

Predicting Food Prices in Nigeria Using Machine Learning: Symbolic Regression

Olawale Basheer Akanbi and *Hameed Olamilekan Ajasa

Department of Statistics, University of Ibadan.

DOI: <https://doi.org/10.51584/IJRIAS.2025.10060074>

Received: 30 May 2025; Accepted: 07 June 2025; Published: 11 July 2025

ABSTRACT

The aim of this study is to predict the prices of local rice, beans, and Garri in the South West (SW) and North Central (NC), Nigeria using economic indicators such as exchange rate, inflation rate, crude oil price, past one month price (lag 1) and past five-month price (lag 5) of the food prices as the predictor variables. The data used were extracted from the website of the National Bureau of Statistics from January 2017 to July 2024. The data were split into training set and testing set.

The study proposed four machine learning techniques; random forest, decision tree, neural network and symbolic regression to model the prices of food, and the root mean square error (RMSE) was used as a criterion for the model evaluation and comparison. Findings showed that symbolic regression outperformed the other models in predicting the prices of beans in the NC with random forest, decision tree, neural network and symbolic regression having the RMSE values: 1365.41, 1348.86, 672.075 and 395.68 respectively. Similarly, symbolic regression outperformed the other models in predicting the prices of rice in the NC with random forest, decision tree, neural network and symbolic regression having the RMSE values: 662.19, 601.74, 1327.951 and 94.39 respectively. Similarly, in predicting the prices of Garri, symbolic regression outperformed others with random forest, decision tree, neural network and symbolic regression having the RMSE values: 442.57, 429.08, 920.8771, and 84.28 respectively.

The results showed that symbolic regression has the least RMSE for predicting the prices of local rice, local beans and gari in the North Central. More so, similar results were obtained for predicting the food prices in the South West. Hence, this study recommends the use of symbolic regression for econometric modelling and also future study to extend on the present work by adopting Bayesian symbolic regression.

Keywords - Food Price Prediction, Symbolic Regression, Machine Learning, North Center and South West

INTRODUCTION

A basic human need exists for food while several people lack sufficient food for their survival. Rice wheat and corn represent the three basic food crops which are globally consumed as the primary sources of grains (Mohapatra et al., 2022). According to Food and Agricultural Organization (2009), food crop consumption will surge quickly in the forthcoming years due to the global population expansion projected to 2.3 billion during the mid-century period (Bruinsma, 2017). As a result, the consumption of food will expand significantly throughout the next 30 years. World population expansion stands as a key determinant behind the rising demand (FAO, 2009).

However, the excruciating rise in food prices mainly impacts the poor while affecting agricultural dependent economies most severely since their incomes are highly devoted to purchasing staple foods (Compton et al., 2010). According to Yamauchi .and Larson (2019), periodic food price surges significantly increase poverty levels in urban food import zones of developing nations. Many agricultural-based economies have received negative impacts from fluctuating food prices that led international aid agencies and governments to take swift actions for lowering staple food prices comprising 60% of worldwide food consumption (Fasanya & Odudu, 2020).

Uncertainty regarding food prices creates substantial risks for all individuals operating within the agricultural sector (Liang et al., 2024). Precise food price predictions support farmers along with consumers and producers and policymakers to execute strategic planning and decision-making (Liang et al., 2024). The prediction process enables valuable discovery of price change drivers in the food market while creating direct impacts on supply chain operations (Liang et al., 2024). Through food budgeting decisions primitive users can determine what to purchase which results in specific nutritional outcomes (Liang et al., 2024). Market trends analysis through forecasting allows primary producers along with processors wholesalers and retailers to anticipate future market conditions while developing their operations plans and marketing production decisions.

However, Nigeria serves as the biggest economy in Africa because it produces oil while having more than 200 million inhabitants whose demographic growth reaches an annual average of 3% (Adeola et al., 2022). The Nigerian economy rests on agriculture since farmers provide sustenance for most citizens despite oil production. However, the sector faces many challenges, notably an outdated land tenure system that constrains access to land (1.8 ha/farming household), a very low level of irrigation development (less than 1% of cropped land under irrigation), limited adoption of research findings and technologies, high cost of farm inputs, poor access to credit, inefficient fertilizer procurement and distribution, inadequate storage facilities and poor access to markets have all combined to keep agricultural productivity low (average of 1.2 metric tons of cereals/ha) with high postharvest losses and waste (FAO, 2016).

Therefore, precise food price forecasts enable policymakers to monitor market tendencies and customer behavior better thus supporting them in implementing prompt policies to guarantee food safety and sustainable markets (Liang et al., 2024). Through predictive models, organizations benefit from guidance which affects all the steps between production and consumption (Oyewole et al., 2024). Acting as vital market tools for market participants including processors and speculators together with hedgers and policymakers (Raihan et al., 2023). The price forecast data assists producers to determine end prices before starting production and exporters need it to fulfill contractual terms while speculators use it for profits and hedgers use it for risk reduction and the implementation of strategic plans by policymakers (Dacha et al., 2021).

Similarly, several research studies presenting different time series methods used for prediction have been published in literature (Jin & Xu, 2024). The literature extensively mentions Vector Autoregressive (VAR) models as well as Autoregressive Integrated Moving Average (ARIMA) models together with Vector Error Correction (VECM) models and multiple model variants. Previous research demonstrates that the Autoregressive Integrated Moving Average (ARIMA) model qualifies as a top selection for different time series forecasting applications. Research findings established that ARIMA provides significantly superior predictions than expert views along with structural model-based forecasts for the US hog and cattle markets (Jin & Xu, 2024).

In addition, ARIMA performs best when used alongside multiple model types because its accuracy benefits from merging different data sources as proven by Elsaraiti and Merabet (2021). The VAR approach stands as a popular econometric technique used for price series forecasting because it understands the connected relations between economic variables. VECM utilizes cointegration to investigate economic variable relationships in a long-term framework because it shares close connections with the VAR model (Andrei & Andrei, 2015). The technique proves beneficial in making extended price predictions. The findings demonstrate that the VECM delivers superior performance to the VAR when modeling global wheat prices according to Tsega and Tsehay (2020). The ARIMA in conjunction with the EGARCH model and GARCH model received analysis by Lama et al. (2015) to forecast domestic and international edible oil prices. Based on volatility pattern detection the EGARCH model proved superior to its competitor models according to their findings. Wang et al. (2013) proved the seasonal VECM as an effective method to forecast prices of soybean and rapeseed oil in China. The GARCH-in-mean model together with volatility impulse response analysis serves Hasanov et al. (2016) to investigate whether crude oil price information can be used for edible oil price prediction.

Recently, scientific investigation into agricultural product price forecasting with machine learning techniques has intensified because researchers now have convenient access to computing resources and technology (Alade et al., 2021; Jin & Xu, 2024). The research field analyzing commodities including soybeans and sugar through neural networks jointly with genetic programming, deep learning, support vector regressions, random forests, K-nearest neighbors, multivariate adaptive regression splines, decision trees, ensembles, as well as boosting

together with deep learning and neural networks (Jin & Xu, 2024). Existing research shows that neural networks lead the popularity rankings for agricultural commodity price prediction models despite all other published findings (Xu & Zhang, 2023). Therefore, the current research implements a modern econometric tool known as symbolic regression to predict selected food prices in Nigeria together with existing machine learning techniques: random forest, decision tree and neural network as a way to contribute to this field.

Food affordability stands as a critical problem throughout Nigeria because a series of fundamental difficulties already stir the nation (Ajibade et al. 2020). The seasonal planting activities of Nigerian farmers deteriorated due to COVID-19 which disrupted their farming practices during the 2019/20 agricultural planting period. Agriculture in Nigeria continues to depend on weather conditions which prevents farmers who missed planting season from enacting catch-up strategies (Abang, 2023). More so, unpredictable droughts together with floods and diseases strongly determine these price levels. The complex nature of agricultural price forecasting exists because such key determinants become challenging to incorporate into food price forecasting models (Ngartera, 2024).

Also, food inflation remains a concern in Nigeria as the nation has dealt with the impact of fuel subsidy removal and security challenges based on findings from Adeniran et al., (2016). The country faces increased challenges due to heavy fuel subsidy spending because it amounted to ₦1.4 trillion (USD 3.7 billion) in 2020. The national government allocated 20% of its annual spending to this sector (Abang et 2024). Also, since removing fuel subsidies, the price of fuel has increased throughout the country. Higher production and transportation expenses have become a common result of the price increase.

Consequently, these frequent food price shocks negatively affect the Nigerians particularly the most vulnerable poor residents. Food inflation persisted because it reached 40.87% in June 2024 with 25.24% from June 2023 and 18.60% from June 2022 according to Balogun (2022). The unending rise of food prices throughout the years has forced many Nigerian people to consume small amounts of food because the cost of food access has become excessively challenging (Balogun, 2022).

Most traditional economic factors play an important role in determining food prices. The entire process begins with standard economic principles of supply and demand acting upon every tradable merchandise category. Also, evidence shows interest rates together with exchange rates have an influence on these prices. According to Ndidi (2013) food prices increase largely because of foreign exchange rates and the petroleum marketing surplus cost but not because of insecurity. According to Ngartera, (2024) the development of higher crude oil prices leads to increased transportation expenses which subsequently affects the prices of certain food commodities.

Therefore, the task of creating econometric models for agricultural commodities prices becomes complex because obscure variables such as weather and drought influence are unmeasurable and the primary price factors exhibit unpredictable volatility patterns. Moreso, the existing forecasting methods primarily utilize time-series models including ARIMA or GARCH-type (Generalized Auto-Regressive Conditional Heteroskedasticity) for volatility modeling according to Atsalakis & Valavanis. (2010). Analysis using neural networks joins other more advanced methods for modeling purposes (Jha & Sinha, 2013). The use of structural models represents another forecasting procedure (Andrei et al, 2015).

The majority of researchers utilize classical regression analysis to explain multiple explanatory variable models (Gorter et al., 2015). Econometric methods show two major limitations when used conventionally. This approach necessitates reducing relevant explanation variables down to select only a few variables. also, during the whole analytical period such methods presume the existence of one single “true” model from between the competing options (Mitra et al., 2017). The selection of the suitable model occurs by fitting the training period data. This method fails to recognize that different periods might require different underlying true models. Traditional models offer limited ability to detect possible non-linear relationships (Mitra et al., 2017).

Furthermore, the existing econometric models require further development because they provide unsatisfactory forecasting outcomes (Anjoy & Paul, 2019). This study proposes a new machine learning approach: symbolic regression algorithm which enables automatic modeling of data through various mathematical functions searched from the infinite mathematical function space to produce an effective result model. Therefore, the result

of the symbolic regression is compared with other machine learning techniques using the root mean square error criterion.

METHODOLOGY

Data Collection

The study used a quantitative method, since it requires the collection of secondary data on food prices from the National Bureau of Statistics (NBS) from January 2017 to July, 2024, equivalent to 91 observations. The data included three food prices selected for South West and North Central of Nigeria: the price of local brown beans, price of white gari and price of local rice. Also, exchange rate (in Naira), inflation rate and crude oil price (USD per barrel) were downloaded from investing.com website and the Central Bank of Nigeria (CBN) respectively.

Data Preprocessing

The data set contain twelve columns and 91 rows including year and date. In the data preprocessing, the 'Year' and 'Date' were converted and set to Datetime and afterwards lag 1 and lag 5 of the food prices were created and used as part of the predictors for each food item. Furthermore, the data was split into training set and testing set. The training set begins from January 2017 to December 2023 while the test set starts from January 2024 to July 2024. Furthermore, the input variables: exchange rate, inflation rate, crude oil price (in Naira), lag_1 and lag_5 of the food items were transformed using the "StandardScaler()" function from the scikit learn package in python before they are used in the model.

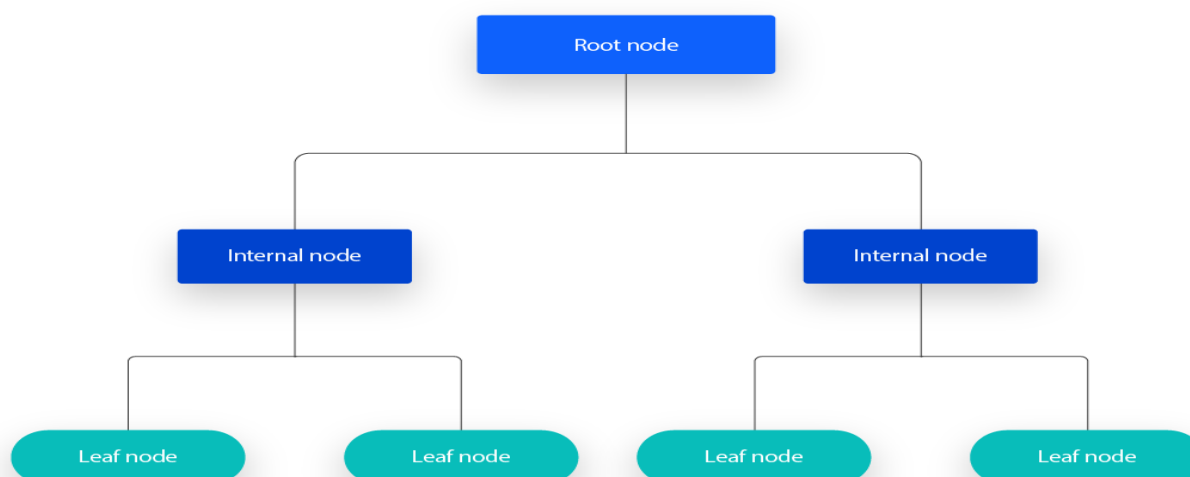
Data Analysis Techniques

The analysis utilized four machine learning algorithms including decision tree (DT), random forest (RF) and neural network (NN) and symbolic regression (SR) to model the food prices data collected for North Central and South West.

Decision Tree

Decision trees represent non-parametric supervised algorithms that support both classification and regression problems (Sonawane & Dhawale, 2016). The decision tree contains four structural elements that form a hierarchical tree through a root node together with branches and internal nodes while having leaf nodes at its end (Ying, 2015). During a decision tree algorithm execution, the dataset gets divided using selected features. The chosen split finds its best option through an impurity measurement approach (Gini Impurity or Entropy or Mean Squared Error for regression tasks). A tree node selects the best feature for splitting the dataset according to an impurity criterion that yields the most optimal partition (Ying, 2015).

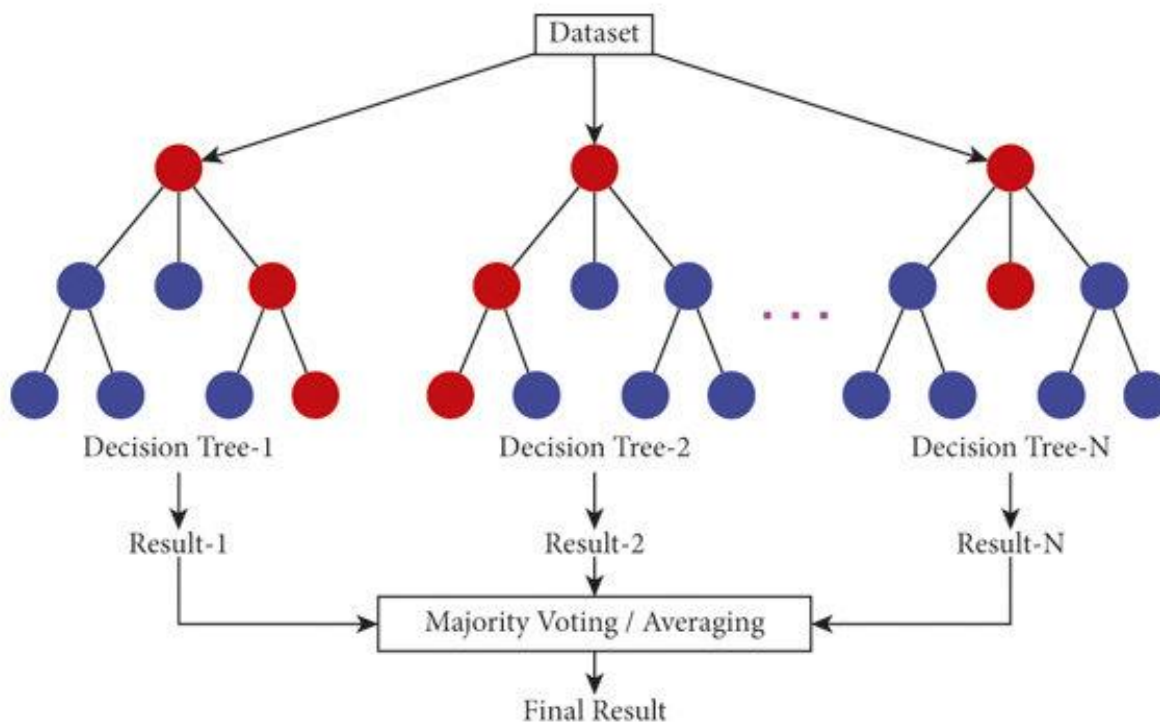
Fig 1: Decision Tree Architecture (Source: Khan et al., 2021)



Random Forest

During training Random Forest builds several decision trees to achieve improved accuracy and minimize overfitting according to Salman et al. (2024). Random Forest functions as an algorithm that handles both classification and regression problem sets. Random Forest functions through generating many decision trees which combine predictions through classification by majority voting or regression using averaging (Salman et al., 2024). Random forest operates as a user-friendly machine learning system which yields superb results naturally even when we avoid hyperparameter optimization (Salman et al., 2024). Random Forest employs bootstrap aggregating (bagging) which generates multiple training subsets from the initial dataset putting it among the top-used algorithms because of its basic structure combined with versatility. Model overfitting prevention and diversity maintenance occur through random selection of features for each tree. During creation each tree follows the CART (Classification and Regression Trees) algorithm. When classifying the Gini impurity or entropy serves as the splitting criterion. When predicting the results Mean Squared Error (MSE) becomes the splitting criterion.

Fig 2: Random Forest Architecture (Source: Khan et al., 2021)



Neural Network

The human brain serves as the inspiration for neural networks which represent a particular machine learning algorithm. The tool provides unrivaled capabilities to solve complicated problems which traditional computer algorithms cannot easily resolve including both image recognition and natural language processing (Stanley, et al., 2019). The network architecture groups numerous connected neurons into separate layers. The neural network enables information from other neurons to reach and transform this data before distributing outputs to different neurons. Among neurons connections in this architecture exist weighting elements that reflect how strongly they connect to each other (Stanley, et al., 2019). When networks undergo training procedures, they modify the weight values to enhance their operational efficiency for a specific task. Through their learning procedure neural networks can predict patterns while recognizing diverse architectural applications from image processing to natural language understanding and machine translation design (Stanley, et al., 2019). A neural network consists of:

Input Layer (X): Contains features used for predictions.

Hidden Layers (H): Process and transform data through multiple neurons.

Output Layer (YYY): Produces the final prediction (food price).

Each neuron in a layer computes: $z_j = \sum_{i=1}^n w_{ij} x_i + b_j$ (3.6)

where:

z_j is the neuron's weighted sum,

w_{ij} are the weights,

x_i are the input features,

b_j is the bias term.

The output of each neuron is passed through an activation function:

$$a_j = f(z_j) \quad (3.7)$$

Then data moves forward through the network, layer by layer:

$$a^l = f(W^l a^{l-1} b^l) \quad (3.8)$$

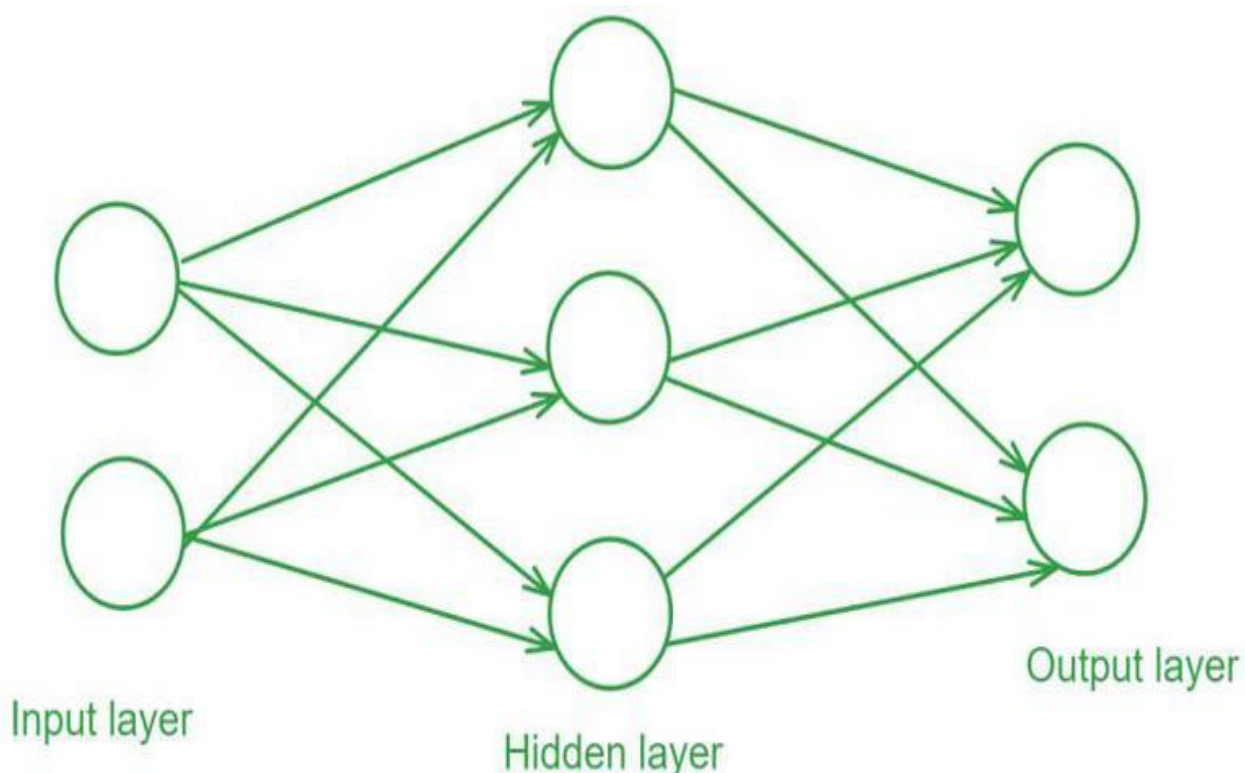
Where

a^l is the activation at layer l

W^l and b^l are the weights and biases at layer l

f is the activation function.

Fig 3: Neural Network Architecture (Source: Khan et al., 2021)



Symbolic Regression

Conducting symbolic regression serves as an artificial intelligence method for finding direct mathematical formulas to model variable linkages in collected data (Makke & Chawla, 2024). The algorithm of symbolic regression differs from traditional regression because it selects the most appropriate functional explanation automatically from the available data. Evolutionary algorithms specifically Genetic Programming enable symbolic regression to find mathematical expressions through repeated cycles of evolution.

Symbolic regression involves:

Generating Random Mathematical Expressions: The initial population consists of randomly generated equations combining variables, constants, and mathematical operators (+, −, ×, ÷, sin, exp, log).

Evaluating Expressions: Each equation is evaluated using a fitness function.

Evolving Better Equations: The best equations undergo selection, crossover, and mutation to produce improved equations in subsequent generations.

Stopping Criteria: The process stops when the model reaches a desired accuracy or after a set number of generations.

Symbolic regression model searches for an equation of the form: $y = f(x_1, x_2, \dots, x_n)$, where f is an unknown function constructed from basic mathematical operations. Afterwards, the adequacy of an equation is assessed using an error or loss function, mostly the mean square error:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.9)$$

Model Evaluation

Regression analysis depends mainly on three performance metrics including Mean Squared Error (MSE) and its Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). Error identification from MSE and RMSE occurs due to their square calculation yet MAE reveals average error amounts in target variable units. The R-squared statistic serves as the last measure because it demonstrates how much the dependent variable can be explained through these attributes while providing the actual overall fit; The Mean Squared Error (MSE) quantifies the average discrepancy between predicted and actual values in regression models. It is computed as the average of the squared deviations between the expected and actual values. The RMSE is the square root of the mean squared error (MSE).

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3.4)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) \quad (3.5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.6)$$

RESULTS

Introduction

The results of the analysis carried out using python are presented below ranging from the descriptive statistics, trend analysis and model evaluation. The descriptive statistic was used to determine the distribution of the food prices and the economic variables, the trend analysis, particularly time plot was used to display the trend of the

food variables and economic variables over the period. While the model evaluation results display the comparison among the machine learning models in accurately predicting the food prices with less error.

Descriptive Statistics

Table 1: Descriptive Analysis on Selected Food Prices in North Central (January 2017 – July 2024)

Measures	Beans NC (₦)	Rice NC (₦)	Garri NC (₦)
Mean	506.9926	470.2926	296.0904
Standard Error	50.14269	34.60932	18.96714
Median	379.18	381.966	256.07
Standard Deviation	478.3307	330.1519	180.935
Minimum	232.26	204.4222	136.63
Maximum	2923.45	1796.88	1112.94

Table 1 depicts the descriptive statistics of the prices of beans, rice and gari between 2017 and July, 2024 for North Central. It is noticed that the mean prices of these food items are ₦506.99, ₦470.29 and ₦196.09 respectively over the period. More so, their standard error of the estimates: ₦50.14, ₦34.61 and ₦18.97 respectively over the period. Furthermore, the minimum prices for each food item are of the means are 232.26, 204.42 and 136.63 respectively while the maximum prices are ₦2923.45, ₦1796.88 and ₦1112.94. The result shows larger disparity between the average value of the price of beans and the maximum price value, indicating the presence of outlier values in the dataset.

Table 2: Descriptive Analysis on Selected Food Prices in South West (January 2017 – July 2024)

Measures	Beans SW (₦)	Rice SW (₦)	Garri SW (₦)
Mean	565.9003	518.7511	305.3929
Standard Error	41.8275	35.86682	24.02279
Median	443.58	395.08	268.92
Standard Deviation	399.0089	342.1477	229.1629
Minimum	263.87	286.42	123.73
Maximum	2539.2	1928.94	1294.51

Table 2 depicts the descriptive distribution of the prices of beans, rice and garri between 2017 and July, 2024 in South West. It is observed that the mean prices of these food items are ₦565.90, ₦518.75 and ₦305.39 respectively over the period. More so, their standard deviations are ₦399.08, ₦342.15 and ₦229.16 respectively over the period. Furthermore, the minimum prices for each food item are 263.87, 286.42 and 123.73 respectively while the maximum prices are ₦2539.2, ₦1928.94 and ₦1294.51. Similarly, the result shows larger disparity between the average value of the price of rice and the maximum price value, indicating the presence of outlier values in the dataset.

Table 3: Descriptive Analysis on Selected Economic Variables (January 2017 – July 2024)

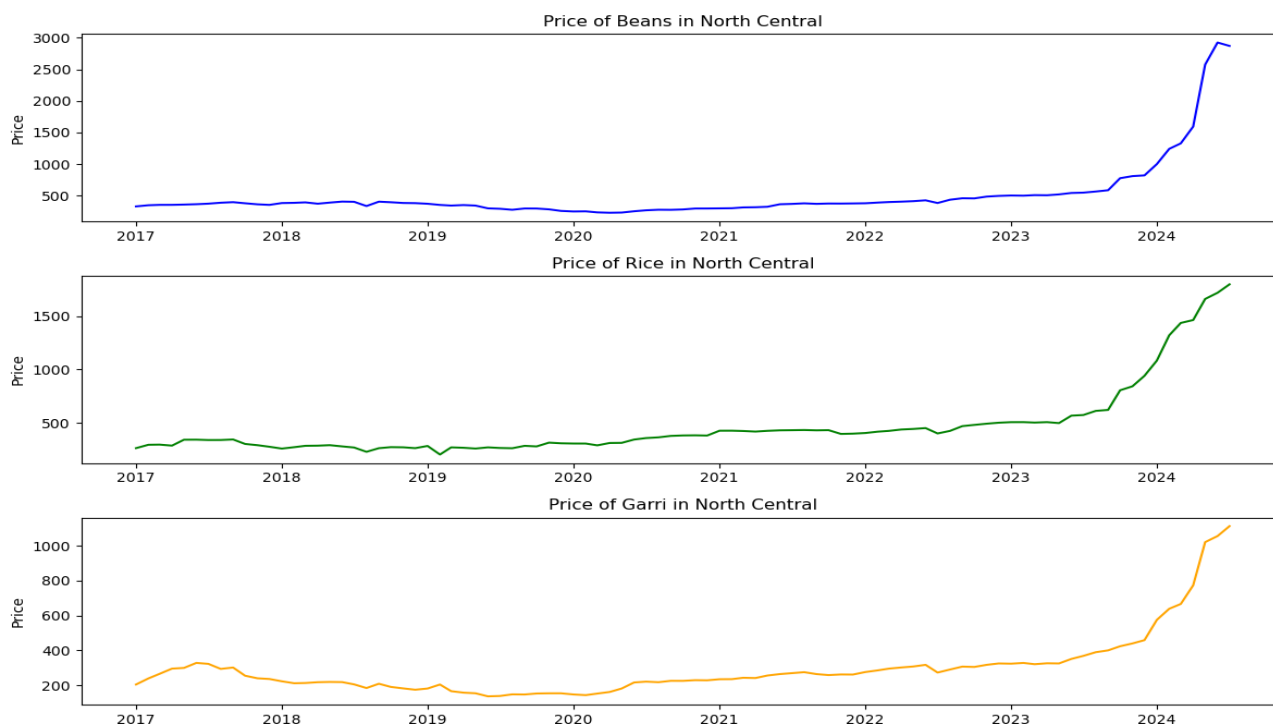
Measures	Inflation Rate	Exchange Rate	Crude Oil (Naira Per Barrel)
Mean	17.50319	474.9571	36152.12
Standard Error	0.653879	32.88521	3181.673
Median	15.99	380.58	22804.54
Standard Deviation	6.237612	313.7049	30351.23
Minimum	11.02	304.25	5140.8
Maximum	34.19	1660	144884.8

More so, the descriptive statistics of the microeconomic variables revealed that the average rate of inflation in Table 3 over this period is 17.50 with standard deviation of 0.653879, the average price of crude oil (in Naira) is ₦36152.12 with standard deviation of 3181.673, and also the average of exchange rate is ₦474.96 with standard deviation of 32.88521. Furthermore, the minimum inflation rate over the period is 11.02 and the maximum inflation rate of 34.19. Also, the minimum value of crude oil is ₦5140.8 while the maximum value is ₦144884.8 and also, the minimum exchange rate value is ₦304.25 and the maximum exchange rate is ₦1660. Furthermore, the trend analysis clearly depicts the price trend of the food items and the economic variables of the period.

Trend Analysis

Time plot displays the trend patterns for the food items as well as the economic variables over the period.

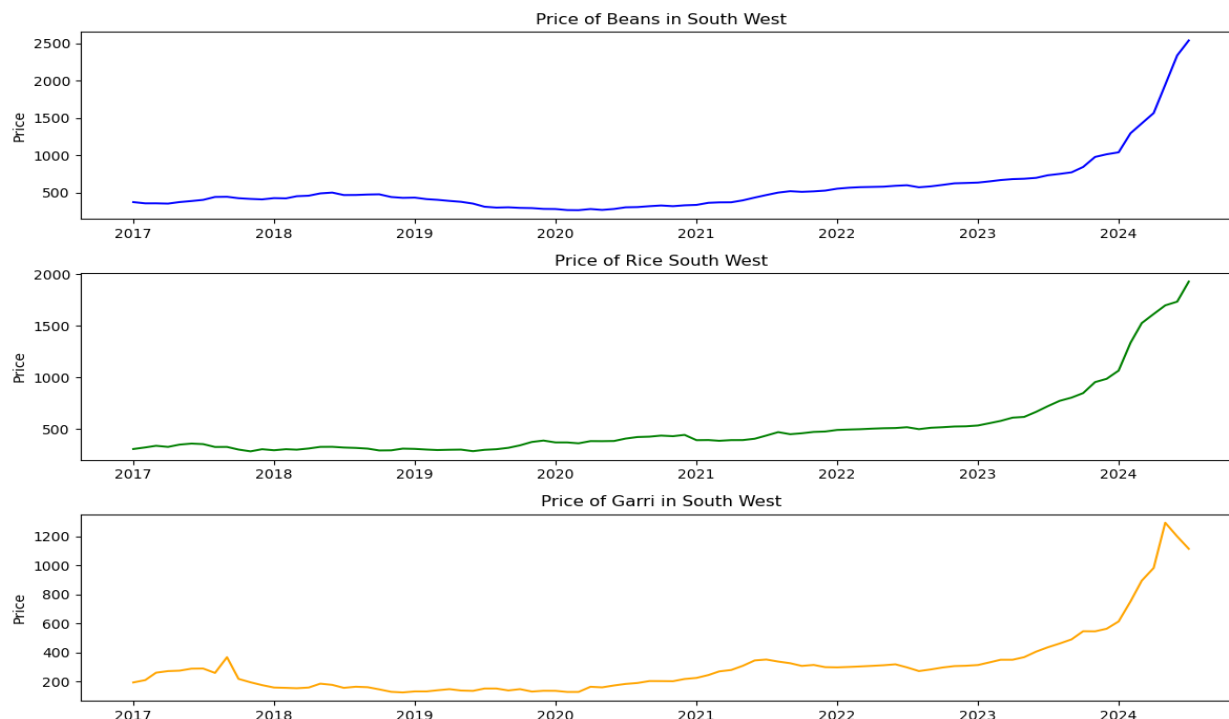
Fig 4: Trend Analysis of Selected Food Prices in North Central



From Figure 4, it is obvious that the price of beans in North central has a stable movement around ₦500 from January 2017 to July 2023 before it experienced a sharp upward movement till 2024 July. Similarly, the price of rice in this region averaged around ₦500 between January 2017 to May 2023 before it experienced a sharp trend

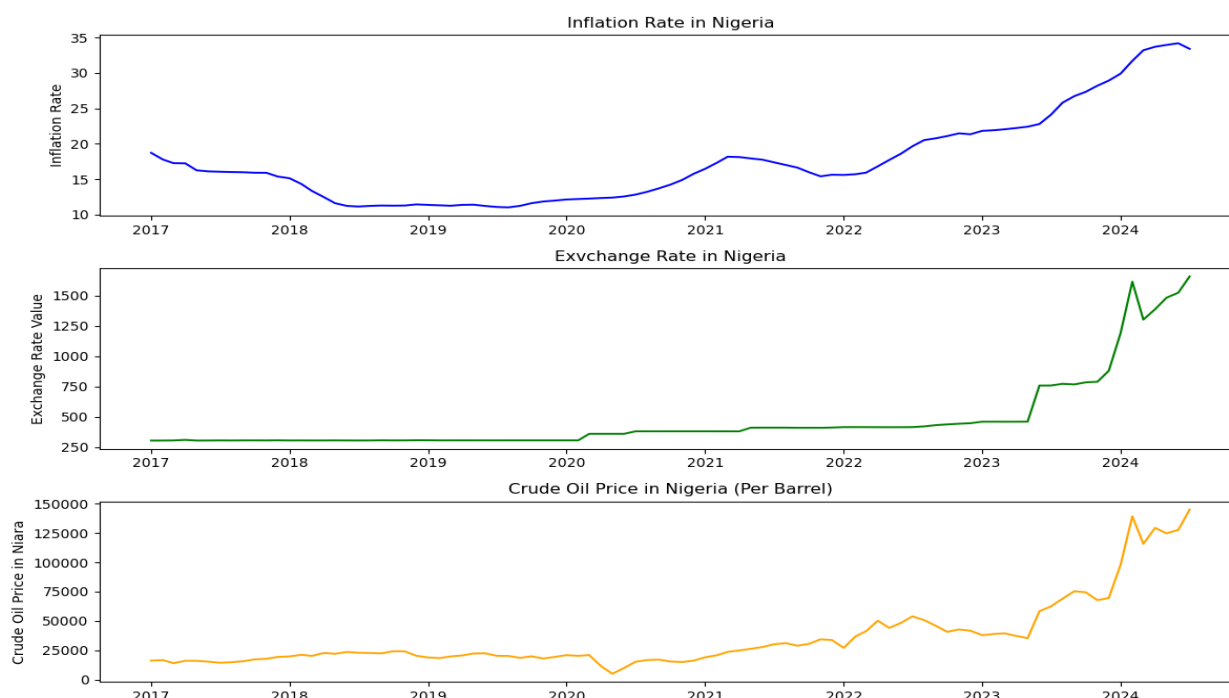
till 2024. More so, the price of garri revolved around the mean of ₦200 between January 2017 and May 2023 and then experienced a sharp upward movement till 2024. These results indicate the presence of structural break in food price as at June 2023. This skyrocket prices could be as a result of the announcement of the removal of fuel subsidy in May by President Bola A. Tinubu.

Fig 5: Trend Analysis of Selected Food Prices in South West



Similarly, Figure 5 depicts the time plot of the food prices in the South Central. It was obvious that the price of rice averaged around ₦500 between 2017 to 2023 May afterwards a sharp upward movement was experienced. Similarly, the price of rice is averaged around ₦500 before a sharp upward movement is experienced. More so, the price of garri is averaged around ₦300 naira between 2017 and 2023 before an upward shift occur in the price of food. This could as be as a result of the removal of subsidy announcement in May 29, 2023.

Fig 6: Trend Analysis of Selected Macro-economic Variables



Similarly, Figure 3 depicts the time plot of the inflation rate from 2017 to July 2024, it is clearly observed that inflation rate was not stable over the years. It declined from 2017 to the mid of 2018 and stable in that position till 2021 before it experienced an upward trend till July 2024. More so, it is observed that the exchange rate experienced a stable movement between 2017 to March 2023 around the average of ₦500 before it later experienced a sharp upward movement till July 2024. Furthermore, it is observed that the crude oil price stable around the average ₦15,000 per barrel between 2017 to 2020, while since then it has been experiencing upward movement till July 2024.

Model Comparison

Afterwards, each food prices in North Central and South West were modelled (univariate analysis) with the proposed four machine learning techniques: random forest, decision tree, neural network and symbolic regression using the economic variables and the lag 1 and lag 5 of the price as indecent variables. Their results of the root means square error and mean absolute error are presented on the following Table.

Table 4: Result and Comparison of Models in Predicting Price of Beans for North Central

Model	RMSE	MAE
Random Forest	1365.41	1131.11
Decision Tree	1348.86	1111.09
Neural Network	672.075	557.69
Symbolic Regression	395.68	287.78

From the result of the model performance carried out on each proposed model. Table 4 depicts that among the models used in predicting the test values of the prices of beans in the North Central, symbolic regression model has the least RMSE value of 395.68 and least MAE of 287.78, followed by the neural network model with RMSE of 672.075 and MAE of 557.69. This results actually showed that the symbolic regression outperformed the other models in predicting the prices of beans for North Central.

Table 5: Result and Comparison of Models in Predicting Price of Rice in North Central

Model	RMSE	MAE
Random Forest	662.19	620.69
Decision Tree	601.74	555.74
Neural Network	1327.951	1228.1407
Symbolic Regression	94.39	85.16

From the result of the model performance carried out on each proposed model. Table 5 depicts that among the models used in predicting the test values of the prices of rice in the North Central, symbolic regression model has the least RMSE value of 94.39 and least MAE of 85.16, followed by the decision tree with RMSE of 601.74 and MAE of 555.74, also followed by the random forest with RMSE of 662.19 and MAE of 620.89. This results actually showed that the symbolic regression outperformed the other models in predicting the prices of rice in the North Central.

Table 6: Result and Comparison of Models in Predicting Price of Gari in North Central

Model	RMSE	MAE
Random Forest	442.57	391.37

Decision Tree	429.08	376.05
Neural Network	920.8771	894.1131
Symbolic Regression	84.28	54.86

From the result of the model performance carried out on each proposed model. Table 6 depicts that among the models used in predicting the test values of the prices of gari in the North Central, symbolic regression model has the least RMSE value of 84.28 and least MAE of 54.86, followed by the Decision Tree with RMSE of 429.08 and MAE of 376.05. This results actually ascertained that the symbolic regression outperformed the other models in predicting the prices of garri in the North Central.

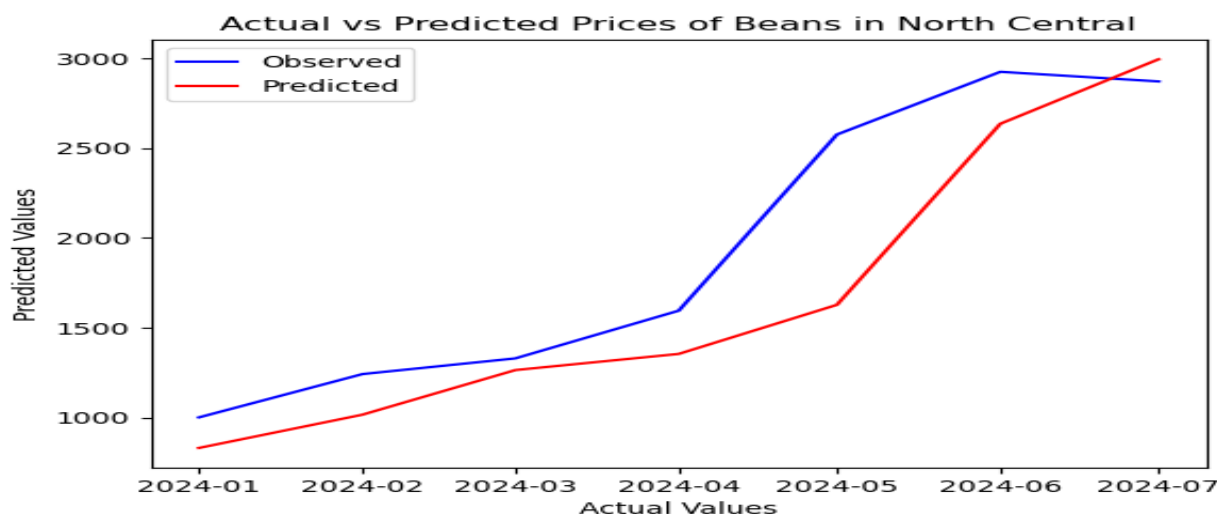
Hence, it is ascertained that symbolic regression consistently outperformed the other models in predicting the prices of rice, beans and gari for North Central. Also, similar results were obtained for the prices of the food stuffs in the South West. The symbolic regression algorithm came up with a mathematical function that best explained the relationship between the economic variables and each food price. Thus, the table below depicts the different mathematical equations generated from the symbolic regression models that predicts each food price.

Table 7: Mathematic Function developed from Symbolic Regression Algorithm in Predicting Food Prices.

S/N	Best Model	Mathematical Function	Model Evaluation
1	Price of Beans (North Central)	$y = lag1 * 108.6110 + 385.0401 - 1038.3701 / (exp(CrudeOilNGN + CrudeOilNGN) - 295.3589)$	RMSE: 395.68 MAE: 287.78 R-squared: 0.73
2	Price of Rice (North Central)	$y = (Exchangerate + (Inflationrate + lag1) * 0.5700 + cos(exp(Exchangerate))) * 80.7200 - (-43.9974) * cos(Exchangerate) + 311.71054$	RMSE: 94.39 MAE: 85.16 R-squared: 0.83
3	Price of Garri (North Central)	$y = Exchangerate * 11.09974 - (-64.6349) * lag1 + 250.5766$	RMSE: 84.28 MAE: 54.86 R-squared: 0.83
4	Price of Beans (South West)	$y = Inflationrate * \left(14.9154 - \frac{0.9962}{lag1 - 2.5548}\right) + lag1 * 143.2763 + 474.4427$	RMSE: 235.26 MAE: 197.42 R-squared: 0.79
5	Price of Rice (South West)	$y = Exchangerate * 23.7343 + lag1 * 127.6259 + 438.5411$	RMSE: 62.82 MAE: 55.71 R-squared: 0.94
6	Price of Garri (South West)	$y = Exchangerate * 26.7912 + lag1 * 81.5568 + exp\left(\frac{0.5781}{lag1}\right) + 246.39738$	RMSE: 119.38 MAE: 94.40 R-squared: 0.72

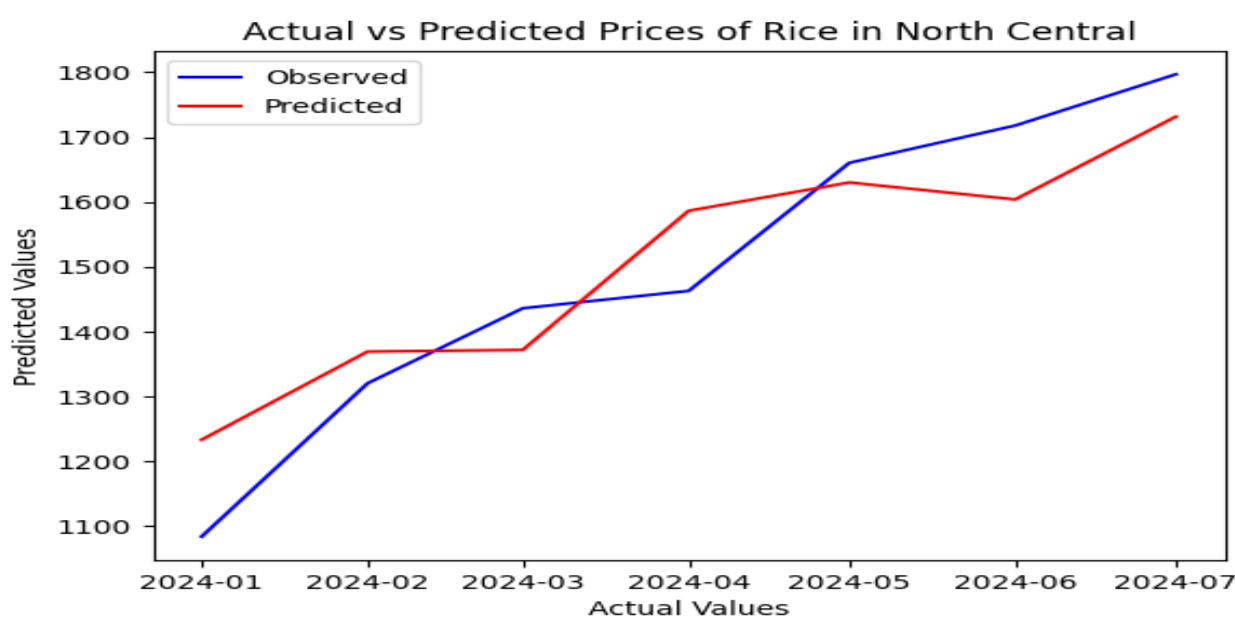
Table 7 presents the mathematic functions derived from the symbolic regression algorithm used in predicting each food prices. It is observed from the table that the price of beans in the North Central was being influenced by the price of last month price and crude oil price according to the symbolic expression. For which this model has the RMSE of 395.68 and MAE of 287.78 and also the R square value is 0.73 indicating that about 73% of the variation in the price of beans in the North Central is being explained by the lag1 of the beans price (last month price of beans) and the price of crude oil. Hence, Figure 4 display the actual and predicted price of beans in North Central from January 2024 to July 2024.

Fig 7: Symbolic Regression Function for Beans Price Prediction in North Central



Furthermore, it is observed from the Table 7, that the symbolic expression for predicting the price of rice in the North Central contains the price of last month value, exchange rate and inflation rate value. The model has the RMSE of 94.39 and MAE of 85.16 and also the R square value is 0.83 indicating that about 83% of the variation in the price of rice in the North Central is being explained by the lag1 of the rice price, the exchange rate and the inflation rate. Therefore, Fig 7 display the actual and predicted price of rice in North Central from January 2024 to July 2024.

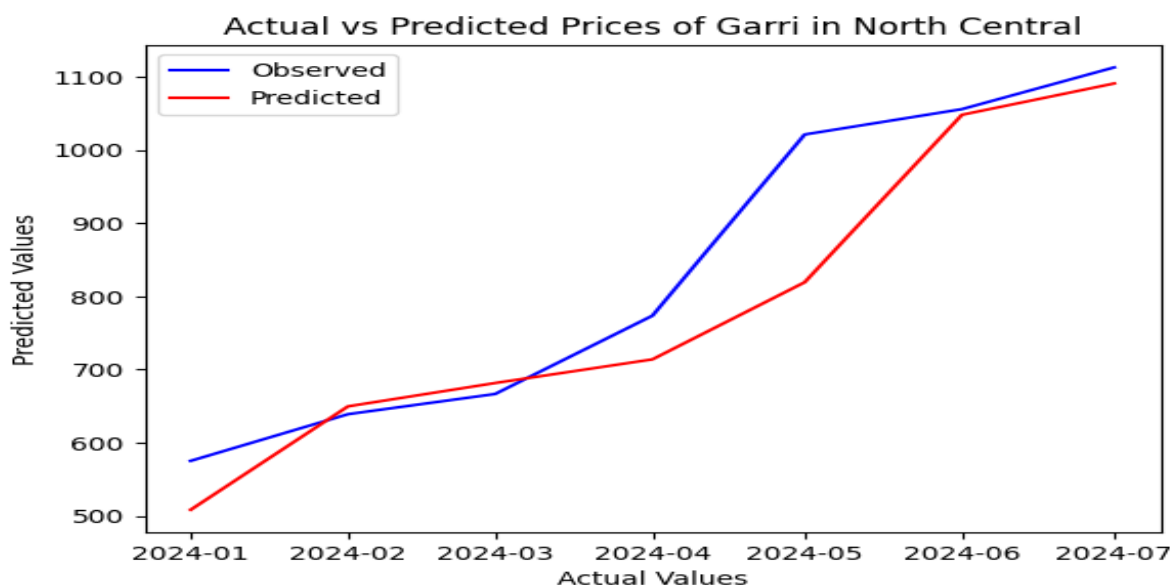
Fig 8: Symbolic Regression Function for Rice Price Prediction in North Central



Furthermore, it is observed from Table 7 that the symbolic expression modelling the price of gari in the North Central contains the price of last month value of gari and the exchange rate. For which this model has the RMSE of 84.28 and MAE of 54.86 and also the R square value is 0.83 indicating that about 83% of the variation in the

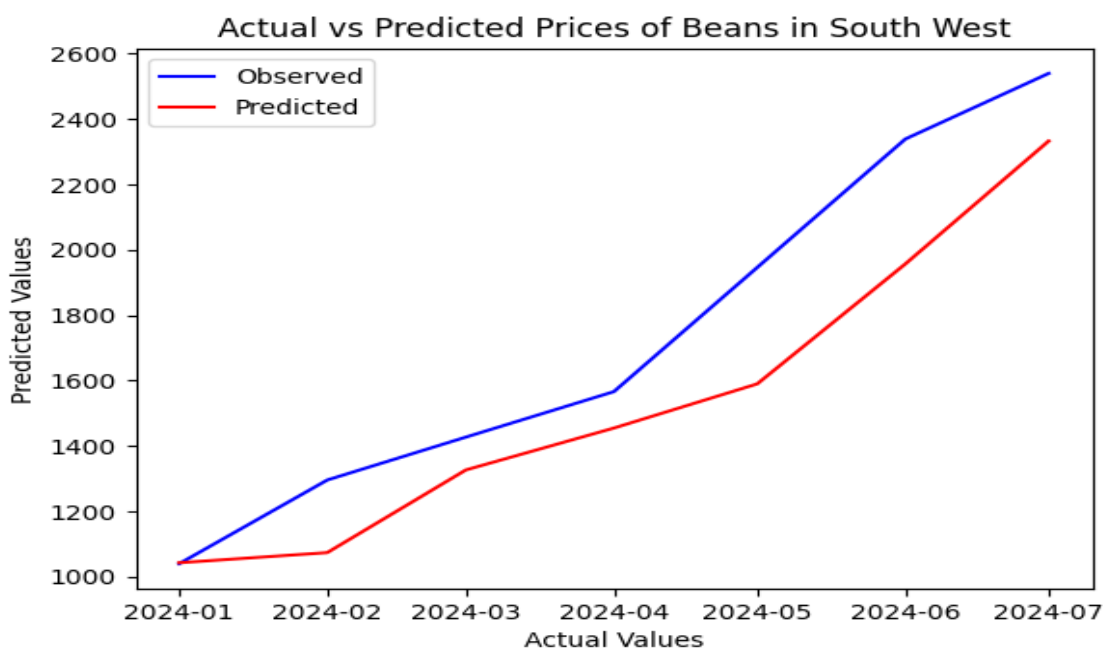
price of gari in the North Central is being explained by the lag1 of the price of gari and the exchange rate. Thus, Fig 8 display the actual and predicted price of gari in North Central from January 2024 to July 2024.

Figure 9: Symbolic Regression Function for Gari Price Prediction in North Central



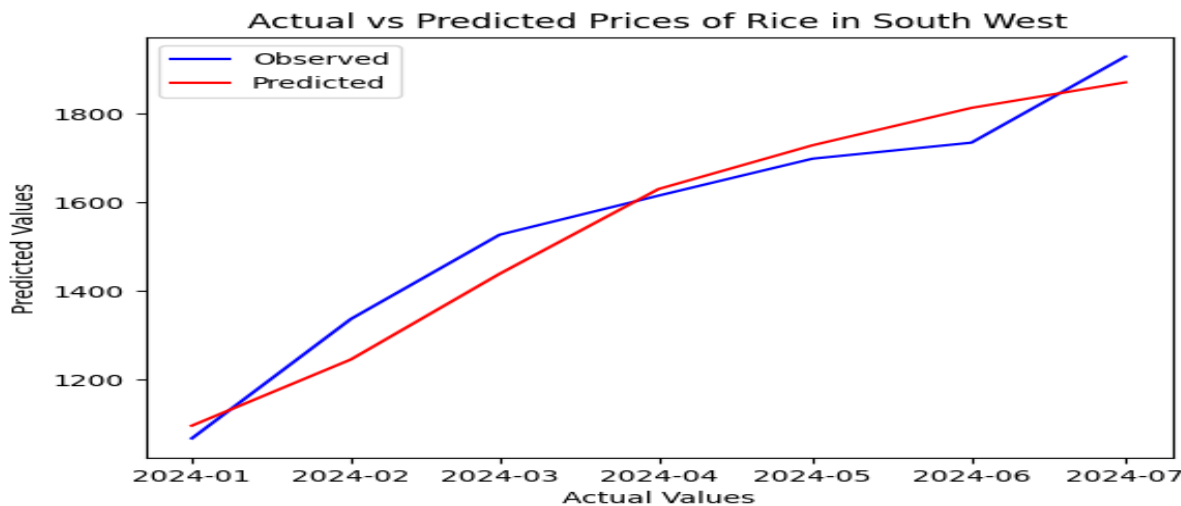
More so, it is observed from the Table 7 that the symbolic expression predicting the price of beans in South West contains lag_1 of beans and inflation rate For which this model has the RMSE of 235.26 and MAE of 197.42 and also the R square value is 0.79 indicating that about 79% of the variation in the price of beans in the South West is being explained by the lag_1 of the beans price and the inflation rate. Thus, Fig 9 displays the actual and predicted price of beans for South West from January 2024 to July 2024.

Fig 10: Symbolic Regression Function for Beans Price Prediction in South West



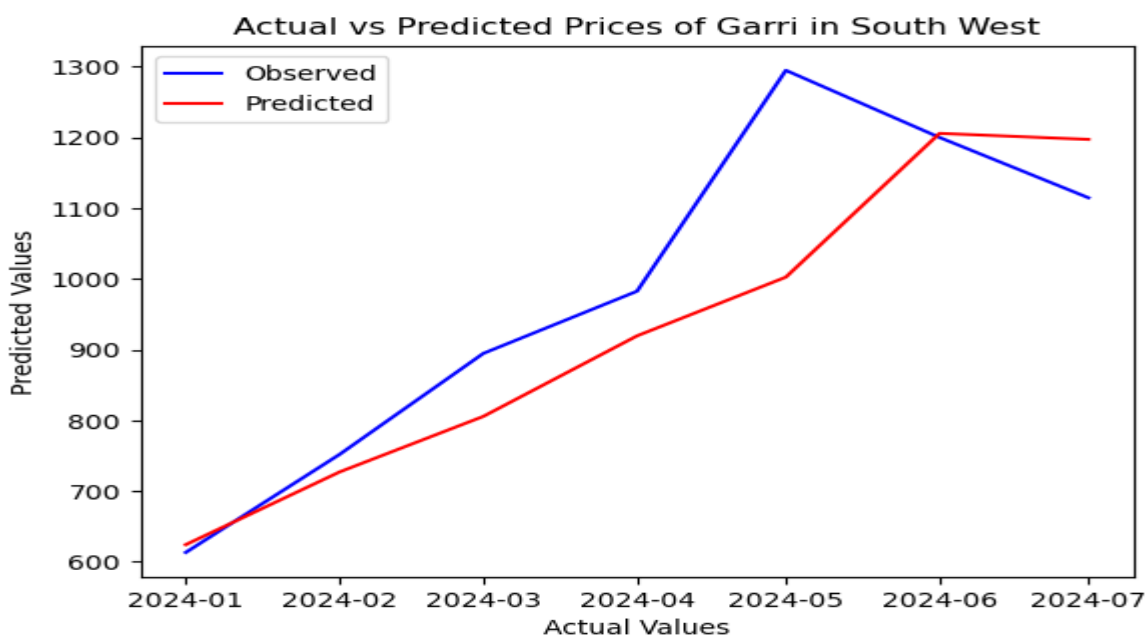
Furthermore, it is observed from Table 7, that the symbolic expression that models the price of rice in the South West contains lag_1 and the exchange rate. For which this model has the RMSE of 62.82 and MAE of 55.71 and also the R square value is 0.94 indicating that about 94% of the variation in the price of rice in the North Central is being explained by the lag1 of the rice price and the exchange rate. Thus, Fig 10 display the actual and predicted price of rice for South West from January 2024 to July 2024.

Fig 11: Symbolic Regression Function for Rice Price Prediction in South West



Furthermore, it is observed from the Table 7, that the symbolic expression that models the price of garri in South West contains the lag 1 and the exchange rate. For which this model has the RMSE of 119.38 and MAE of 90.40 and also the R square value is 0.72 indicating that about 72% of the variation in the price of garri in the South West is being explained by the lag 1 of garri and the exchange rate. Thus, Fig 11 display the actual and predicted price of garri for South West from January 2024 to July 2024.

Figure 12: Symbolic Regression Function for Garri Price Prediction in South West



DISCUSSION

This study findings demonstrated that machine learning achieves exceptional effectiveness when predicting food prices particularly within the South West and North Central regions of Nigeria. The research established symbolic regression as the optimal machine learning mechanism to forecast food prices and expose food price inflating variables throughout the nation. The study confirms Mishra's (2021) findings about machine learning time series forecasting methods that exceed the prediction performance of SARIMA statistical models in food price analysis. The findings of this research study back Wirfelt and Björklund (2024) who analyzed traditional models and XGBoost alongside SARIMA and SARIMAX for Swedish public sector food price forecasting. Machine learning techniques achieved superior predictive results compared to traditional statistical algorithms according to their analysis. The study from Chaparro et al. (2024) demonstrated how small-scale machine

learning tools effectively detect municipalities at risk of food price disruptions by delivering predictions with 79% accuracy points.

The results of model evaluation demonstrate that symbolic regression achieved the most accurate predictions for every food item through its lowest Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) results. Symbolic regression serves as a flexible forecasting method in economic simulations because it discovers optimal mathematical models of food price relationships automatically from input data variables beyond standard hypothesis limitations.

The prediction of food prices needs to include multiple widespread economic indicator variables including inflation metrics alongside exchange rates and crude oil prices statistics. Research from Ajibade et al. (2020) and Shan (2024) received confirmation through the symbolic regression models which established the top position held by macroeconomic variables in forecasting food price volatility. Staple food prices experienced direct effects from both inflation and exchange rates when these economic indicators affected regions that depended on trade-sensitive imports for their food supply. More so, the symbolic regression models showed that previous month price values (lag_1 price) strongly determined food prices in subsequent months. The research indicates that food prices tend to rise in the subsequent month when current prices elevate.

Numerous differences in how prices for food evolve across Nigerian regions emerge as the key finding from the study. The research fills a missing gap in previous works about regional-level forecasting because it studies the price behavior differences between South West and North Central areas (Guo et al., 2022; Chaparro et al., 2024). Statistics show food prices together with forecasting models exhibit unique patterns between different Nigerian regions which supports the need for location-based policy measures and region-tailored prediction tools. The ability of symbolic regression to create separate models for each region demonstrates its conclusion.

Furthermore, the examination of this study proves symbolic regression as a ground-breaking model which provides strong capabilities for forecasting food prices in Nigeria. The adoption of this method by economists together with statisticians and policymakers will help them reach policy decisions through data analysis to stabilize food prices while achieving economic development.

REFERENCES

1. Abang, S. O., Arasomwan, K. O., & Ayodele, O. (2024). Fuel subsidy removal, insecurity, the impact on rising food inflation in Nigeria: A comparative of time series analysis and machine learning techniques.
2. Adeniran, A. O., Azeez, M. I., & Aremu, J. A. 2016. External debt and economic growth in Nigeria: A Vector Auto-Regression (VAR) approach. *International Journal of Management and Commerce Innovations*, 4(1), 706-714.
3. Adeola, A. O., Akingboye, A. S., Ore, O. T., Oluwajana, O. A., Adewole, A. H., Olawade, D. B., and Oguntimele, A. C. (2022). Crude oil exploration in Africa: socio-economic implications, environmental impacts, and mitigation strategies. *Environment Systems and Decisions*, 42(1), 26-50.
4. Ajibade, T. B., Ayinde, O. E., & Abdoulaye, T. (2020). Food price volatility in Nigeria and its driving factors: evidence from garch estimates. *International Journal of Food and Agricultural Economics (IJFAEC)*, 8(4), 367-380.
5. Alade, I.O., Zhang, Y. & Xu, X. (2021). Modeling and prediction of lattice parameters of binary spinel compounds (am₂x₄) using support vector regression with Bayesian optimization", *New Journal of Chemistry*, Vol. 45 No. 34, pp. 15255-15266, doi: 10.1039/d1nj01523k.
6. Andrei, D. M., & Andrei, L. C. (2015). Vector error correction model in explaining the association of some macroeconomic variables in Romania. *Procedia Economics and Finance*, 22, 568-576.
7. Anjoy, P., & Paul, R. K. (2019). Comparative performance of wavelet-based neural network approaches. *Neural Computing and Applications*, 31, 3443-3453.
8. Atsalakis, G. S., & Valavanis, K. P. (2010). Surveying stock market forecasting techniques-Part I: Conventional methods. *Journal of Computational Optimization in Economics and Finance*, 2(1), 45-92.
9. Balogun, E. D. (2025). The short-term effects of gasoline price subsidy removal in Nigeria: an analysis of the economic and social Impacts.
10. Bruinsma, J. (2017). *World agriculture: towards 2015/2030: an FAO study*. Routledge.

11. Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., ... & Crouch, R. C. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of educational psychology*, 102(2), 32
12. Drachal, K. (2021). Forecasting selected energy commodities prices with Bayesian dynamic finite mixtures. *Energy Economics*, 99, 105283.
13. Elsaraiti, M., & Merabet, A. (2021). A comparative analysis of the arima and lstm predictive models and their effectiveness for predicting wind speed. *Energies*, 14(20), 6782.
14. Fasanya, I. O., & Odudu, T. F. (2020). Modeling return and volatility spillovers among food prices in Nigeria. *Journal of Agriculture and Food Research*, 2, 100029.
15. Food and Agricultural Organization. (2016), FAO's Food Price and Output Watch Database. Rome, Italy: Food and Agricultural Organization. Available from: <http://www.fao.org/statistics/databases/en>. [Last accessed on 2019 Mar 16].
16. Gorter, R. R., Eker, H. H., Gorter-Stam, M. A., Abis, G. S., Acharya, A., Ankersmit, M., ... & Bonjer, J. (2016). Diagnosis and management of acute appendicitis. EAES consensus development conference 2015. *Surgical endoscopy*, 30, 4668-4690.
17. Jha, G. K., & Sinha, K. (2013). Agricultural price forecasting using neural network model: An innovative information delivery system. *Agricultural Economics Research Review*, 26(2), 229-239.
18. Jin, B. & Xu, X. (2024). Predicting wholesale edible oil prices through Gaussian process regressions tuned with Bayesian optimization and cross-validation. *Asian Journal of Economics and Banking*.
19. Lama, A., Jha, G. K., Paul, R. K., & Gurung, B. (2015). Modelling and forecasting of price volatility: an application of GARCH and EGARCH models §. *Agricultural Economics Research Review*, 28(1), 73-82.
20. Liang, W., Liu, Y., Somogyi, S., and Anderson, D. P. 2024. A Multi-Model, Ensemble Approach to Forecasting United States Food Prices.
21. Mitra, D., & Paul, R. K. (2017). Hybrid time-series models for forecasting agricultural commodity prices. *Model Assisted Statistics and Applications*, 12(3), 255-264.
22. Mohapatra, P. K., Sahu, B. B., Mohapatra, P. K., & Sahu, B. B. (2022). Importance of rice as human food. Panicle architecture of rice and its relationship with grain filling, 1-25. *Natera*, 2024
23. Ngartera, L., Issaka, M. A., & Nadarajah, S. (2024). Application of Bayesian Neural Networks in Healthcare: Three Case Studies. *Machine Learning and Knowledge Extraction*, 6(4), 2639-2658.
24. Raihan, A., Voumik, L. C., Mohajan, B., Rahman, M. S., & Zaman, M. R. (2023). Economy-energy-environment nexus: the potential of agricultural value-added toward achieving China's dream of carbon neutrality. *Carbon Research*, 2(1), 43.
25. Wang, J., Dharmasena, S. & Bessler, D.A. (2013), "Price dynamics and forecasts of world and China vegetable oil markets". doi: 10.22004/ag.econ.151150
26. Yamauchi, F., & Larson, D. F. (2019). Long-term impacts of an unanticipated spike in food prices on child growth in Indonesia. *World Development*, 113, 330-343.

Ethical Consideration

This research did not involve human or animal participation. The data used was collected from the National Bureau of Statistics through their website.