed those shocks directly through retail price surges hitting the poorest households hardest [6] (#ref-6).

Climate exposure compounds the challenge: an average of 20 typhoons per year strike the Philippines, with 8–9 making landfall [1] (#ref-1). Super Typhoon Odette (Rai) in December 2021 destroyed over ₱20 billion in crops across Visayas and Mindanao [10] (#ref-10). The 2023–2024 El Niño reduced rice yields by 5–8% in Central Luzon and Cagayan Valley [12] (#ref-12). Meanwhile, PSA food inflation data reveals the CPI food component surged from 2.6% (2020) to 10.9% (January

2023) [2] (#ref-2)—eroding purchasing power for the 18.1% of Filipinos below the poverty line [9] (#ref-9). Red onion prices in late 2022 briefly exceeded imported beef prices at ₱600–700/kg in Metro Manila wet markets [10] (#ref-10).

Population growth of 1.5 million annually (UN projects 115.6 million by 2026) [9] (#ref-9), rising urbanization at 48%, and shifting consumption toward animal protein further strain domestic production. Rice self-sufficiency remains aspirational: domestic output of ~20 million metric tons falls short of the 22–23 million required [7] (#ref-7) [12] (#ref-12). This convergence—import dependence, climate exposure, inflation volatility, population growth—makes food price prediction a critical input for national planning [11] (#ref-11). ML models trained on granular historical data offer anticipatory food policy that reduces the lag between price shocks and government response.

II. BACKGROUND AND RELATED WORK

The World Food Programme’s Vulnerability Analysis and Mapping (VAM) unit maintains one of the most

comprehensive food price monitoring systems in the developing world [1] (#ref-1). For the Philippines, this dataset spans January 2000 through December 2023—a 24-year longitudinal record of 153,404 price observations across 73 commodities, 17 administrative regions, and multiple province/locality levels [5] (#ref-5). Data is collected from public wet markets, trading posts, and wholesale centers throughout the archipelago.

Each record contains 12 fields: Month, Day, Year, Region, Province, Locality, Location (lat/long), Category, Commodity, Priceflag, Pricetype, and Price (PHP) [5] (#ref-5). The six food categories—cereals and tubers, meat/fish/eggs, miscellaneous food, oil and fats, pulses and nuts, and vegetables and fruits—cover the Filipino diet comprehensively. Retail prices dominate at 146,478 observations (95.5%), with Wholesale at 6,262 (4.1%) and Farm Gate at 664 (0.4%). This class imbalance must be handled through stratified sampling or categorical encoding.

Geographic coverage spans all 17 regions, from NCR (Metro Manila) to ARMM (now BARMM) [12] (#ref-12). Data density varies: urbanized regions like NCR and Region III have consistently dense series, while conflict-affected regions like ARMM have sparser coverage in earlier years due to security-related collection challenges.

Twenty-four years of WFP price data reveal striking patterns that inform both feature engineering and model architecture [5] (#ref-5). The most apparent trend is the secular upward movement in staple food prices, driven by domestic inflation, global commodity dynamics, and structural agricultural shifts [6] (#ref-6).

Rice—the most politically sensitive Philippine commodity—saw retail prices double from ₱17.87/kg (2000) to ₱41.41/kg (2023), a 131.7% increase [5] (#ref-5). This was non-linear: the 2008 global food crisis produced a 31% single-year jump (₱22.95 to ₱30.07) [6] (#ref-6), followed by a plateau at ₱31–33/kg for five years, then a second structural shift in 2014 to the ₱37–41 range. The Rice Tariffication Law (RA 11203, 2019) briefly lowered prices—from ₱41.43 (2018) to ₱37.45 (2020)—before pandemic disruptions and the 2023 Indian export ban reversed those gains [7] (#ref-7).

Pork prices surged 244% over the full period (₱91.75 to ₱315.73/kg) [5] (#ref-5). African Swine Fever, confirmed in September 2019, devastated the domestic hog industry: prices jumped 36.6% from ₱214.24 (2019) to ₱292.64 (2021) as the national herd declined by ~2.5 million heads [10] (#ref-10) [12] (#ref-12). Government price ceilings and expanded import quotas stabilized but couldn't reverse the escalation.

Vegetables show different behavior: less dramatic long-term trends but far higher intra-year volatility [5] (#ref-5). Tomato prices can swing 40%+ within a year due to concentrated growing regions and limited cold-chain infrastructure. A typhoon hitting Benguet or Bukidnon during harvest can double Metro Manila retail prices within weeks [10] (#ref-10)—presenting both a modeling challenge (external shock noise) and opportunity (strong seasonality signal).

III. METHODOLOGY

Transforming raw WFP records into an ML-ready feature matrix requires engineering temporal, geographic, and domain-specific features [3] (#ref-3). The raw 12 columns expand to ~120 features after encoding—a dimensionality that favors tree-based ensembles, which handle high-dimensional sparse features without curse-of-dimensionality issues [3] (#ref-3).

Temporal features form the predictive backbone. Beyond raw Year, Month, and Day, we extract Quarter, Day-of-year, and Week-of-year to capture planting and harvest cycles [8] (#ref-8). Month is sine/cosine-transformed to preserve cyclical continuity (December is adjacent to January). The Philippines' wet season (June–November) and dry season (December–May) create predictable supply patterns these cyclical features encode.

Lagged price features are the single most important input, contributing 42% of feature importance [5] (#ref-5). For each observation, we compute the same commodity's price in the same region at t-1, t-3, t-6, and t-12 months. These autoregressive features capture momentum and mean-reversion. The t-12 lag is especially valuable for seasonal commodities where the best comparator is the same month in the prior year.

Rolling statistics capture trend and volatility: 3-month moving average (short-term smoothing), 6-month moving average (medium-term trends), and 12-month rolling standard deviation (price instability signal) [8] (#ref-8). **Categorical encoding** converts Region (17), Category (6), Commodity (73), and PriceType (3) into one-hot binary vectors—99 binary features preserving non-ordinal structure. Target encoding was tested but discarded due to leakage during cross-validation.

Geographic features use Location latitude/longitude to capture spatial price gradients, including a distance-to-NCR feature encoding logistical cost as a continuous variable [12] (#ref-12). **Interaction features** (Region × Commodity, Month × Commodity) capture commodity-specific regional effects and seasonality—pork prices in ARMM behave differently from pork prices in Region III beyond additive main effects.

The Random Forest regressor (scikit-learn [8] (#ref-8)) serves as the primary production model. Random Forests handle mixed feature types natively, provide built-in feature importance for interpretability, and tolerate non-stationarity in price time series without explicit differencing [3] (#ref-3). The final model uses 500 trees, max depth 20, and minimum 5 samples per leaf—hyperparameters selected via 5-fold time-series cross-validation with temporal ordering (each validation fold strictly follows its training fold).

The temporal train/test split uses 2000–2019 for training (~122,000 records) and 2020–2023 as the held-out test set (~31,000 records) [5] (#ref-5). This tests generalization to an out-of-sample period containing two unprecedented disruptions: COVID-19 (2020–2021) and the ASF aftermath (2020–2022). The R^2 of 0.91 indicates 91% of price variance is explained—a strong result given test-period disruptions.

Mean Absolute Error of ₱12.34 indicates typical prediction miss; RMSE of ₱18.76 penalizes large errors more heavily. The ₱6.42 gap between RMSE and MAE indicates some high-error outliers, likely corresponding to extreme ASF-era and pandemic-era price spikes [10] (#ref-10).

A feedforward neural network complements the Random Forest with capacity for nonlinear feature interactions [4] (#ref-4). Architecture: Input(4) → FC(50, ReLU) → FC(50, ReLU) → Output(3), trained with Adam (lr=0.01) and CrossEntropyLoss over 100 epochs. The implementation was first smoke-tested on Iris classification (90.00% validation, 93.33% test accuracy); those classification scores are an implementation check only and do not validate the food-price regression model. The price-regression adaptation uses MSELoss.

The training loss curve shows rapid convergence: 92% reduction from 0.2778 (epoch 10) to 0.0218 (epoch 30), continuing to 0.0004 at epoch 100 [4] (#ref-4). The smooth decay without spikes confirms stable optimization. Neural networks complement Random Forests by learning smooth decision boundaries that extrapolate more gracefully outside

training distribution [3] (#ref-3); combining both via stacking yields more robust predictions.

Analysis was conducted using Python 3.13.2 with PyTorch 2.x [4] (#ref-4), scikit-learn 1.x [8] (#ref-8), pandas 2.x, and numpy 1.x. The WFP dataset from dataviz.vam.wfp.org [5] (#ref-5) required minimal preprocessing: 342 null-price records dropped (0.22%), and 17 clearly erroneous prices (e.g., ₱0.01/kg or ₱999/kg rice) removed via IQR filtering.

The pipeline enforces strict temporal ordering: all feature engineering uses only historically available data at each time step. The 2019/2020 train/test boundary is fixed; all hyperparameter selection uses only 2000–2019 data with time-series CV folds. Model serialization: joblib for Random Forest (~450 MB for 500 trees), PyTorch state_dict for the NN (~50 KB). Serving can be exposed through Flask/FastAPI; this study does not report a controlled serving-latency benchmark.

IV. RESULTS

TABLE I. Dataset Composition by Price Type

TABLE 1		
Price Type	Observations	Share
Retail	146,478	95.5%
Wholesale	6,262	4.1%
Farm Gate	664	0.4%

TABLE II. Random Forest Evaluation

TABLE 2	
Metric	Value
R ²	0.91
Explained price variance	91%
Mean Absolute Error	₱12.34
RMSE	₱18.76
RMSE-MAE gap	₱6.42

TABLE III. Random Forest Feature Importance (%)

TABLE 3	
Feature Group	Importance
Lagged Prices	42%
Commodity Type	18%
Region	12%
Month	11%
Year	8%
Other Features	9%

TABLE IV. PyTorch Validation and Training Checkpoints

TABLE 4	
Measure	Value
Iris validation accuracy	90.00%
Iris test accuracy	93.33%
Training loss, epoch 10	0.2778
Training loss, epoch 20	0.0726
Training loss, epoch 30	0.0218
Training loss, epoch 40	0.0088
Training loss, epoch 70	0.0010
Training loss, epoch 100	0.0004

TABLE V. 2023 Regional Rice Price Points (Retail PHP/kg)

TABLE 5	
Region	Price
ARMM (now BARMM)	₱48.04/kg

Region VII (Central Visayas)
NCR (Metro Manila)
Region I (Ilocos)
Region II (Cagayan Valley)

₱44.03/kg
₱40.44/kg
₱38.99/kg
₱37.96/kg

Fig. 1. Average Price by Food Category (PHP)

Fig. 2. Top 10 Commodities by Data Point Count

Fig. 3. Random Forest Feature Importance (%)

Fig. 4. PyTorch Training Loss by Epoch

The 17 administrative regions exhibit persistent, economically significant price divergences reflecting structural differences in logistics, production capacity, and market integration [5] (#ref-5) [12] (#ref-12). The 2023 rice breakdown reveals a ₱10+/kg spread between highest and lowest-cost regions—meaningful for households spending 40–60% of income on food [9] (#ref-9).

ARMM (now BARMM) records the highest rice prices at ₱48.04/kg—16% above the national average [5] (#ref-5). This premium reflects limited road and port infrastructure, security-related transport disruptions, small local rice production, and multi-middleman supply chains [11] (#ref-11). Region VII (Central Visayas) at ₱44.03/kg faces island logistics costs: limited rice paddy area means dependence on inter-island shipments from Mindanao and Luzon, with port handling fees and perishability markups.

At the other end, Region II (Cagayan Valley) records the lowest at ₱37.96/kg—a direct consequence of being a top rice-producing region where consumer proximity to farms eliminates intermediaries [7] (#ref-7). Region I (Ilocos) similarly benefits from its Luzon rice belt location at ₱38.99/kg. NCR (Metro Manila) achieves moderate ₱40.44/kg despite being the largest consumer market—reflecting efficient port infrastructure, competitive retail, and high-volume price transparency [2] (#ref-2).

A. Recent Updates (March 18, 2026)

Updated: March 18, 2026 20:00 PHT — Six major enhancements have been implemented, cross-validated by independent AI agents, and deployed. For the complete technical reference of the upgraded system, see the upgraded system post (/ph-food-price-after-march-2026-upgrade/).

B. 1. LSTM Deep Learning Model (March 18, 2026 08:15 PHT)

A PyTorch 2-layer LSTM network (hidden_size=128, dropout=0.2) with early stopping and gradient clipping captures non-linear temporal patterns missed by traditional models. The architecture uses cyclical month encoding (sin/cos transforms), normalized year features, and region/commodity encodings over 12-month sliding windows. Per-commodity models are trained with AdamW optimizer, HuberLoss (robust to price spike outliers), and ReduceLROnPlateau scheduling. 639 lines. *Replaces the original simple feedforward neural network.*

C. 2. Exogenous Feature Integration (March 18, 2026 08:15 PHT)

Real-time external data sources are now integrated into the forecasting pipeline: NOAA Oceanic Niño Index (ENSO climate state), USD/PHP exchange rates via the Frankfurter API, and the FAO Food Price Index. Derived features include lagged values (1–6 months), ENSO state dummies (El Niño/La Niña/Neutral), and momentum indicators. All external fetches use local JSON caching with configurable TTL. *Previously, the model used WFP price data only.*

D. 3. Climate Scenario Analysis (March 18, 2026 08:15 PHT)

ENSO-price correlation modeling across 7 climate states (Strong/Moderate/Weak El Niño and La Niña + Neutral) quantifies how climate events impact Philippine food prices. Historical price change analysis (mean, median, standard deviation) validates price differences across ENSO phases, with granular per-commodity and per-region breakdowns. Impact lag testing at 0, 3, and 6 months captures delayed climate effects. *Previously listed as future work.*

E. 4. Early Warning System (March 18, 2026 08:15 PHT)

A 4-detector anomaly detection system identifies price spikes, year-over-year surges, regional divergence, and model-forecast divergence using a FEWS NET-inspired 4-level severity classification (low/medium/high/critical). Each severity level maps to specific policy recommendations. An interactive Leaflet.js map dashboard visualizes 949 generated alerts geographically. *Previously listed as future work.*

F. 5. Ensemble Stacking Model (March 18, 2026 08:15 PHT)

A scikit-learn StackingRegressor combining Gradient Boosting (200 trees, depth 4), Extra Trees (200 trees, depth 10), and Random Forest (200 trees, depth 10) as base estimators with a Ridge Regression (alpha=1.0) meta-learner, using 5-fold cross-validation. *Previously, only single-model predictions were available.*

G. 6. Comprehensive Test Suite (March 18, 2026 10:30 PHT)

174 automated tests (verified via pytest) across 4 test files cover data quality validation, model accuracy benchmarks, API endpoint testing, ensemble integration, and exogenous feature pipeline integrity. Cross-validation caught and fixed 3 bugs: La Niña ONI mask inversion, ensemble NameError, and CV fold incompatibility. *Previously, no automated tests existed.*

Additional infrastructure upgrades: Dark/light theme toggle, PWA support, Excel/CSV/PNG/JSON export, zoom/pan charts, REST API (7 endpoints), parallel model training (25 variants), automated daily updates with retry logic and data quality gates, 355-line README with full API reference. See the upgraded system reference (/ph-food-price-after-march-2026-upgrade/) for the complete capability stack.

V. DISCUSSION

ML price prediction enables the shift from reactive crisis management to anticipatory intervention [11] (#ref-11). An **early warning system** can flag emerging price anomalies 1–3 months before retail peaks, giving the Department of Agriculture time to authorize emergency imports or release NFA buffer stocks before prices destabilize [10] (#ref-10).

Targeted regional intervention becomes possible with disaggregated models. The ₱10/kg gap between ARMM (₱48.04) and Cagayan Valley (₱37.96) suggests a targeted transport subsidy in BARMM would deliver more food security per peso than blanket national price ceilings [5] (#ref-5) [11] (#ref-11). **Buffer stock optimization** benefits from 3–6 month price forecasts: release stock at predicted peaks, replenish at seasonal lows—reducing both taxpayer cost and price volatility [10] (#ref-10).

Import timing optimization addresses a recurring challenge [6] (#ref-6). When models predict domestic prices will exceed import parity, the government can tender import contracts weeks ahead rather than scrambling for emergency imports after prices surge—as happened in the 2023 rice crisis [7] (#ref-7). **Climate adaptation planning** represents the frontier: integrating ENSO forecasts (predictable 6–9 months ahead) as exogenous features enables climate-conditional scenarios like “If El Niño reaches +1.5°C, Region III rice prices projected to rise 8–12% in Q3,” supporting pre-positioned interventions [1] (#ref-1).

VI. THREATS TO VALIDITY

Limitations include data sparsity in early years for some regions (ARMM pre-2008) and measurement inconsistencies across collectors. Exogenous features (weather, global indices, exchange rates) have now been integrated as described above.

Future work: satellite imagery (NDVI) integration for agricultural yield prediction, real-time API with rolling forecasts served via FastAPI, Philippine ePrice system integration for automated early warning, and transformer-based architectures (temporal fusion transformers) for multi-horizon forecasting.

The article also notes the original price-type imbalance, with Retail at 146,478 observations (95.5%), Wholesale at 6,262 (4.1%), and Farm Gate at 664 (0.4%), and it notes that conflict-affected regions such as ARMM have sparser coverage in earlier years.

VII. CONCLUSION

The study presents Philippine food-price forecasting as an application of longitudinal WFP monitoring data, temporal and categorical feature engineering, Random Forest regression, and neural-network modeling. On the held-out 2020–2023 period, the Random Forest achieves an R^2 of 0.91, while lagged prices contribute 42% of feature importance. The historical and regional analyses show that national averages conceal substantial commodity-specific and location-specific variation. The March 18, 2026 upgrades extend the original workflow with LSTM modeling, exogenous data, climate scenarios, anomaly detection, ensemble stacking, and automated tests. The article therefore frames forecasting as a tool for earlier and more targeted intervention, while retaining explicit limitations around sparse early regional data and measurement inconsistencies.

KEY TAKEAWAYS

- A Random Forest model trained on 153,404 WFP price observations achieves an R^2 of 0.91 on held-out 2020–2023 data, demonstrating that Philippine food prices are highly predictable from historical patterns, commodity type, and regional factors.
- Rice prices increased 131.7% over 23 years (₱17.87 to ₱41.41/kg), but the increase was concentrated in two structural step-changes: the 2008 global food crisis and the 2014–2018 domestic supply tightening.
- Regional price disparities of up to ₱10/kg for rice (ARMM at ₱48.04 vs. Cagayan Valley at ₱37.96) reflect persistent infrastructure and logistics gaps that policy interventions should target directly.
- Lagged price features accounted for 42% of the Random Forest feature-importance mass in this run, suggesting strong short-term autoregressive signal. This descriptive feature-importance result does not establish that any particular data-collection investment is causally optimal.
- The African Swine Fever crisis (2019–2021) produced the largest commodity-specific price shock in the dataset, with pork prices surging 36.6% in two years—a magnitude that highlights the need for disease surveillance integration into food price early warning systems.

REFERENCES

- [1] World Food Programme, “Philippines Country Page,” [Online]. Available: wfp.org
- [2] Philippine Statistics Authority, “Consumer Price Index,” [Online]. Available: psa.gov.ph
- [3] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. [Online]. Available: springer.com
- [4] A. Paszke et al., “PyTorch: An Imperative Style, High-Performance Deep Learning Library,” *NeurIPS*, 2019. [Online]. Available: neurips.cc
- [5] WFP VAM, “Food Prices Data,” [Online]. Available: dataviz.vam.wfp.org

- [6] FAO, “Food Price Monitoring and Analysis,” [Online]. Available: fao.org
- [7] Philippine Rice Research Institute, “Rice Statistics,” [Online]. Available: philrice.gov.ph
- [8] F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” JMLR , vol. 12, pp. 2825–2830, 2011. [Online]. Available: jmlr.org
- [9] World Bank, “Philippines Overview,” [Online]. Available: worldbank.org
- [10] Department of Agriculture, “Price Monitoring,” [Online]. Available: da.gov.ph
- [11] Asian Development Bank, “Philippines Economy,” [Online]. Available: adb.org
- [12] Philippine Statistics Authority, “OpenSTAT: Agricultural Indicators,” [Online]. Available: openstat.psa.gov.ph

AUTHOR BIOGRAPHY

EXZIL CALANZA is an independent researcher and software builder based in Iloilo City, Philippines. His work focuses on browser and desktop automation, machine learning, reproducible evaluation, workflow verification, and AI-assisted systems engineering.