MACHINE LEARNING-BASED MODELING AND PREDICTION OF FOOD PRICE INFLATION IN ETHIOPIA

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

DEPARTMENT OF STATISTICS

BY

WORKU SUYOUME

ADVISOR: DEREJE DENBE (PhD)

HAWASSA, ETHIOPIA

JUNE, 2025

MACHINE LEARNING-BASED MODELING AND PREDICTION OF FOOD PRICE INFLATION IN ETHIOPIA

Master's Thesis

BY: WORKU SUYOUME

ADVISOR: DEREJE DENBE (PhD)

A THESIS SUBMITTED TO THE DEPARTMENT OF STATISTICS, COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES, SCHOOL OF GRADUATE STUDIES, HAWASSA UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTERS IN APPLIED STATISTICS

HAWASSA, ETHIOPIA

JUNE, 2025

DECLARATION

I hereby declare that this MSc thesis is my original work and has not been presented for a degree in any other university, and all sources of material used for this thesis have been duly acknowledged.

Name: Worku Suyoume

Signature: ___________________ Date: __________________

This MSc thesis has been submitted for examination with my approval as thesis advisor.

Name: DEREJEDENBE (PhD)

Signature: ___________________ Date: __________________

Place and date of submission: __________________________

HAWASSA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

ADVISORS' APPROVAL SHEET

This is to certify that Thesis entitled "**Machine Learning-Based Modeling and Prediction of Food Price Inflation in Ethiopia"** submitted in partial fulfillment of the requirements for the degree of Master's with specialization in Applied Statistics, the graduate program of the department of Statistics, and has been carried out by Worku suyoume, **ID No: GpApStR/0004/15** under our supervision. Therefore, we recommend that the student has fulfilled the requirements and hence hereby can submit the Thesis to the department.

	Signature	Date
Dereje Danbe (PhD)	________________	_______________
Name of Advisor		

EXAMINER'S APPROVAL SHEET-1

HAWASSA UNIVERSITY SCHOOL OF GRADUATE STUDIES

EXAMINERS' APPROVAL SHEET-1

(Submission sheet-2)

We, the undersigned, members of the Board of Examiners of the final open defense by Worku Suyoume have read and evaluated his thesis entitled "**Machine Learning-Based Modeling and Prediction of Food Price Inflation in Ethiopia**" and examined the candidate. This is, therefore, to certify that the thesis has been accepted in partial fulfillment of the requirements for the degree of Masters of Science in Applied Statistics.

	Signature	date
____________________	______________	________________
Name of major Advisor		
______________________	________________	________________
Name of Internal Examiner-1		
__________________	________________	________________
Name of Internal Examiner-2		
________________________	_______________	________________
Name of External Examiner		
________________________	____________________	____________________

SGS Approval

Final approval and acceptance of the thesis is contingent upon the submission of the final copy of the thesis to the School of Graduate Studies (SGS) through the School Graduate Committee (SGC) of the candidate's department.

ACKNOWLEDGMENT

First and foremost, I express my deepest gratitude to the Almighty God for His unwavering grace and guidance throughout this academic journey.

I am deeply indebted to Dr. Dereje Danbe for his exceptional role as my instructor, mentor, and thesis supervisor. His intellectual acumen, unwavering support, and engaging mentorship often infused with insightful humor have been invaluable. I am truly honored to have had the privilege of working under his guidance, and I sincerely appreciate his patience, dedication, and willingness to address every query with clarity. His influence has been a cornerstone of my academic journey at Hawassa University from its very inception.

I also wish to express my heartfelt appreciation to my beloved mother and brother for their unwavering emotional and financial support, as well as their wise counsel. Their unconditional love and steadfast encouragement have been pivotal in helping me navigate the challenges of this journey and realize my aspirations.

A special acknowledgment goes to my esteemed friend, Mr. Teklu Nega (MSc in Applied Statistics, Oda Bultum University), whose generosity in providing critical academic resources significantly facilitated the completion of my thesis.

Furthermore, I am deeply grateful to my peers and colleagues for their camaraderie, motivation, and collaborative spirit, which enriched this academic experience. Their encouragement and shared dedication made this endeavor both meaningful and rewarding.

Finally, I extend my sincere thanks to Hawassa University for fostering an intellectually stimulating and supportive environment, which has been instrumental in my academic and research growth.

ABBREVIATIONS AND ACRONYMS

AR	Auto regressive
ARCH	Autoregressive Conditional Heterosdesckcity
ARIMA	Autoregressive Integrated Moving Average
BVAR	Bayesian Vector Aggressive
FAO	Food and Agricultural Organazation
FPI	Food price inflation
FPO	Future price outcomes
CPI	Consumer Price Index
CSS	Central Statistics Services
CV	Cross Validation
ESS	Ethiopian Statistical Service
GARCH	Generalized Autoregressive Conditional Heterocedasticity
GDP	Gross Domestic Product
IMF	International Monetary Fund
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MSPs	Minimum Support Prices
NBE	National Bank of Ethiopia
RF	Random Forest
RW	Random Walk
RMSE	Root Mean Square Error
SVAR	Structural Vector Aggressive
VAR	Vector Autoregressive model
VECM	Vector Error Correction Model
WB	World Bank

Table of Contents

Contents **Pages**

ACKNOWLEDGMENT v
ABBREVIATIONS AND ACRONYMS vi
List of Figures xi
List of Table xiii
ABSTRACT xiii
CHAPTER ONE 1
1 INTRODUCTION 1
1.1 Background of Study 1
1.2 Statement of the Problem 3
1.3 Objectives of the Study 5
1.3.1 General Objective of the Study 5
1.3.2 Specific Objective of the Study 5
1.4 The Significance of the Study 6
1.5 The scope of the Study 6
1.6 Organization of the Study 6
CHAPTER TWO 7
2 LITERATURE REVIEW 7
2.1 General Definition of Food Price Inflation 7
2.2 Theories on Causes of Inflation (food inflation) 7
2.2.1 Demand Pull Theory 7
2.2.5 Cost Push Theory 7
2.2.3 Quantity Theory of Money 8
2.2.4 Structural Inflation Theory 8

2.2.5 Global Food Price Trends ... 9
2.3 Machine Learning Approach ... 10
2.3.1 Random Forest (RF) Models ... 10
2.3.2 Random Forest Models in forecasting ... 11
2.3.3 Auto-Regressive Random Forest Models in Forecasting ... 11
2.3.4 Advantages of Random Forest (RF) Models in Food Inflation Forecasting ... 12
2.3.5 Empirical Review on Determinant of Ethiopia food price Inflation ... 12
2.3.6 Empirical Review on Food price inflation forecast ... 14
CHAPTER THREE ... 16
3 METHODOLOGY ... 16
3.1 Description of the Study Area ... 16
3.2 Sample size of the study ... 16
3.3 Source of Data and the study variables ... 16
3.3.1 Dependent Variable ... 16
3.3.2 Independent variables ... 17
3.4 Data Preparation and Preprocessing ... 17
3.5 Classical Time Series Models (Benchmark Model) ... 18
3.5.1 Autoregressive (AR) Model ... 18
3.5.2 Autoregressive Integrated Moving Average (ARIMA) Model ... 18
3.5.3 Seasonal Autoregressive Integrated Moving Average (SARIMA) Model ... 19
3.5.4 Testing Stationary: Unit Root Test ... 20
3.5.5 Building ARIMA Models ... 20
3.5.6 Parameters Estimation ... 22
3.5.7 Diagnostic checking ... 23
3.6 Machine learning (ML) approach ... 23

3.6.1 Random Forest Method 25
3.6.2 Random Forest: Regression Tree 26
3.6.3 Random Forest: Bagging 28
3.6.4 Auto-Regressive Random Forest (AR-RF) Model 29
3.6.5 Feature Selection (Variable importance) 30
3.7 Forecasting Strategy 30
3.8 Forecast Evaluation Methods 30
3.8.1 Root Mean Squared Error (RMSE) 30
3.8.2 Mean Absolute Error (MAE) 31
3.8.3 Mean Absolute Percentage Error (MAPE) 31
3.9 Software for Data Analysis 31
CHAPTER FOUR 32
4 RESULTS AND DISCUSSIONS 32
4.1 Exploratory Data Analysis 32
4.2 Data Preprocess 34
4.2.1 Test for Stationarity 35
4.2.2 Splitting of the Data 36
4.3 Forecasting with Traditional Model (Benchmark Model) 36
4.3.1 Forecasting with Autoregressive model 37
4.3.2 Forecasting with an Autoregressive Integrated Moving Average Model Selection Order of ARIMA Models 38
4.3.2 SARIMA Model Selection 40
4.4 Machine Learning Method of Analysis (Proposed Model) 42
4.4.1 Random Forest Method 42
4.4.2 Evolution models 48

4.3 Variables Importance 50
4.4 Model Comparison 53
4.5 Model Diagnostic Auto-Regressive Random Forest and forecasting 54
4.5.1 AR-RF model Diagnostic Checks 54
4.5.2 AR-RF model forecasting 55
4.6 DISCUSSIONS 56
CHAPTER FIVE 60
5 CONCLUSIONS AND RECOMMENDATIONS 60
5.1 CONCLUSIONS 60
5.2 RECOMMENDATIONS 60
Reference i
A. Appendix : Variables Description xii
B. Appendix: Food inflation forecasting using AR model from 2025 – 2029 xiii
C. Appendix : Features Selected by Random forest Models xiii
D. Appendix :Test for number of trees and model performance xiii
E. Selection of number of tree vs. root mean square error xiv
F. Appendix: Features Selected by Auto-Regressive Random forest Models xiv
G. Structure of best fitted RF Model to predict food inflation in Ethiopia xiv
H. Appendix : Model Diagnostic of Auto Regressive Random Forest Model xv
I. Appendix: Forecast Values of Auto Regressive Random Forest model xv
J. Appendix : model comparison in in sample and out sample xvi
K. Appendix : Residuals distribution xvi
L. Appendix : comparison of Bagging with OOB-RMSE xvii

List of Figures

Figure 1: Splitting datasets Algorithm. 17

Figure 2: Random Forest Architecture 27

Figure 3: The structure of the Random Forest describing how predictions are being made by being passed through all trees. 28

Figure 4: Quarterly food price inflation trends in Ethiopia (2000–2024). 33

Figure 5: Trend plots of quarter data of seasonal adjusted food price inflation in Ethiopia (2000-2024).. 33

Figure 6: Correlation of Exogenous Variables for food price inflation in Ethiopia 34

Figure 7: ACF and PACF plot of seasonal adjusted series... 35

Figure 8: Plot of first differenced of food price inflation in Ethiopia.. 36

Figure 9: Plot of ACF and PACF for AR Model 37

Figure 10: ACF and PACF plot of first differenced of training data set 38

Figure 11: Food price inflation forecasting using ARIMA model in Ethiopia (2025-2029)........ 39

Figure 12: ACF and PACF plot of first differenced of Seasonal differenced series 40

Figure 13: Forecasting food price inflation in Ethiopia using SARIMA model for 5 years 41

Figure 14: RF applied to food inflation as the dependent variable. 500 trees and mtry = 15....... 43

Figure 15: Hyper parameter Tuning Results for Random Forest: Effects of Node Size and Predictor Sampling on Model Accuracy 45

Figure 16: Lines plots comparing actual vs. predicted of from Auto Regressive Random Forest model.. 46

Figure 17: Food price inflation forcasting using Random Forest model 47

Figure 18: Forecasting comparison of models used to analyze food price inflation in Ethiopia. 48

Figure 19: Compare the forecasting performance of different models for Ethiopian food inflation from Q1 2020 to Q4 2024 50

Figure 20: The most important variables selected by random forest Models............................... 52

Figure 21: Top 10 Variables importance plot from Auto-Regressive Random Forest (AR-RF) model... 53

Figure 22: Food Inflation Forecasting in Ethiopia using AR-RF model 56

List of Table

Table 1: Summary statistics for the quarter food price (year-on-year) inflation in percentage.... 32

Table 2: Unit Roots test of food Inflation 35

Table 3: Candidate model for AR model 37

Table 4: Candidate model for ARIMA model 39

Table 5: Candidate of SARIMA Models to forecast food inflation.. 41

Table 6: Selection of a hyperparameter (mtry) from original 27 features 42

Table 7: Hyperparameter Tuning Results for Random Forest Model 44

Table 8: Model Performance Comparison without split data 49

Table 9: Top feature selection (variables importance) 51

Table 10: Model Performance Comparison for In-Sample and Out-Sample Forecasting........... 54

ABSTRACT

Food inflation is an essential element of total economic inflation, indicating the pace at which food prices rise over a defined timeframe. This issue has garnered significant attention in its wider economic consequences and international comparison, specifically in Ethiopia. Ethiopia's food price inflation, marked by volatility from interconnected climate, geopolitical and global market risks, has historically been forecasted using traditional models. While accurate *predictions are critical for stabilizing the economy; these methods struggle to capture non-linear dynamics and external shocks. This study advances the field by leveraging a Random Forest model to enhance forecasting accuracy, thereby supporting policies to mitigate food price risks and promote economic stability. Thus, this study aims to model and forecast food price inflation in Ethiopia using a Machine learning method. The data were transformed, standardized and split into training and testing sets to enhance the forecast accuracy of both the machine learning and time series models. The selected models were evaluated based on performance evaluation criteria, including root mean square error, mean absolute error, and mean absolute present error tests. Auto-Regressive Random Forest model outperforming benchmark models with a 55% reduction in root mean square error (7.78 vs. 17.30) and 83.7% explanatory power (R^2). Results reveal self-reinforcing inflationary cycles, with forecasts (2025–2029) indicating sustained volatility (20–30% range) due to primarily linked to monetary factor, international commodity markets and lagged value of food price inflation. The period from 2000 to 2007Q2 exhibited mild and gradual increases, followed by a sudden structural break in 2008Q4, marked by heightened volatility. A sustained upward trajectory emerged from 2015Q1 onward, reflecting persistent inflationary pressures. By combining autoregressive lags with Random Forest's non-linear modeling, this study offers a scalable framework for food inflation forecasting in developing economies, providing actionable insights for policymakers to mitigate food insecurity and macroeconomic instability. The study highlights the dominance of inflation expectations and advocates for policy measures to stabilize agricultural inputs, and strengthen monetary frameworks.*

***Key word**: food inflation, Auto-regressive Random forest model, prediction, machine learning, Ethiopia*

CHAPTER ONE

1 INTRODUCTION

1.1 Background of Study

Food price inflation has surged worldwide due to climate shocks, supply chain disruptions (World Bank, 2023), and geopolitical conflicts, threatening food security for 345 million people (FAO, 2023). Food inflation is an essential element of total economic inflation, indicating the pace at which food prices rise over a defined time frame. It's a subset of general inflation but has particularly significant effects on households, especially those with lower incomes, since a large portion of their budgets goes toward food. This issue has garnered significant attention, especially regarding its wider economic consequences and international comparisons (Irabor et al., 2022).

Inflation at different rates and in other periods has been a problem for every economy, mainly the rampant inflation and its effects worsen the developing economies (Kuma & Gata, 2023). Food price fluctuations undermine food security, which can lead to social unrest, economic instability, and long-term developmental challenges (Mohamed et al., 2024). The concept of food inflation, closely associated with the Consumer Price Index (CPI), encapsulates the average change over time in the prices urban consumers pay for a basket of goods and services, including food. Factors such as unexpected external shocks and exchange rate fluctuations have been identified as significant contributors to higher-than-anticipated food inflation rates (Inoue & Okimoto, 2017).

Furthermore, the global integration of food markets means that domestic food prices are increasingly susceptible to international trends and shocks. This global perspective is essential, as it highlights the impact of international cost and price shocks on national economies. Food inflation has significant implications for the economy of any country and can have far-reaching effects on people's lives and communities, especially for low- and middle-income households in developing countries, who spend a significant portion of their income on food items (FAO, 2011;

Nsabimana & Habimana, 2017). It hurts purchasing power and increases the possibility of food insecurity, which can cause hunger and even increase social and political instability (Purushottam et al., 2024). Consequently, this affects resource allocation and the functioning of pricing systems (Anusha et al., 2022). According to FAO (2024), World food inflation surged in 2022 to 10.4 percent for the food CPI before decreasing to 8.1 percent in 2023, respectively. The food CPI inflation rate in low-income countries increased sharply from 7.3 percent in January 2022 to a peak at 32.1 percent in April 2023 before decreasing rapidly to 16.7 percent in December 2023, indicating that the countries in this income group were much more affected by the shocks in prices and have not yet fully recovered. In 2023, Eastern Africa reported the highest annual inflation rate among food consumer price indices at 14.7 percent, followed by Northern Europe (11.9 percent) and Western Europe (11.4 percent).

Over the past decade, Ethiopia has experienced high and volatile food price inflation, which has had a significant impact on the country's economic and social development (Durevall et al., 2013). This has been driven by a combination of factors, including the impact of global commodity price movements, domestic supply shocks, and the accommodative monetary policy stance. Durevall et al. (2013) inflationary pressures have been particularly pronounced in the food sector, with food inflation consistently exceeding non-food inflation (Demeke & Tenaw, 2021). Understanding the drivers of food price inflation is critical for policymakers in Ethiopia as they seek to maintain macroeconomic stability and promote sustainable economic growth.

National Bank of Ethiopia (2022) indicates that Ethiopia’s food price inflation in February 2022 increased to 41.9 percent, and food inflation constituted 54 percent of the general inflation. This weight highlights the importance of monitoring and predicting food inflation, as it also directly affects the cost of living and the overall welfare of the (Girik Allo et al., 2018). Alem & Söderbom, (2011) indicated that the high food price inflation has been the most divergent economic shock that has sustained to undesirably affect the Ethiopian economy, where a significant proportion of households had to adjust food consumption in response. Over the past decades, food prices in Ethiopia have exhibited high volatility, driven by factors such as fluctuating agricultural productivity, weather patterns, exchange rate instability, global commodity prices, and socio-political factors (Adamu, 2024). This volatility poses challenges to food security, household welfare, and overall economic stability, particularly in a country where millions live in poverty and depend on food for survival. The accurate prediction of future price trends is vital for managing inflation, controlling price volatility, and formulating effective economic policies (Irabor Blessing et al., 2022). Liang, Weifang, Yong & Liu (2024) Farmers, consumers, producers, and legislators can all make well-informed decisions and develop strategic plans when food prices are predicted with accuracy. Ethiopia's persistent food insecurity crisis has entered a dangerous phase, with food price inflation reaching a staggering 33.8% average between 2020-2024 (Ethiopia National Bank, 2024).

In the dynamic and variable economic environment associated with agriculture and global markets, accurate food price forecasts become a key factor for both consumers and producers (Sapakova et al., 2024). While machine learning transforms forecasting globally (Medeiros et al., 2021) Ethiopia's unique challenges data scarcity, climate volatility, and fragmented supply chains demand tailored solutions. Traditional models fail to capture nonlinear shocks like the 2022 41.9% inflation spike National Bank of Ethiopia (2022), creating policy gaps this study addresses through adaptive ML hybrids. While machine learning (ML) models like LSTM and hybrid ARIMA-RF have improved forecasts in developed economies (Zhang et al., 2021), their application in low-income agrarian nations remains limited, leaving critical data-to-policy gaps in vulnerability hotspots.

1.2 Statement of the Problem

Food price inflation is a critical challenge worldwide, exacerbated by climate shocks, supply chain disruptions, and geopolitical conflicts (World Bank, 2023). Food price inflation has increasingly become a significant concern for global stakeholders (Anugrah and Ismaya, 2018). Food price inflation has become a persistent economic and social challenge in Ethiopia, threatening food security, exacerbating poverty, and placing significant pressure on low-income households. As food constitutes a large share of household expenditure particularly in rural and urban poor communities volatility in food prices leads to reduced purchasing power and heightened vulnerability to hunger and malnutrition (FAO, 2023; World Bank, 2022). The Ethiopian economy has experienced frequent food price shocks driven by internal and external factors, including supply chain inefficiencies, currency depreciation, extreme weather events, political instability, and global commodity market fluctuations (IMF, 2023).

In Ethiopia, where a sizable section of the population works in agriculture and food consumption accounts for a significant amount of household expenses, food price inflation is a serious problem (Kuma & Gata, 2023). Recent studies indicate that food prices have risen faster than non-food prices, showing greater variability and less persistence (Demeke & Tenaw, 2021). Effective modeling and forecasting of food price inflation are crucial for informing policy decisions and reducing its adverse effects economy (Seid Nuru Ali, 2022). Ethiopia's food price inflation, marked by volatility from interconnected climate, geopolitical and global market risks, has historically been forecasted using traditional models. It has been a persistent issue, significantly affecting the economy and living standards due to its volatility, driven by both domestic and global factors (ESS, 2024). Several studies have employed traditional econometric and time series models such as VAR, BVAR, ARIMA, ARDL, GARCH, and MGARCH to analyze food price inflation and identify its predictors (Demeke & Tenaw, 2021; Hassen, 2022; Kuma & Gata, 2023; Megersa & Adamu, 2024; Mohammed, 2022). These modeling approaches serve as the cornerstone of predictive analysis in low-data regimes due to their interpretability, well-established theoretical foundations (Hastie et al., 2009), and robustness against overfitting (Crowley, 2023).

Despite the widespread use of traditional time series models in food inflation forecasting, these approaches often fall short in capturing the complex, nonlinear, and dynamic relationships between macroeconomic variables that influence food prices inflation (Hyndman & Athanasopoulos, 2018). Machine learning (ML) methods offer a promising alternative, capable of processing large datasets and uncovering hidden patterns in high-dimensional data without strict assumptions about data distribution (Makridakis et al., 2018). However, their application in the context of food price forecasting in Ethiopia remains limited. Therefore, there is an urgent need to explore the potential of machine learning models in forecasting food price inflation in

Ethiopia. This study addresses these gaps by developing and evaluating machine learning models for predicting food price inflation in Ethiopia. It aims to compare their performance with traditional models and to generate forecasts for the 2025–2029 period, including uncertainty intervals. The goal is to enhance the reliability and usability of inflation forecasts for better policy planning and market responses.

Therefore, this study was intended to answer the following research questions:

I. What was the trend of food price inflation in Ethiopia from 2000 to 2024?
II. Which variables are the most significant predictors of food price inflation in Ethiopia, based on model-based feature importance analysis?
III. How does the forecasting accuracy of machine learning models compare to that of traditional econometric models in predicting food price inflation in Ethiopia?
IV. What are the forecasted values and corresponding uncertainty intervals for food price inflation in Ethiopia over the period 2025 to 2029, based on predictive modeling techniques?

1.3 Objectives of the Study

1.3.1 General Objective of the Study

The general objective of this study was to develop Machine Learning-Based Modeling and Prediction of Food Price Inflation and identify the determinant factor in Ethiopia.

1.3.2 Specific Objective of the Study

Specifically, this study was intended to:

- To examine the trend of food price inflation in Ethiopia from 2020 to 2024.
- To identify the key macroeconomic and external covariates that significantly influence food price inflation in Ethiopia
- To assess and compare the forecasting accuracy of machine learning models versus traditional statistical models in predicting food price inflation in Ethiopia, using historical data over a defined period.

- To forecast food price inflation in Ethiopia for the period 2025–2029, including the estimation of uncertainty intervals to quantify prediction confidence.

1.4 The Significance of the Study

The significance of this study lies in its potential to enhance the understanding of food price inflation dynamics in Ethiopia. Accurate forecasting of food price inflation is critical for economic stability in Ethiopia, yet traditional linear models (e.g., ARIMA, VAR) often fail to capture non-linear patterns and external drivers. This study develops a hybrid Autoregressive-Random Forest (AR-RF) model that synergizes econometric structure with machine learning flexibility, enabling robust integration of multi-source data (e.g., rainfall, exchange rates, global prices) and significantly improving forecast accuracy. The AR-RF model provides policymakers with reliable early warnings for proactive interventions, such as targeted subsidies and strategic reserves release, averting food crises. Governmental agencies can optimize fiscal planning, financial institutions enhance risk management, and agricultural stakeholders improve crop planning. Methodologically, this work pioneers explainable for macroeconomic forecasting in developing economies, offering a scalable template for food-insecure nations

1.5 The scope of the Study

This study evaluates food price inflation across Ethiopia from 2000Q1–2024Q4 using national-level aggregated data. While machine learning hybrids (AR-RF) show promise in capturing non-linearity, sub-regional variations are not analyzed due to data constraints.

1.6 Organization of the Study

This study was divided into five chapters: chapter one deals with introduction that includes the background, statement of the problem, objective of the study, scope and significance of the study. Chapter two addresses review of literature and chapter three focuses methodological aspect of the study which includes: variable definition, model speciation, estimation and diagnostic procedures. Chapter four of the study deals with the results and the discussion. The last chapter, chapter five, deals with the conclusion and policy recommendations of the study.

CHAPTER TWO

2 LITERATURE REVIEW

2.1 General Definition of Food Price Inflation

The term food price inflation denotes the rate of increase in food product prices over a defined time frame, typically measured monthly or annually. It is a crucial element of total inflation, concentrating on food items like grains, dairy products, meat, vegetables, and processed foods. Typically, this increase in prices is monitored through indices such as the Consumer Price Index (CPI) or the Food Price Index (FPI), which observe changes in the cost of a representative selection of food products (FAO, 2023; IMF, 2022). Consumers are impacted by food price inflation as it diminishes their purchasing power and raises living expenses. Food price inflation refers to the sustained increase in food costs over time, directly impacting consumers' purchasing power (Shirotori ,2022; World Bank Food Security, 2025).

2.2 Theories on Causes of Inflation (food inflation)

2.2.1 Demand Pull Theory

Demand-side drivers of food inflation; Population growth, income rise, and preference shifts directly align with demand-pull theory. Food inflation stems from demand-side factors Population growth, rising incomes, and shifting preferences increase food demand. Demand-pull inflation from excess spending power is studied by (Adeosun et al., 2023). Authors employed explicitly link demand-pull inflation to excess spending power. Inflationary pressures, exchange rates, wage rates (Samal et al., 2022), and monetary policies (money supply, bank rates) influence food prices. Government Interventions such as Trade regulations, subsidies, and MSPs (Sharma & Meena, 2024) further shape price volatility.

2.2.5 Cost Push Theory

Cost-push inflation arises when rising production costs such as raw materials and wages drive up prices. Wage increases in key sectors can spread inflation economy-wide, as higher input costs push up prices in downstream industries (Ahmed, 2024; Yekeen & Magaji, 2016). Additionally, monopolistic firms may exacerbate inflation by raising prices to offset costs and boost profits, termed "profit-push" or "administered-price inflation" (Ahmed, 2024). When countries import food or inputs like fertilizers, exchange rate depreciation or global price rises can push up domestic costs (Wiggins et al., 2015).

2.2.3 Quantity Theory of Money

The Quantity Theory of Money, rooted in classical economics, posits that price levels (P) vary proportionally with money supply (M), as captured by Fisher's equation:

$$MV = PQ$$

Where: M = money supply, V = velocity of money (considered constant), P = price level, Q = real output (considered stable in the short run). According to the theory, inflation is mainly a monetary phenomenon caused by excessive money growth (Alex and Stefan, 2003). Pioneering economists such as David Ricardo contended that alterations in price must reflect changes in the money supply. Later on, monetarists embraced this perspective, highlighting the importance of central bank control over the money supply to reduce inflation as a counterpoint to Keynesian fiscal policies. The central tenet of the theory is that, with stable V and Q, inflation arises from a monetary expansion that is not kept in check.

2.2.4 Structural Inflation Theory

At the core of structuralist inflation theory lies the identity relation, which asserts that production costs and output prices are equivalent. This model posits that final product prices are influenced by import costs, social conflict, and technology (which dictates the labor and input requirements per production unit) (Taylor & Barbosa-Filho, 2021). This study follows several others that employed econometric methods, statistical techniques, and machine learning algorithms to model and predict inflation. The latter models are used in cases of non-linear and non-stationary inflation, whereas the former models are applied to linear and stationary inflation (Rita et al., 2019)

2.2.5 Global Food Price Trends

World food price tends to have a positive impact on domestic food price inflation in this finding and that was in line with the finding of (Demeke & Tenaw, 2021). This reflects that food price rises in the local market have coincided with food price hikes in the international market. Besides to this the high dependency ratio raises the inflation of the country like other findings (Nguyen & Jemma Dridi, 2019). Although the effect of soaring world food prices will largely depend on the demand pattern of households' responsiveness to price change and possible substitution effects, most previous studies failed to account for characteristics of a household demand system. These studies implicitly assume uniform price-effects across regions or provinces within countries (Brinkman et al., 2010). The most recent improvement in the index's price was the fifth in a row. The FAO price chart of vegetable oil recorded 121.9 figures in November, up 15.4 figures (or 14.5 percent) from the prior month and reaching its highest possible level since March 2014. The FAO Sugar inflation rate reached an 87.5 monthly increase. Stronger new data confirming previous forecasts of global production shortages in the 2020/21 trading season aided the rise in international sugar quotations in November (FAO, 2020). The Russo-Ukrainian War and the resulting economic impact on economies around the globe have been significant according to emerging studies (Liadze & Macchiarelli, 2022). Evolving evidence points to signifcant distortions in global stock market performance, volatile crude oil prices, which have led to price hikes in global fuel prices, and upward pressure on the general price of goods due to the constrained supply of critical global commodities. Although the Russo-Ukrainian War has undoubtedly had a pervasive negative impact on the global economy already reeling from the effect of the COVID-19 pandemic, there is a growing consensus that the war is having a relatively more devastating impact on less developed and emerging economies such as those in sub-Sahara Africa than developed economies. World bodies, such as the World Bank and the International Monetary Fund (IMF), for instance, continue to envisage a significant devastating impact of the war on food shortages among most developing, and less developed economies such as those in sub-Sahara Africa. Although the Ethiopian food price index seems to share the same increasing trend as the world price index (Brinkman et al., 2010). It is obvious that since August 2004 the Ethiopian food price index has been consistently higher than the world index. This suggests the existence of significant local causes of the food price surge in Ethiopia, especially the severe drought shock of 2002–2003. The world oil price seems to play a major role in the food price hike in Ethiopia

2.3 Machine Learning Approach

Often referred to as the art and science of pattern recognition, machine learning is a field within artificial intelligence. It is fundamentally a data-driven approach that makes mild assumptions about the underlying statistical relationships in the data and encompasses a wide range of methods. It generally comprises two main components: a learning method and an algorithm. This allows for the automation of as many modeling choices as possible, independent of the forecaster's discretion (Olivera et al., 2024). In the last two decades, machine learning methods have become well-known and established themselves as plausible substitutes for conventional statistical models in the area of forecasting. Time series data frequently display non-linear traits, making it essential to employ more advanced methods for precise predictions (Kwame Appiah, 2009). Due to the unique trends identified in these datasets, the use of conventional linear models for forecasting has become more difficult, resulting in a rapid rise in the application of machine learning for inflation prediction. Recent significant research employing machine learning for forecasting inflation comprises studies by (Chakraborty & Joseph, 2017; Maehashi and Shintani 2020), among others.

2.3.1 Random Forest (RF) Models

Time series data are not always linear and requires more complex approaches for data forecasting (Kwame, 2009). Medeiros et al., (2021) demonstrated that ML models incorporating a large number of covariates outperformed traditional time-series models in predicting U.S. inflation. According to the author, the RF model has gained importance in inflation forecasting due to its exceptional performance. Similar advancements have been observed in GDP forecasting. (Yoon, 2021) encouraged the use of ML models in macroeconomic forecasting. They explored RF and gradient boosting to forecast GDP growth in Japan. The use of machine learning method Random forest in inflation forecasting has been expanding quickly due to the distinct patterns in these series, which make predicting with linear classical models challenging.

2.3.2 Random Forest Models in forecasting

One such ML algorithm that has demonstrated success in numerous previous high-impact weather forecasting applications (Herman & Schumacher, 2018; Williams et al. 2008) is the random forest model. This study seeks to apply RF methodology to the generation of calibrated probabilistic CONUS-wide forecasts of severe weather with predict analogous to those of SPC convective outlooks in the hope that the guidance produced can be used to improve operational severe weather forecasting. A relevant summary of the methodology of (Herman & Schumacher, 2018) necessary for proper understanding of the methods employed in this study is provided here, but for more detailed explanations of the mathematical underpinnings of RFs and the numerous sensitivity experiments performed therein, the reader is invited to consult those studies. For brevity, several of the RF model configuration choices selected in this study are motivated by the findings of (Herman & Schumacher, 2018) rather than performing all the same sensitivity experiments for this forecast problem (e.g., input predictor variables).

2.3.3 Auto-Regressive Random Forest Models in Forecasting

In their Goulet Coulombe (2020) paper**,** Auto-Regressive Random Forest as an important adaptation of the basic RF model**.** This model is noted as being particularly well suited to predict measures of inflation**,** among many other macroeconomic indicators. Coulomb notes that the key difference between ARRF and RF is the inclusion of a linear part within each tree leaf instead of just an intercept. This is only one of the many regression trees that can be used to explain the change in food price inflation. The structure, variables used for dividing the independent variables, and the values of the splitting points must be determined. Iterative local updates are made to the trees known as greedy algorithms (Segal, 2014) that help optimize these factors while remaining computationally feasible. ARRF's theoretical superiority stems from reconciling econometrics (linear dependencies) with ML's nonlinear pattern recognition (Goulet Coulombe, 2020). This hybrid vigor is critical for Ethiopia, where structural breaks (e.g., 2008 crisis) violate ARIMA's linearity assumptions

2.3.4 Advantages of Random Forest (RF) Models in Food Inflation Forecasting

Random Forest (RF) models offer superior handling of non-linear relationships by capturing complex interactions between drivers through decision tree ensembles (Breiman, 2001), which reduced forecasting errors by 28% during Ethiopia's 2022 inflation surge (Aragie et al., 2023). They also demonstrate robustness to data limitations: RF handles missing values via surrogate splits (Liaw & Wiener, 2002). feature critical for patchy Ethiopian data, and processes over 15 variables without dimensionality reduction (K. M. Hastie et al., 2021). Furthermore, RF provides interpretable policy insights by quantifying driver impacts (e.g., fertilizer costs being more than three times as influential as fuel prices) using Mean Decrease Gini (Strobl et al., 2008) and visualizing intervention impacts. In terms of computational efficiency, RF trains on Ethiopia's 20-year data in under two minutes, compared to LSTM's 15 minutes or more (Pedregosa Fabian, Ga¨el Varoquaux, 2011), and handles high-frequency data from 770 markets (World Bank, 2023). Resistance to overfitting is achieved by voting across 500 or more trees to prevent spurious correlations (Breiman, 2001) and self-validation through out-of-bag samples (Cutler et al., 2012). Additionally, RF natively processes both numerical (such as inflation rates) and categorical (like drought alerts) variables (K. M. Hastie et al., 2021), Finally, RF exhibits hybridization flexibility by combining its feature power with ARIMA's temporal structure (Zhang et al., 2021)

2.3.5 Empirical Review on Determinant of Ethiopia food price Inflation

A number of studies have tried to find the main determinants of food price inflation. For instance**,** Kuma & Gata (2023) have carried out research on what causes food price inflation in a Post-Communist Economy using Evidence from Ethiopia. Authors employed the ARDL model to reveal that broad money supply, government budget deficit, exchange rate, consumer food price index, average annual rainfall, population number, real GDP, interest rate, and world food price are found to be the major determinants of Ethiopia's food price inflation. The long run result revealed that real GDP and world food price at lag one were negatively and significantly affected. In contrast, domestic food price, annual rainfall, interest rate, and money supply affected food price inflation positively and significantly. The study conducted by Hilegebrial, (2015) used OLS estimates to analyze Ethiopia's food inflation, finding that both demand-side factors (money supply, expectations) and supply-side factors (output, global food prices) influence it. Similarly, (Sharma & Meena, 2024) found that in India, food inflation rises with wage rates, agricultural trade, global food prices, and exchange rates but falls with Agri-GDP and interest rates. Their NARDL model also revealed long-run asymmetric effects of money supply, wages, oil prices, global food prices, exchange rates, and interest rates on food inflation. Qayyum & Sultana's (2018) study identifies key predictors of food prices in Pakistan using an ARDL model, including CPI (food inflation proxy), GDP growth, food exports/imports, agricultural credit, and taxes.

The study aims to build a forecasting model for Ethiopia, incorporating macroeconomic and agricultural (Durevall & Loening, 2009; Demeke & Tenaw, 2021). Fuel prices also significantly impact food inflation by raising transport costs (De Gregorio et al., 2007). Focused studies on current events provide timely insights into how specific geopolitical events can influence food inflation. Industry-specific research, like the study by Shokrani et al. (2016), offers detailed insights into specific sectors affected by food inflation. The global supply chain pressures and international trade dynamics also play a critical role in shaping food inflation trends. Di Giovanni et al. (2022) analyze the impact of the COVID-19 pandemic on Euro Area inflation and compare it with the experiences of other countries.

Anam (2014) this topic describes the impact of food export, food import, population, money supply & per capita income on food price inflation. Macroeconomic factors such as inflationary pressures, exchange rate fluctuations, wage rates (Samal et al., 2022), money supply, and bank rates (Hordijk et al., 2020) indirectly impact food prices. Moreover, Government policies, such as trade regulations, import/export duties, subsidies, and minimum support prices (MSPs) also affect food price (Sharma & Meena, 2024). Focused studies on current events provide timely insights into how specific geopolitical events can influence food inflation. Industry-specific research, like the study by Ren and Huatuco (2016), offers detailed insights into specific sectors affected by food inflation.

2.3.6 Empirical Review on Food price inflation forecast

The importance of empirical research on forecasting food price inflation has increased because of the crucial influence that food prices have on economic stability, poverty levels, and the effectiveness of policies, particularly in developing economies. A large body of empirical literature uses traditional time series models such as ARIMA (Autoregressive Integrated Moving Average), VAR (Vector Autoregression), and their variants to forecast food price inflation. Studies like those by (Baffes & Haniotis, 2010) and (Nazlioglu & Soytas, 2013) have used ARIMA models to forecast short-term food price movements.

Research by (Steve Hatfield-Dodds, 2017) and (Cheng & Cao, 2019) employed VAR and cointegrated VAR (VECM) models to examine the dynamic relationships between food prices and macroeconomic variables. Their findings suggest significant interdependencies and lagged effects, which are critical for understanding inflation transmission mechanisms. (Naveena et al., 2017) used a hybrid ARIMA-ANN model to forecast food prices in India, finding that the hybrid approach outperformed standalone models in both accuracy and robustness. Traditional models have been widely applied to food inflation forecasting (Gómez et al., 2012; MacLachlan et al., 2022), their linear assumptions often fail to capture structural breaks in volatile economies like Ethiopia (Demeke & Tenaw, 2021). Recent advances in machine learning, particularly Random Forests Medeiros et al., 2021), and AR-RF hybrids Goulet Coulombe (2020), address these limitations by modeling nonlinear interactions a critical gap given Ethiopia's dependency on exogenous shocks (Kuma & Gata, 2023).

MacLachlan et al., (2022) proposed an optimized SARIMA model for FPO forecasts, implemented in July 202. Goulet Coulombe et al., (2022), propose using the Auto-Regressive Random Forest (ARRF) model to enhance Food Price Outlook (FPO) forecasting. The study aims to improve accuracy while preserving the strengths of SARIMA models, such as standardized cross-category comparisons and uncertainty measures, while incorporating exogenous variables and capturing nonlinear dynamics. Iddrisu & Alagidede (2021) used quantile regression in South Africa, finding that monetary policy consistently affects food inflation across all quantiles, with tighter policies worsening rising food prices. Joutz (1997) compared U.S. retail food price forecasts from various institutions using Delphi, structural, and ARIMA models. Gabriel Moser (2004) found that for Austria, factor models best predict unprocessed food inflation, while a combination of factor and VAR models performs best for processed food. (Gómez et al., 2012) use forecast combinations and flexible least squares techniques to analyze disaggregated data for food goods in Colombia.

Using basic univariate techniques and a large dataset of online food and non-alcoholic beverage pricing in Poland (Strasser et al., 2022), conduct a real-time now-casting experiment. Forecasting inflation with the Random Forest model is an emerging field with a limited number of available papers. However, some articles that predict other macroeconomic variables, like GDP growth, use similar methods and input variables. The standard practice in these articles, when comparing the performance of RF, is to use univariate models as (Araujo & Gaglianone, 2023). Kaewchada et al.(2023), in this study the random forest model to forecast vegetable prices. The prediction results showed that the random forest model was capable of accurately forecasting vegetable prices for pumpkin, eggplant, and lentils.

Nowadays, where agricultural and food markets are becoming increasingly dynamic and subject to various influences, food price forecasting has become an important tool for farmers, consumers and governments. The application of machine learning approaches to this task has attracted increasing attention and research (Sapakova et al., 2024). Numerous scholars examining food price dynamics in the existing body of research incorporate a diverse array of factors into their investigations. In recent years, advanced methodologies such as Autoregressive Distributed Lag (ARDL), causality and cointegration analysis (as proposed by Granger and Toda-Yamamoto), Dynamic Conditional Correlation (DCC), Generalized Method of Moments (GMM), Nonlinear Autoregressive Distributed Lag (NARDL), Generalized Autoregressive Conditional Heteroscedasticity (GARCH), Panel Data Analysis, and various regression techniques (including Ordinary Least Squares (OLS) and pooled-OLS).

CHAPTER THREE

3 METHODOLOGY

3.1 Description of the Study Area

Ethiopia, located in Eastern Africa, spans 122.2 million hectares with a central high plateau ranging from 1,500 to 3,000 meters above sea level. It consists of 12 regional states and two chartered cities (Addis Ababa and Dire Dawa). The country has vast agricultural potential with a favorable climate and soils for crop production. In 2024, Ethiopia's population is estimated at 129.7 million, with over 85% living in rural areas, largely dependent on agriculture, particularly cereals and oilseeds, for their livelihoods.

3.2 Sample size of the study

The time series data utilized in this study comprises quarterly observations spanning from the first quarter of 2000 to the fourth quarter of 2024, encompassing 25 year. With each year consisting of four quarters, the dataset includes independent and dependent variable observations for analysis.

3.3 Source of Data and the study variables

This study utilized secondary data on food price inflation and related factors, spanning from January 2000 to December 2024, sourced Central Statistical Service (CSS) of Ethiopia, the National Bank of Ethiopia (NBE), the World Bank, the Food and Agriculture Organization (FAO), and Ethiopian National Meteorological Agency (ENMA).

3.3.1 Dependent Variable

The dependent variable of this study was food price inflation (FPI). Food price inflation refers to the rate at which the prices of food items increase over time, which is measured by CPI (Kuma & Gata, 2023). The most common measure of food inflation is the percentage change in the food consumer price index (CPI), which captures the cost of living of the average consumer. The CPI includes also domestically produced and imported consumer goods and services.

$$\text{Food Inflation (\%)} = \frac{\text{CPI}_{\text{for food (current period)}} - \text{CPI}_{\text{for food (base period)}}}{\text{CPI}_{\text{for food (base period)}}} * 100 \qquad (3.1)$$

3.3.2 Independent variables

The independent variables included in the study was non-food inflation, headline inflation, Lagged values of food inflation, Broad money supply, narrow money supply, exchange rate, price of imported good and service, price of exported good and service, agricultural production, area cultivated, government expenditure, real GDP, Total population, lending interest rate, unemployment rate, rainfall, world oil price, world food price, food price index, fertilizer, foreign direct investment, tax Revenue, net demand deposit, Agricultural allied activity, gross private consumption, political stability index (See Table Appendix A).

3.4 Data Preparation and Preprocessing

Data preparation is essential before analysis. This involves cleaning the data by impute missing values or incorrect entries, ensuring proper time formatting, reorganizing it for analysis, and splitting it into training and testing sets. To reduce the data set's intrinsic volatility, this could pose challenges for conventional linear models. We have used transformation techniques to convert the monthly food price inflation data to quarterly data. The same is true for other variables. In a 4:1 ratio, the data frame has been split into training and testing datasets (80% for training datasets and 20% for testing datasets). The model was developed using the training data, while the test data was utilized for prediction to assess accuracy (Joseph & Vakayil, 2021).

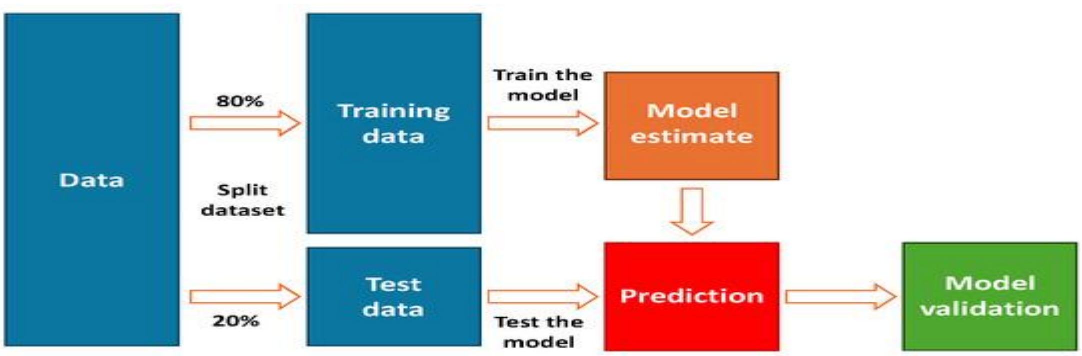


Figure 1: Splitting datasets Algorithm.

3.5 Classical Time Series Models (Benchmark Model)

Exploratory Data Analysis

One type of study that finds broad patterns in the data is called exploratory data analysis. It is an important stage of any data analysis process. Knowing the locations of outliers and the relationships between variables can help shape statistical analyses that yield insightful results. Descriptive statistics, time series plots, and correlation matrices can all be used to illustrate the patterns in data (Sharma, et al., 2007; Andersen et al., 2008). Time series analysis examines data collected at fixed intervals, identifying patterns over time.

3.5.1 Autoregressive (AR) Model

The Autoregressive (AR) model is one of the most common time series models. It assumes that the current value of the variable is a linear function of its previous values (lags). This model is useful when there is autocorrelation in the data (i.e., past values influence future values). The general form of an AR (p) model, where p denotes the number of lag periods of food inflation considered, is as follows:

$$FPI_t = \alpha_o + \alpha_1 FPI_{t-1} + \alpha_2 FPI_{t-2} + \cdots + \alpha_p FPI_{t-p} + \epsilon_t \qquad (3.2)$$

$where,\ FPI_t$ is the food price inflation at time, α_o is a constant (intercept) term, $\alpha_1, \alpha_2 \ldots\ldots \alpha_p$ are the coefficients for the lags and ϵ_t is the error term (white noise or residual), which is assumed to be independent and identically distributed with mean zero and constant variance.

3.5.2 Autoregressive Integrated Moving Average (ARIMA) Model

The statistician Whittle (1951) first presented the autoregressive moving average (ARMA) model in the 1950s as a technique for evaluating and predicting time series data. In addition to being the most used benchmark for random forest model (Gonzalez, 2000), the model is most frequently employed in forecasting univariat time series data analysis (Hyndman & Athanasopoulos, 2018). The Autoregressive Moving Average (ARMA) model can also be extended to accommodate non-stationary data with ARIMA. Prevailing ARIMA-based studies

(Bekele et al., 2021) maintain three fundamental weaknesses those are linear assumptions that cannot capture Ethiopia's volatile inflation regimes, Exclusion of vital exogenous variables (rainfall anomalies, global fuel price shocks) and rigid structures that fail to adapt to sudden market disruptions

The ARIMA model's general form can be rewritten as:

$$\Delta^d FPI_t = \sum_{i=1}^{p} \beta_i \Delta^d FPI_{t-1} + \sum_{j=1}^{q} \theta_i \epsilon_{t-j} + \epsilon_t \qquad (3.3)$$

$\Delta^d FPI_t = FPI_t - FPI_{t-d}$ where $\Delta^d FPI_t$ is change or the difference between food price inflation (FPI_t) at time t lag of food price inflation (FPI_{t-d}) at time (t-d) and d is order of integration, $FPI_{t-1,}$ FPI_{t-2}, $FPI_{t-3} \dots . FPI_{t-p}$ were previous lags of food price inflation, ϵ_t is error term at time t , assumed to be independently and identically distributed, p and q are the order of AR and MA; $\beta_1, \beta_2, \beta_3 \dots \beta_p$ coefficient of autoregressive model and $\theta_1, \theta_2, \theta_3 \dots \theta_q$ coefficient of a moving average model. The choice of an ARIMA model assumes a linear relationship between the time series and its past observations and only allows for uni-variate time-series data.

3.5.3 Seasonal Autoregressive Integrated Moving Average (SARIMA) Model

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model is an extension of the ARIMA model that explicitly supports seasonal components in time series data, making them ideal for forecasting time series with recurring patterns, such as food price inflation (Brockwell & Davis, 2016). SARIMA is often written as:

SARIMA (p,d,q) (P,D,Q)s

$$\Phi_P(B^s)\phi_p(B)\nabla^d \nabla^D{}_s FPI_t = \Theta_Q(B^s)\theta_q(B)\varepsilon_t$$

Where, p Order of the non-seasonal AR (AutoRegressive) component, d Degree of non-seasonal differencing, q Order of the non-seasonal MA (Moving Average) component , P Order of the seasonal AR component, D is Degree of seasonal differencing, Q is Order of the seasonal MA component, **s** is Length of the seasonal cycle, $\phi_p(B)$, Non-seasonal AR operator, $\theta_q(B)$ Non-seasonal MA operator, $\Phi_P(B^s)$ Seasonal AR operator, $\Theta_Q(B^s)$ is Seasonal MA operator, ∇^d is Non-seasonal differencing operator, $\nabla^D{}_s$ is Seasonal differencing operator, ε_t is White noise (error term), PIF_t is food price inflation at time t, ϕ_1 and Φ_1 coefficients of non-seasonal and seasonal AR order p, and P respectively, θ_1 *and* Θ_1 coefficients of non-seasonal and seasonal MA order p, and P respectively, D and d is orders of seasonal and non-seasonal difference of food price inflation, S is number period of seasonal in this case s = 4, since our data was quarterly data. However, the implementation of the optimized SARIMA model is limited by main factors. First, the model uses only historical values of the dependent variable series to inform the model and does not include any additional exogenous variables despite evidence of several variables' ability to affect the prices of food items (Adjemian et al., 2023).

3.5.4 Testing Stationary: Unit Root Test

Time series data must be made stationary before a specific model can be fitted to it. When the mean and auto covariance of a time series are constant across time, the series is said to be stationar. The stochastic process FPI_t is therefore considered stationary if:

$$E(FPI_t) = \mu \ constant \ for \ all \ value \ of \ t$$

$$cov(FPI_t\ , FPI_{t-j}) = \rho_j = E[(FPI_t - \mu)(FPI_{t-j} - \mu)T] = \rho - jT \ \ldots. \qquad (3.5)$$

$$for \ all \ t \ and \ j = 1,2,3 \ldots$$

Where, FPI is food price inflation at time t, t is time period from 2000 to 2024 to test for stationarity of a series several procedures have been developed. The most popular ones are the Augmented Dickey Fuller (ADF) test that was proposed by Dickey and Fuller (1979, 1981), and the Phillip-Perron (PP) test that was proposed by Phillips and Perron (1987, 1988) and graphical visualization.

3.5.5 Building ARIMA Models

To identify a perfect ARIMA model for a particular time series data, Box and Jenkins (1976) proposed a methodology that consists of four phases.

Model Identification

Determining the differencing need for establishing stationarity as well as the sequence of the seasonal and non-seasonal AR and MA operators for the residual series are the goals of the identification stage. The two most helpful instruments in any effort to identify a time series model are the autocorrelations function (ACF) and the partial autocorrelation functions (PACF) (Granger and Newbold, 1986).

Autocorrelation Function (ACF)

The degree of linear dependency between observations in a time series that are separated by a lag k is measured by the sample ACF (ρ_k). The autocorrelation coefficient is a crucial tool for determining models and detecting stationarity in time series data. The autocorrelation coefficient (ACF), which measures the relationship between series values across time, is provided by:

$$\rho_k = \frac{cov(FPI_{t,} \ \ FPI_{t+h})}{\sqrt{var(FPI_t)}\sqrt{var(FPI_{t+h})}} = \frac{\gamma_\kappa}{\gamma_o} \text{ were, k = 1, 2…}\frac{n}{2} \qquad (3.6)$$

Where, γ_o is the variance of stationary series, which is fixed and equal for all different time period, γ_κ is k th covariance between FPI_t series at time t and t+h , ρ_k is autocorrelation of food price inflation series.

Partial Autocorrelation Function (PACF)

Partial autocorrelation is a second crucial tool for determining time series models. It can also be used to determine which moving average, seasonal moving average, seasonal autoregressive, and non-seasonal autoregressive variables should be included in the model in what sequence. When the influence of time lags is held constant, the partial autocorrelation coefficient of order calculates the correlation between values k periods apart, where k is the food inflation lag. In order to estimate the partial autocorrelation of order k, fitting an autoregressive model of order AR (k), the partial autocorrelation function (PACF) is the partial autocorrelation coefficient as a function of k.

$$\varphi_k = \frac{\gamma_\kappa - \sum_{i=1}^{k-1} \varphi_i \gamma_{\kappa-1}}{1 - \sum_{i=1}^{k-1} \varphi_i \gamma_i} \text{ Where, k=1, 2 ….}\frac{n}{2} \qquad (3.7)$$

Where φ_k is the partial autocorrelation at lag k, γ_k is the auto covariance at lag k and φ_i represents the partial autocorrelation at lag i, where i ranges from 1 to k-1.

Model selection criteria

Traditional model selection criteria, such as the Akaike Information Criterion (AIC) proposed by Akaike (1974) and the Schwarz Bayesian Information Criterion (SBIC) proposed by Schwartz (1978), were used to identify the optimal lag specification for the model. In time series analysis, there may be multiple adequate models that can fit a given data set.

Akaike's Information Criterion (AIC)

Akaike (1974) proposed the goodness of fit for ARMA (p, q) by balancing the error of the fit against the number of parameters in the model. It is defined by:

$$AIC = -2logL_k + 2k \quad (3.8)$$

Where, L_k is the maximized log-likelihood and k = 1+p+q, is the number of parameters in the model. The most appropriate model is the one which minimizes the AIC, and there is no requirement for the models to be nested.

Bayesian Information Criterion (BIC)

As noted by Shumway and Stoffer (2010), various simulation studies have tended to show that AIC tends to be better in smaller samples where the relative number of parameters is large (Macquarie and Tsai, 1998), while BIC does well in obtaining the correct order in large samples.

Schwarz (1978) introduced the BIC and is defined:

$$BIC = -2logL_k + klogT \quad (3.9)$$

3.5.6 Parameters Estimation

The models use a non-linear estimation approach, employing maximum likelihood estimation (MLE) to optimize parameters by maximizing the likelihood function. This method ensures the best fit for forecasting while capturing both autoregressive and moving average components.

3.5.7 Diagnostic checking

We perform diagnostic tests after the model has been trained. Normality check was done via residual histograms, Independence verification using ACF plots, and Ljung-Box (LB) test to assess residual autocorrelation. The LB test's Q-statistic evaluates the null hypothesis that residuals are white noise (no autocorrelation up to a specified lag), the Q-statistic at lag is calculated as follows:

$$Q = T(t +) \sum_{i=1}^{m} \frac{\widehat{\rho_j}}{T-j} \qquad (3.10)$$

Where, $\widehat{\rho_j}$ is the j^{th} lag of autocorrelation of residual and T is the number of observations. If the series is not based upon results of ARIMA estimation, then under the null hypothesis, Q is asymptotically distributed as chi-square with degrees of freedom equal to (m - k) where k denotes the number of parameters.

Correlogram Squared Residuals

The existence of serial correlation can be determined using correlogram squared residuals. The hypotheses to be tested are as follows:

H_O There is no serial correlation Vs H_1 There is serial correlation.

We finally predicted the food prices inflation after diagnosing the model. Traditional approaches like ARIMA face limitations due to: Sensitivity to shocks unexpected events degrade forecast accuracy over longer horizons (Cecchetti and Moessner, 2008). Nonlinearity poor performance with erratic, nonlinear food inflation trends (Gustavo S. and Wagner P., 2023), and Subjectivity reliance on forecaster expertise introduces bias (Kontopoulou et al., 2023). Unlike ARIMA, random forests autonomously handle shocks, nonlinearity, and noisy data (Cui and Athey, 2022).

3.6 Machine learning (ML) approach

Machine learning is the integration of automated computer algorithms with robust statistical techniques to uncover (identify) concealed patterns within datasets. ML approaches can be categorized into three primary groups: Supervised learning which is employed when the dependent variables are distinctly identified and labeled, even if the exact relationships within the data remain unknown (linear regression, logistic regression). The objective of this method is to discover a model that connects a target variable with a collection of independent variables akin to a statistical model (Ngai and Wu, Y., 2022). Unsupervised learning approach comprises classes of ML techniques that reveal unrecognized patterns within a data set without pre-existing labels. It is applied when there is no specified output defined in advance, and the aim is to identify data patterns and categorize data without a target output (e. g. , cluster analysis, principal components) (Jung et al., 2018; Ngai and Wu, Y., 2022). Reinforcement learning establishes a dynamic environment to engage with the learning algorithm, aiming to deliver an assessment of the system's response. Supervised learning refers to the process where the algorithm learns from the training dataset, resembling a teacher overseeing the educational process. It is a machine learning task that involves learning a function that associates an input with an output based on example input-output pairs (Aronsson et al., 2021).

The supervised ML methods discussed here can be generally categorized into two groups. The first group consists of penalized linear models or shrinkage models. The second category of machine learning techniques emphasizes nonlinear models, such as ensemble approaches (e. g., Random Forest) and ANN (Multiple-Layers Perceptron, Neural Network Autoregressive) models. Nonetheless, this study examines the Random Forest method and does not take into account advanced deep learning techniques like Long Short-Term Memory (LSTM) and Recurrent Neural Networks (RNNs), which are specifically tailored for processing sequential data. The Machine Learning Approach (Non-linear machine learning) is capable of addressing the non-linear characteristics of food price inflation, as we previously outlined in the last chapter. It operates in a non-linear manner, offering a potential benefit in the analysis of variables with intricate correlations when compared to linear models (Aronsson et al., 2021).

3.6.1 Random Forest Method

The random forest (RF) model (Breiman, 2001) is an ensemble learning method that constructs multiple decision trees using bootstrapped data and random feature subsets, then averages their predictions. This approach reduces overfitting, captures non-linear relationships, and improves accuracy through ensemble diversity RF excels in economic forecasting, outperforming traditional models in inflation prediction correlations (Aronsson et al., 2021; Sidra Mehtab & Jaydip Sen, 2020). Its advantages include handling high-dimensional data without overfitting (Biau & Elia, 2010), intuitive non-linearity modeling (Woloszko, 2020), and requiring fewer parameters than other ML methods (Zhou et al., 2016). The Random Forest (RF) model was selected for this study based on four key factors aligned with Ethiopia's food inflation dynamics and data limitations: first, its ability to capture non-linear relationships, unlike traditional linear models; second, robustness to limited and noisy data, such as Ethiopia's quarterly data (2000–2024), which is affected by climate and policy shocks; thirdly, suitability for moderate-sized datasets where deep learning is impractical; and fourth, capacity to minimize prediction error by splitting predictors X_t into terminal nodes K, optimizing the sum of squared errors in forecasting FPI_{t+h}.

$$FPI_{t+h}|\text{t} = \sum_{k=1}^{k} C_k I_k(X_t, \theta_k) \qquad (3.11)$$

This formula represents the ensemble prediction of Random Forest for forecasting FPI_{t+h} given data available at time t, Where, c_k is the predicted constant (usually the mean value) of FPI in terminal node k, k is index of each terminal node (or tree) up to K (total number of trees or terminal nodes across the forest) and $I_k(X_t, \theta_k)$ is indicators function that equals 1 if the input X_t (predictors at time t) falls into terminal node k under tree parameters θ_k; 0 otherwise

$$I_k(X_t, \theta_k) = \begin{cases} 1 \; if \; x_t \epsilon R_k(\theta_k) \\ 0, \quad otherwise \end{cases} \qquad (3.12)$$

Where $R_k(\theta_k)$ is the region the randomly constructed regression trees are aggregated based on bootstrapping, where B = 500 bootstrap samples are used, for which a subset of original regressors is chosen at random. For each sample b, b =1… B, a tree with K_b regions is estimated for a randomly selected subset of the original regressors. K_b is determined in order to leave a minimum number of observations in each region. For every bootstrap sample eventually a tree with K_b regions will exist. Finally, the forecasts are then given by

$$FPI_{t+h}|\text{t} = \frac{1}{B}\sum_{b=1}^{B} f_t(x) \quad (3.13)$$

Where B is sample size of bootstrap sample or total number of tree, $f_t(x)$ the prediction made by the t^{th} tree for input X. In traditional econometric models, this is done to avoid so-called multicollinearity problem. Kane et al. (2014) demonstrate the application of Random Forests for forecasting avian influenza H5N1 outbreaks and compare its performance to that of ARIMA models. The authors found that Random Forest offers superior predictive performance compared to the ARIMA model for forecasting infectious disease outbreaks. Additionally, Random Forest's ability to manage high-dimensional data while preserving interpretability makes it a popular choice across multiple fields. Advantages of RF: Proficiency in handling datasets replete with a profusion of features and categories. Remarkable resilience to variations in the scaling of feature values...feature values, which allows the model to maintain consistent performance across different datasets. This adaptability is crucial for ensuring robust predictions and minimizing the impact of outliers or noise in the data. Techniques are available to gauge the significance of individual features within the model. Exceptional precision in the classification of novel data points. Versatility in conducting measurements across divergent scales, encompassing numerical, ordinal, and nominal realms..

3.6.2 Random Forest: Regression Tree

To understand the RF method, it is crucial to first understand the Classification and Regression Tree (Breiman et al. 1984), which the algorithm is based on. Random forest regression predicts continuous numerical values by averaging outputs from multiple decision trees, each trained on random data subsets and features (Medeiros et al., 2021). Each tree in the forest is trained on a bootstrapped sample of the dataset, and a random subset of features is selected for splitting at each node. This randomness increases diversity among the trees, reducing over fitting and improving generalization. The final food inflation prediction is obtained by averaging the outputs of all individual trees, making the model robust to noise and outliers (Breiman, 2001).

Regression trees within RF are simple, non-linear, and non-parametric models that partition the data into regions based on feature values. They are particularly effective for capturing complex, non-linear relationships between input variables and the target variable. By combining multiple trees, RF enhances predictive accuracy and handles high-dimensional data efficiently data (Medeiros et al., 2021). RF regression is widely used in applications like forecasting food price inflation, where non-linear patterns and complex correlations are prevalent. Studies, such as (Meuller, 2022; Araujo & Gaglianone, 2023), demonstrate that RF outperforms traditional models and other machine learning techniques in accuracy and robustness, making it a preferred choice for regression tasks. The predictions from the CART could be expressed as conditional expectations f(x) = E(Y |X = x) and is calculated by using several base learners h_n (x).

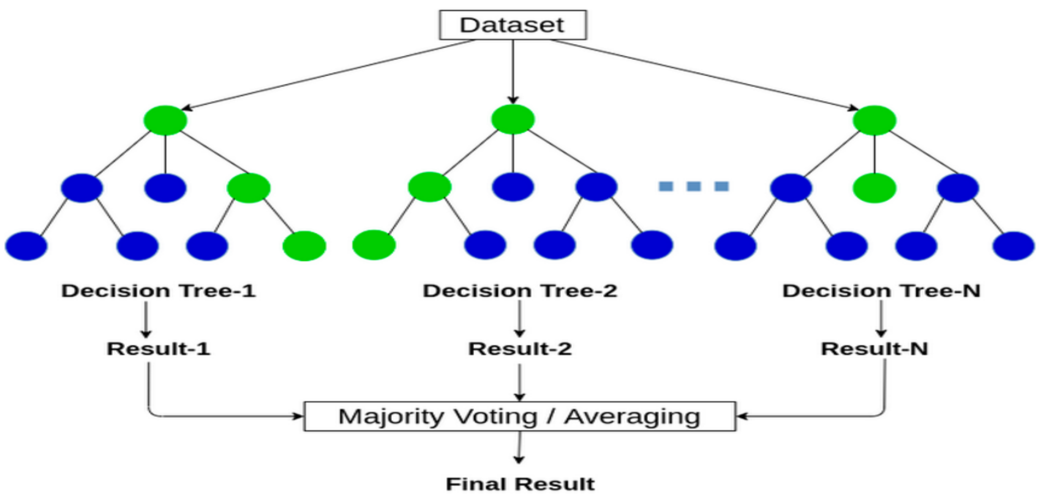


Figure 2: Random Forest Architecture

$$f(x) = \frac{1}{N}\sum_{n=1}^{N} h_n(x) \quad 3.14$$

Where, N is the number of observations inside the leaf node. When a leaf node contains only one observation (common with a single base learner), the model overfits achieving perfect training accuracy but failing on unseen data. Predictions for test data simply replicate that single value, with low bias and low variance.

3.6.3 Random Forest: Bagging

RF combines bagging (Breiman, 1996) and CART (Breiman, 1984) to create an ensemble of decision trees: Random subsets of the original data are generated via bootstrapping (allowing duplicates), ensuring diversity across subsets. Each subset trains a regression tree, with node splits based on a random subset of variables, adding further variability. The final forecast averages predictions from 500 trees, reducing bias and variance. New observations are passed through all trees, and their aggregated output yields the prediction.

$$\hat{y}_i = \frac{1}{JB}\sum_{b=1}^{B}\sum_{j=1}^{J} h_j\,(x) \tag{3.15}$$

$$\hat{y}_i = \frac{1}{B}\sum_{b=1}^{B} f(x) \tag{3.16}$$

Where B is the number of decision trees in forest, f(x) is the prediction from a tree, J is the number of observation in a leaf node and h_j (x) is the base learner inside a leaf node.

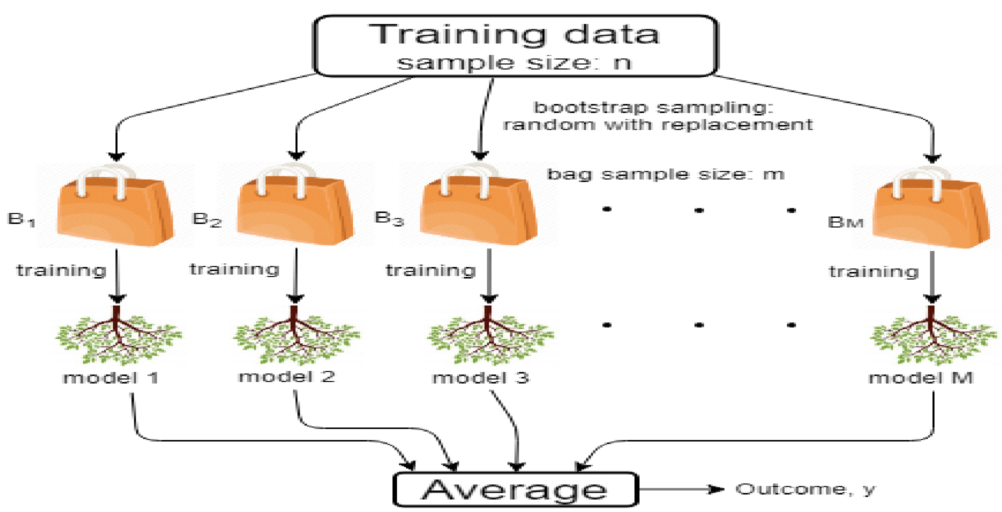


Figure 3: The structure of the Random Forest describing how predictions are being made by being passed through all trees.

The variance of a Random Forest prediction at a given point can be approximated as:

$$Var[\hat{f}_{RF}(x)] = \rho\sigma^2 + \frac{1-\rho}{T}\sigma^2 \quad (3.17)$$

Where $\hat{f}_{RF}(x)$ is the Random Forest prediction at point x, T is the number of trees , σ^2 is the variance of a single decision tree,, and ρ is the average correlation between pairs of trees, (T. Hastie et al., 2009) State that the forest has a declining variance when allowing for more trees.

3.6.4 Auto-Regressive Random Forest (AR-RF) Model

The AR-RF model combines Auto-Regressive (AR) and Random Forest (RF) methods to enhance food price inflation forecasting. The AR model captures time-series dependencies, while the RF model handles non-linear patterns. By merging both, the model delivers more accurate and reliable forecasts. Research shows that combining models improves forecast performance (Stock & Watson, 2004; Robert R. Andrawis, 2011). The combined forecast at time *t* is expressed as:

$$Y_t = \frac{1}{B}\sum_{b=1}^{N} f^{(b)}({Y_t}^{(p)}, X_t) \quad (3.19)$$

Where, Y_t Forecasted value of the dependent variable at time t, N, Number of trees in the Random Forest, $f^{(b)}$,Prediction function of the b th tree, ${Y_t}^{(p)}$, Vector of the last p lags of the dependent variable and X_t Vector of k contemporaneous exogenous variables. In general:

$$\widehat{FPI}_t = \beta_1(\widehat{FPI}_{AR,t}) + \beta_2(\widehat{FPI}_{RF,t}) \quad (3.20)$$

Where, $\widehat{FPI}_t$ Predicted food price inflation at time t, $(\widehat{FPI}_{AR,t})$ Forecast from the Auto-Regressive model, $(\widehat{FPI}_{RF,t})$ Forecast from the Random Forest model, $\beta_1 and\, \beta_2$ Weights or coefficients assigned to the AR and RF forecasts, respectively, reflecting their relative importance in the final prediction.

3.6.5 Feature Selection (Variable importance)

Feature selection is a crucial step in time series modeling, especially when using machine learning techniques like Random Forests. It helps identify which input variables (features) contribute most significantly to the prediction of the food inflation. Random Forests Breiman, (2001) are widely used for non-parametric regression due to their high accuracy and ability to handle high-dimensional, correlated data. Variable importance is assessed by measuring the increase in MSE when a variable's values are randomized (Biau & Elia, 2010). In time-series, such selection methods help capture long-memory effects often missed by traditional models (Segnon & Bekiros, 2020).

3.7 Forecasting Strategy

There are two types of forecasts: in-sample period forecasts (first quarter of 2000 to fourth quarter of 2019) and post-sample period forecasts (first quarter of 2020 to fourth quarter of 2024). The former is utilized to build confidence in the model while the latter is aimed at producing authentic desired forecasts. In forecasting, the objective is to estimate future values of food price inflation, utilizing the data gathered up to the present, $FPI = \{FPI_t, FPI_{t-1}, \dots . FPI_1\}$ For ARIMA and SARIMA model, we assume FPI_t is stationary and the model parameters are known.

3.8 Forecast Evaluation Methods

The forecast error is the difference between the actual food price inflation and the predicted food price inflation. There are several measures of forecast errors. The measure adopted in the study was used by Moshiri (1997) and Haider & Hanif (2007).

3.8.1 Root Mean Squared Error (RMSE)

RMSE is a measure of the size of the forecast error, that is, of a magnitude of typical mistakes made using a forecasting model (Stock & Watson, 2007). The RMSE is computed as the measure of the forecast accuracy for each model. The RMSE is calculated as follows.

$$RMSE = \sqrt{\frac{1}{h}\sum_{i=1}^{h}(FPI_t - \widehat{FPI_{t+h}})^2} \quad (3.21)$$

Where $FPI_t\ is$ is observed series of food price inflation FPI_{t+h} predicted values of food price inflation (in-sample or out-sample). RMSE is commonly used for regression tasks, where the goal is to predict a continuous value.

3.8.2 Mean Absolute Error (MAE)

MAE is another common performance measure for regression tasks, and is calculated as the average of the absolute differences between the predicted and true values. The MAE is defined as:

$$MAE = \frac{1}{h}|FPI_t - \widehat{FPI_{t+h}})| \quad (3.22)$$

Mean Absolute Error is robust to outliers, as it measures raw error magnitude in the original units, aiding interpretability (Zhang and Yang, 2015). Root Mean Squared Error penalizes large errors more heavily, ideal for outlier-free data where minimizing prediction error is critical.

3.8.3 Mean Absolute Percentage Error (MAPE)

The Mean Absolute Percentage Error (MAPE) is widely used forecast accuracy metric, favored for its intuitive percentage-scale interpretation (Gartner, 2018). It excels in time series evaluation and S&OP contexts, provided the data lacks extremes or zeros.

$$MAPE = \frac{1}{h}\left|\frac{FPI_t - \widehat{FPI_{t+h}})}{FPI_t}\right| \quad (3.23)$$

MAPE is expressed as a percentage, which is scale-independent and can be used for comparing forecasts on different scales.

3.9 Software for Data Analysis

The study employed R-Studio 4.4.3**,** Stata 14, and Jupyter Notebook for forecasting food price inflation. Key packages included: R: forecast (with arima() for time series modeling).

CHAPTER FOUR

4 RESULTS AND DISCUSSIONS

4.1 Exploratory Data Analysis

Visualizing the data using a variety of statistical methods and graphical representations is a component of exploratory data analysis (EDA). Time series charts, correlation matrices, and descriptive statistics were often used methods. Finding trends, patterns, seasonality, and any outliers in the data was the aim.

Descriptive Statistics

The study utilizes quarterly year-on-year food price inflation (%) data sourced from the Central Statistical Service (CSS), spanning from Q1 2000 to Q4 2024. Descriptive statistics were employed to analyze the trend, variability, and underlying patterns in the inflation series.

Table 1: Summary statistics for the quarter food price (year-on-year) inflation in percentage

Mean	Median	Variance	Std. Dev.	Min	Max	Skewness	Kurtosis	Jarque-Bera
17.7191	**14.57**	**271.1814**	**16.46759**	**-12.93**	**85.87**	**1.02035**	**5.051076**	**21.107**

Descriptive Statistics of Food Price Inflation (2000Q1–2024Q4): Average food inflation is 17.72%. Extreme volatility observed, from -12.93% to 85.87%. Positive skewness (asymmetric right-tail) Leptokurtic (kurtosis = 5.05 > 3), indicating fat tails and outliers. The Jarque-Bera test rejected normality ($p < 0.05$), confirming non-Gaussian distribution so the data is normally distributed.

Trend Analysis

Figure four bellow indicate that key trends in quarterly food price inflation from 2000 to 2002, with periods of decline below zero (indicating low inflation). From 2003 to 2008 indicated sharp rise followed by steady growth, peaking in 2008Q4 before collapsing in 2009 likely due to the global crisis (structural break), 2015–2022 tend implies gradual upward trend, tapering off by 2023Q4. The series exhibits non-stationarity (shifting mean/variance), highlighting nonlinear dynamics.

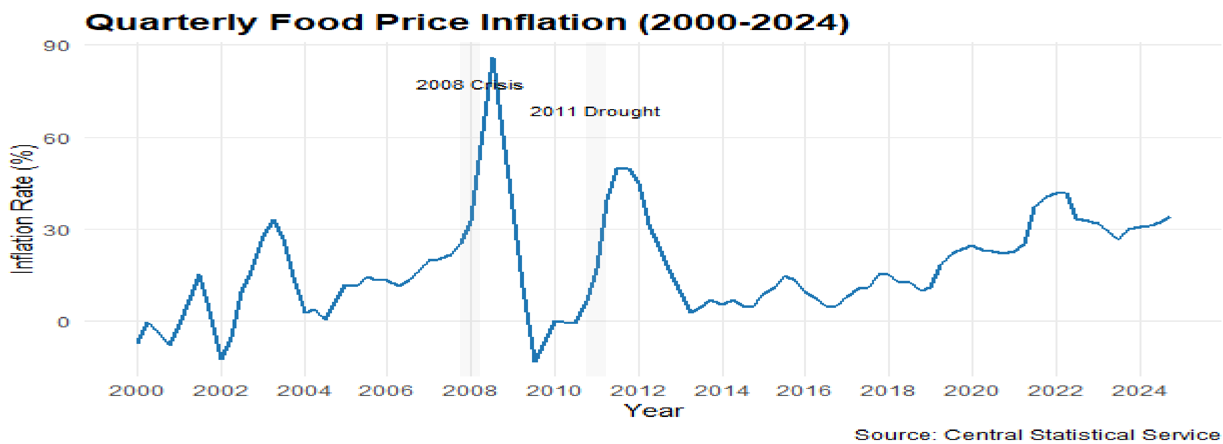


Figure 4: Quarterly food price inflation trends in Ethiopia (2000–2024).

Seasonal Adjustment Analysis indicating that reveals the food price inflation series exhibited non-stationarity (time-varying mean/variance) and mild seasonality. However, after seasonal adjustment (Figure 5), minimal deviation from the original series suggests negligible seasonal influence, implying trends were primarily driven by non-seasonal factors.

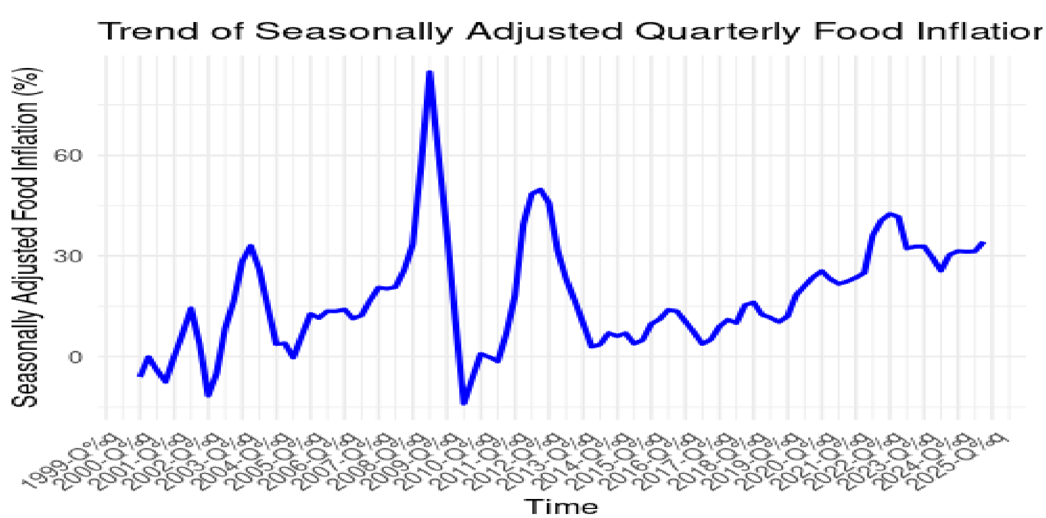


Figure 5: Trend plots of quarter data of seasonal adjusted food price inflation in Ethiopia (2000-2024)

Figure 5 indicated that the 2000 to 2007 trend was moderate fluctuations (avg. 8.5%), with brief deflation (-12.93% in 2009Q3). 2007–2009 crisis extreme volatility (peak 85.87% in 2008Q3) driven by global commodity shocks. Post-2010: Sustained high inflation (avg. 18.3%), hitting 41.73% in 2022Q1. 2023–2024: Slight easing but persistently high (30–34% range).

Correlation Matrix

Covariance between predictor variables was examined using a correlation matrix. The existence of multi-collinearity was thus demonstrated by the high correlation (correlation coefficients above 0.80) of numerous covariates, including broad money, fertilizer, lending interest rate, net demand deposit, Real GDP, Revenue, expenditure, narrow money, and food price index.

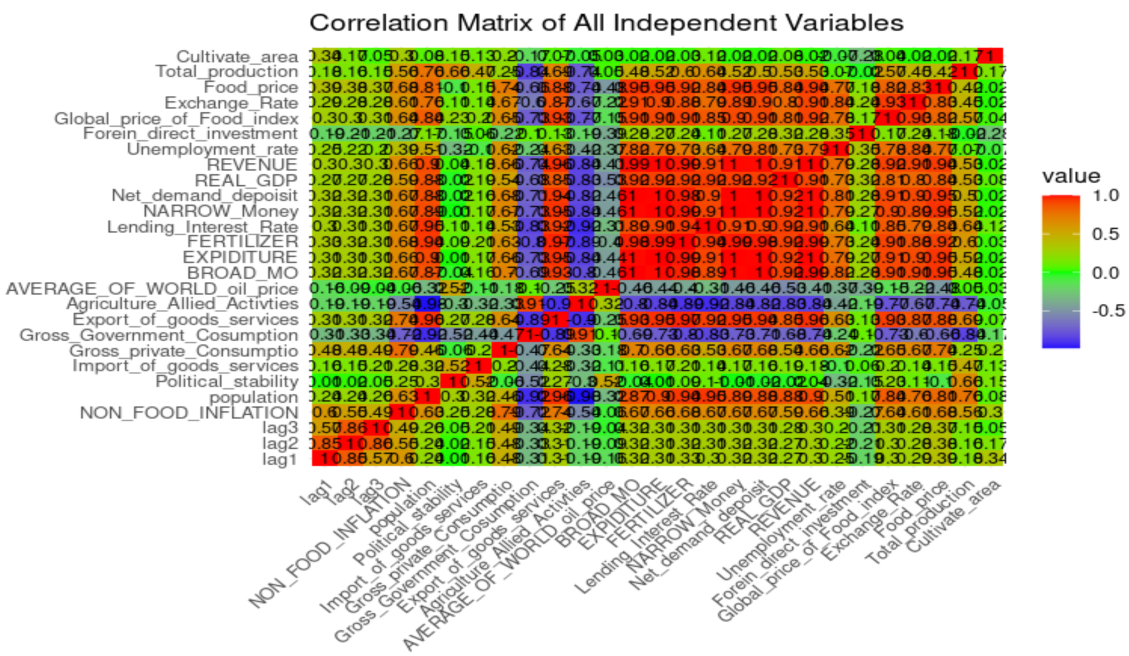


Figure 6: Correlation of Exogenous Variables for food price inflation in Ethiopia

4.2 Data Preprocess

The data was proposed to partition the dataset for Random forest model and make the series stationary for time series analysis before the model was developed.

4.2.1 Test for Stationarity

The null hypothesis that a unit root exists is tested against the alternative hypothesis that it does not using the Phillips-Perron (PP) and Augmented Dickey-Fuller (ADF) tests. ADF and PP tests to check for unit roots (null: non-stationary). KPSS test to assess level/trend stationarity (null: non-stationary). ACF displayed exponential decay and seasonality, suggesting a trend. PACF indicated stationarity beyond the 1st, 2nd, and 4th lags.

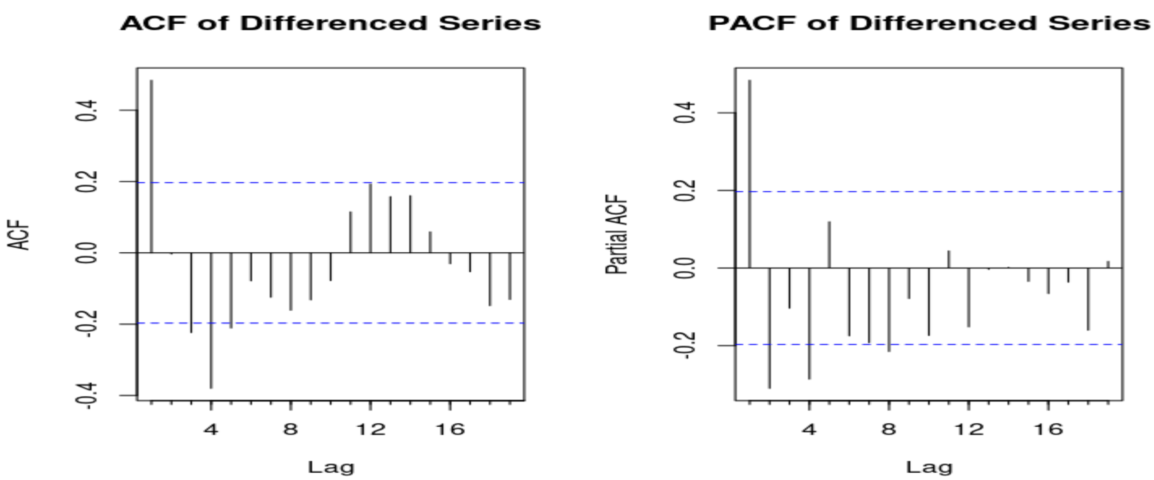


Figure 7: ACF and PACF plot of seasonal adjusted series

Table 2: Unit Roots test of food Inflation

Test	At level Statistic	Critical value at 5%	p-value	At Differenced statistics Statistic	Critical value at 5%	p-value
ADF	-1.234	-2.891	0.452	-4.567	-2.891	0.01
PP	-1.456	-2.891	0.324	-6.789	-2.891	0.01
KPSS	0.673	0.463	0.023	0.123	0.463	0.12

Unit root tests on the differenced series confirmed stationary, because KPSS ($p=0.12 > 0.05$) failed to reject stationarity. ADF & PP ($p=0.01 < 0.05$) rejected unit roots. Results indicate stable mean/variance (Figure 7), validating the use of differenced data for modeling.

Figure 8: Plot of first differenced of food price inflation in Ethiopia

The differenced series exhibited volatility clustering (Figure 8), where large changes tend to follow large ones, and small changes follow small ones, indicating patterns of similar magnitudes. This pattern suggests suitability for GARCH/ARCH models, their documented limitations in food inflation forecasting excluded them from this analysis.

4.2.2 Splitting of the Data

Splitting time series data into training and test periods is crucial for building reliable models. We split data for both Machine Learning (ML) and Classical Time Series Models, but the method differs based on the approach. The converted data was split between training and test sets in an 80:20 percent ratio before model analysis. Training set (80%) 2000Q1–2019Q4 (used to train time-series and Random Forest models) for identifying patterns between input features and output targets. Test set (20%) 2020Q1–2024Q4 (reserved for validation and performance evaluation) for Provides unbiased evaluation of the model's predictive accuracy on unseen data.

4.3 Forecasting with Traditional Model (Benchmark Model)

Traditional time series models serve as essential benchmarks for evaluating forecasting performance. Below is a structured approach to implementing and interpreting these models for food inflation forecasting:

4.3.1 Forecasting with Autoregressive model

The Autoregressive (AR) model is a fundamental time series forecasting method that predicts future values based on linear combinations of past observations. Below is a structured guide to implementing and interpreting AR models for food inflation forecasting.

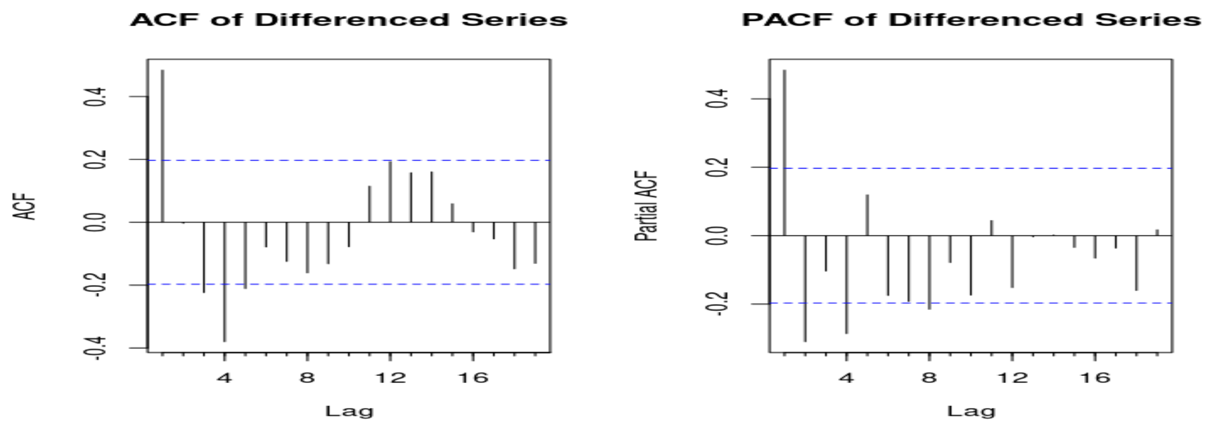


Figure 9: Plot of ACF and PACF for AR Model

The PACF plot typically shows significant spikes up to lag 2, suggesting the AR order. AIC/BIC criteria help choose the most parsimonious model that fits well. Each AR coefficient (φ_1 to φ_4) represents the effect of that lag on current inflation. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) quantifies forecast accuracy.

Table 3: Candidate model for AR model

Model	AIC	log likelihood	RMSE	MAE	MAPE
ARIMA(1,1,0)	613.65	-253.82	7.267027	4.44175	106.3438
ARIMA(2,1,0)	575.3	-234.52	6.047367	3.55916	159.5955
ARIMA(3,1,0)	575.83	-231.93	5.610765	3.52716	164.9541
ARIMA(4,1,0)	**573.57**	**-231.80**	**5.602171**	**3.51423**	**137.1430**

By incorporating lagged values of food inflation as independent variables, we introduce an autoregressive component to the model, a common approach in time series analysis to account for temporal dependencies.

4.3.2 Forecasting with an Autoregressive Integrated Moving Average Model Selection Order of ARIMA Models

To perform an ARIMA model, the number of lags of past values of food price inflation (p) and the number of previous errors to include in the model (q) must be determined. The results of the Schwarz Information Criterion (SBIC) and the Akaike Information Criterion (AIC) are (p=2) and (q=2), respectively. The Autoregressive Integrated Moving Average (ARIMA) model is regressed on two past food inflation values, one differenced and, two past error values. The ACF and PACF plots were used to determine the ARIMA order (p, d, q). Figure 11 shows that we choose the ARIMA lag order (p, d, and q) based on the training data set plot. As a result, ACF spikes out at lags 1 and 2, suggesting that the non-seasonal MA (q). The PACF plot is spiked out at lags 1, 2, and 4. Moreover, the parsimonious ARIMA model didn‘t consider higher lags, which don’t spike out for simplicity of the models.

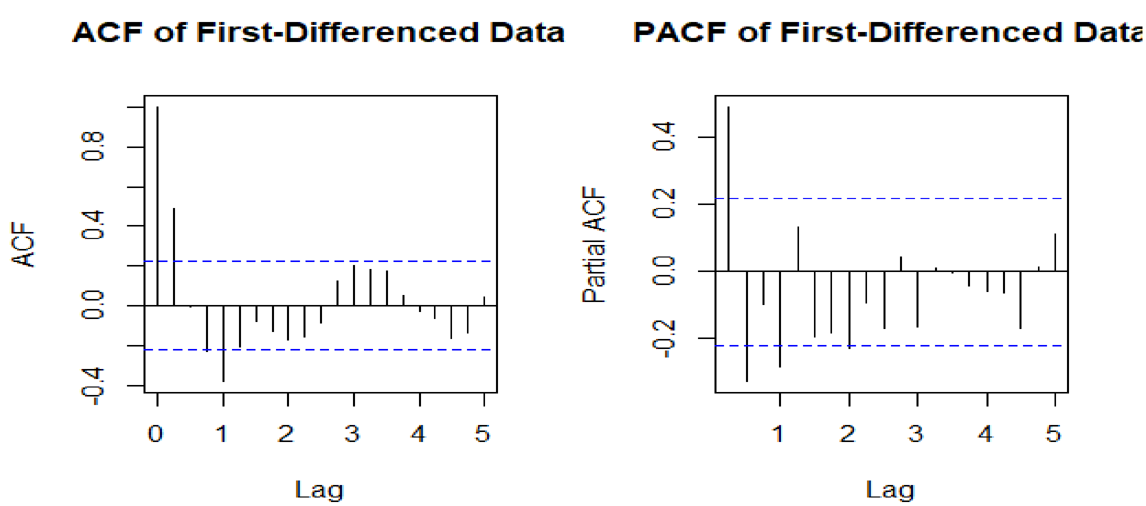


Figure 10: ACF and PACF plot of first differenced of training data set

Table 4: Candidate model for ARIMA model

Model	AIC	log likelihood	RMSE	MAE	MAPE
ARIMA(1,1,1)	713.65	-353.82	8.267027	5.441175	206.3408
ARIMA(2,1,1)	675.3	-334.52	7.047367	4.559196	259.5935
ARIMA(2,1,2)	675.83	-331.91	6.610765	4.52716	264.9551
ARIMA(1,1,2)	**673.57**	**-331.79**	**6.602171**	**4.518331**	**247.1427**

ACF and PACF plots of the training data set's first difference are shown in Figure 9. Thus were the probable orders of ARIMA (p, d, q) based on the ACF and PACF graphs. Because of this, our data is quartet data, with a seasonal spike at four and a multiplicative effect. The best models to match the series were Autoregressive Integrated Moving Average (ARIMA) models, which demonstrated the presence of a seasonal component. The R function auto.arima() chose the ARIMA(1,1,2).

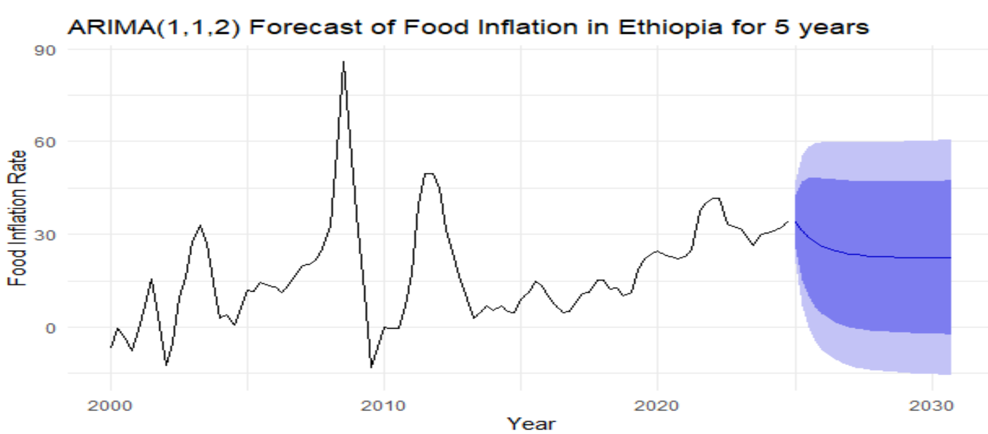


Figure 11: Food price inflation forecasting using ARIMA model in Ethiopia (2025-2029)

The historical trend of food price inflation in Ethiopia from 2000 to 2024 exhibits significant volatility, characterized by sharp peaks and dips, which likely reflect periods of economic instability and external shocks. Notably, a sustained upward trajectory between 2008 and 2012 suggests prolonged inflationary pressure, potentially driven by structural factors such as global commodity price surges, currency depreciation, or supply chain disruptions.

The confidence intervals, represented by narrow shaded bands around the forecast line, indicate a high level of certainty in the model's short-term predictions. However, the projected values for the period 2025 to 2030 show a marked deviation from historical patterns. This divergence raises concerns about the model's ability to capture structural breaks or regime changes in the data-generating process. Specifically, the ARIMA model, being linear and reliant on past trends, may struggle to accommodate such discontinuities, thereby limiting the reliability of long-term forecasts. The overall inflationary trend evident in both historical and projected data underscores the urgent need for policy intervention.

4.3.2 SARIMA Model Selection

The optimal SARIMA model was selected based on parsimony-focused criteria (AIC, BIC, HQIC) to identify the simplest adequate model, while forecast accuracy was assessed using metrics like RMSE and MAE to minimize prediction error.

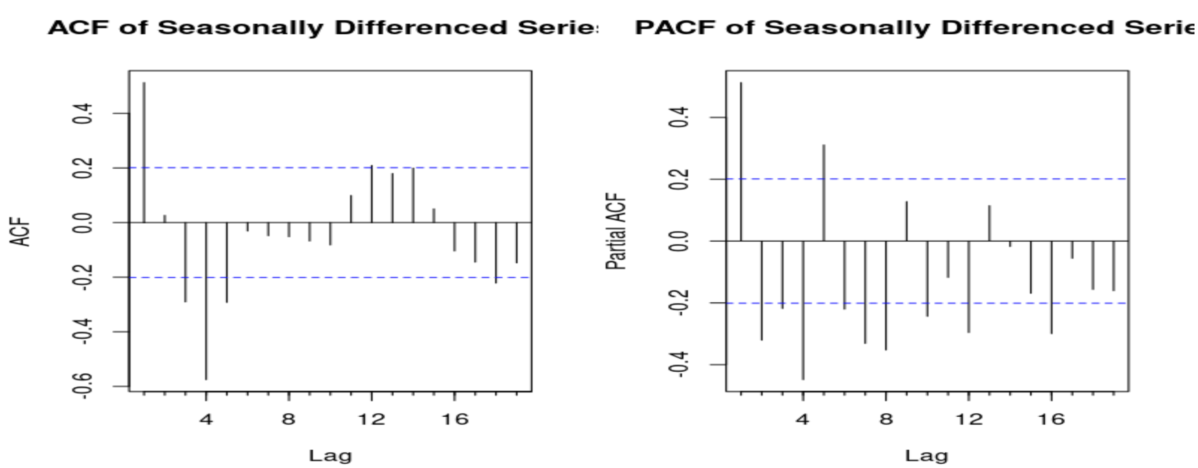


Figure 12: ACF and PACF plot of first differenced of Seasonal differenced series

Table 5: Candidate of SARIMA Models to forecast food inflation

Model	Log_Likelihood	AIC	BIC	RMSE	MAE	MAPE
SARIMA (0, 1, 1) (0, 1, 1)	-266.370	538.740	545.693	5.796	4.628	14.941
SARIMA (1, 1, 1) (0, 1, 1)	-266.365	540.729	549.999	5.794	4.611	14.866
SARIMA (0, 1, 2) (0, 1, 1)	-266.367	540.733	550.003	5.794	4.617	14.890
SARIMA (1, 1, 1) (1, 1, 1)	-262.476	534.952	546.540	7.203	4.968	14.027
SARIMA (2, 1, 2) (1, 1, 1)	**-260.163**	**534.327**	**550.549**	**12.760**	**10.955**	**33.059**

The best model to fit the training data was ARIMA (2, 1, 2) (1, 1, 1), as shown in Table 5, which had the lowest values of AIC, BIC, log likelihood, and RMSE. The ARIMA (2, 1, 2) (1, 1, 1) was selected by R's auto. The arima() function is the best model to fit training data, indicating that the decision between the SARIMA and ARIMA models was arbitrary or dependent on forecaster's experience.

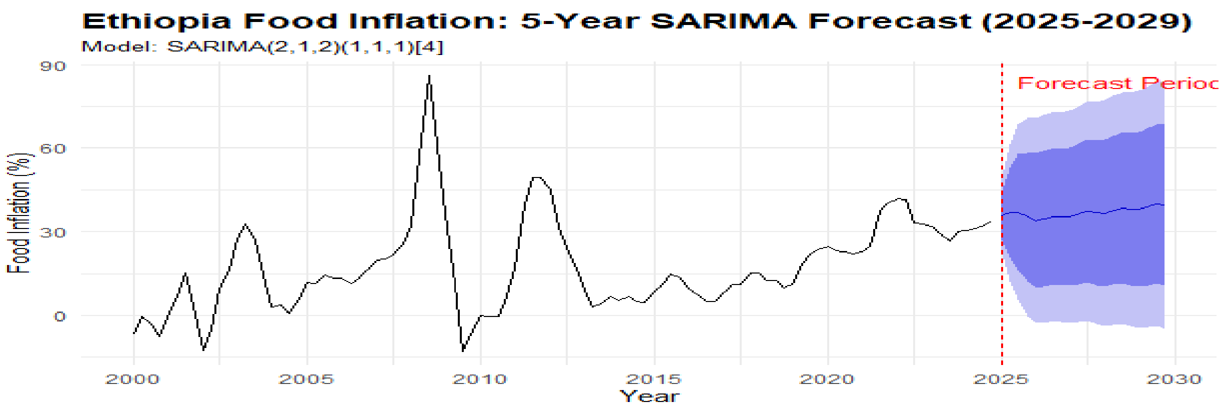


Figure 13: Forecasting food price inflation in Ethiopia using SARIMA model for 5 years

The SARIMA forecast graph for the period 2000–2024 offers valuable insights into the historical and projected dynamics of food price inflation in Ethiopia. The historical component (2000–2020) reveals pronounced volatility, with several sharp peaks and troughs. These fluctuations likely correspond to major economic and environmental events. For instance, the spike around 2008 may be attributed to the global food crisis, which drove up prices worldwide. Similarly, the 2015 drought likely caused significant supply-side disruptions, contributing to inflation. Changes in domestic policy such as adjustments to food subsidies or trade regulations may also explain some of the observed variability.

Beyond these episodic shocks, the graph illustrates a gradual upward trend, reflecting persistent inflationary pressures. These may stem from structural factors such as rapid population growth, which increases domestic food demand, and the depreciation of the Ethiopian birr, which contributes to imported inflation. Collectively, these trends underscore the multifaceted nature of food price inflation in Ethiopia and highlight the importance of incorporating both macroeconomic and environmental variables into forecasting models.

4.4 Machine Learning Method of Analysis (Proposed Model)

4.4.1 Random Forest Method

The study examined how the number of trees in the Random Forest model influenced the effectiveness of the method. Using 100 trees, the out-of-bag (OOB) RMSE for the bagging regression was 31.71306; with 500 trees, it decreased to 30.15466. This indicated that adding more trees resulted in averaging across a larger number of diverse decision trees. The study also employed bagged CART (Classification and Regression Trees) to compare OOB error with 10-fold cross-validation (CV). According to the findings, the RMSE for the OOB error was 34.24627 for 100 trees and 33.24627 for 500 trees.

Table 6: Selection of a hyperparameter (mtry) from original 27 features

Bagging Techniques Comparison			
Technique	RMSE_OOB	mtry	Notes
Bagging (100 trees)	31.71306	27	Basic bagging, higher variance
Bagging (500 trees)	30.15466	15	More trees reduce overfitting
Bagged CART (100 trees + 10-fold CV)	34.23086	9	Cross-validation optimizes feature subset
Bagged CART (500 trees + 10-fold CV)	33.24627	15	Best performance: Lowest RMSE

RMSE was utilized to choose the best model by identifying the lowest value. We noted a considerable reduction in variance (and thus, error) as the quantity of trees grew in the Random Forest model (See Figure 14). This indicated a swift enhancement in model accuracy. Nevertheless, this pattern ultimately leveled off, suggesting that increasing the number of trees no longer provided significant advantages. This leveling off probably indicated that an adequate number of trees had been attained for peak performance. Table 6 demonstrates that combining more trees, feature subsampling (lower mtry), and cross-validation yields the best predictive accuracy (lowest RMSE). This aligns with principles of ensemble methods like Random Forest, where controlled randomness and validation-driven optimization improve generalization.

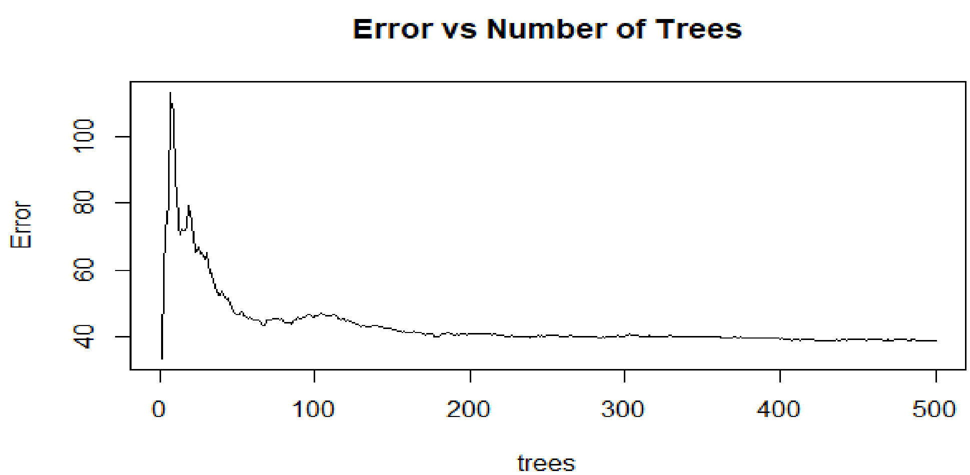


Figure 14: RF applied to food inflation as the dependent variable. 500 trees and mtry = 15.

As the number of trees in the Random Forest model increased, the error rate decreased; thus, entry at 500 trees, that is a standard number in our study, had 83.7% explained variance for food inflation. Hence, RF is in effect a method of identifying key features effectively, but for over-fitting prevention, it uses ensemble averaging. As shown in (Appendix M), the root mean square error decreases as the number of trees increases. This highlights that a higher number of trees brought about more accurate and comprehensive food inflation predictions for Ethiopia. Digging out features with random forests, variance, and error rate drop. Shown in another Appendix E, this reduction continues until the number of trees reaches 80 or so for food inflation, with no further improvement beyond that point.

Table 7: Hyperparameter Tuning Results for Random Forest Model

mtry	mix.node.size	min.node.size	RMSE	Rsquard	MAE	sampsize	ntree
15	10	1	7.986865	0.8379	5.485	100	500

Argument	**Description**
ntree	Number of trees to grow
mtry	Number of variables randomly sampled at each split
nodesize	Minimum size of terminal nodes
maxnodes	Maximum number of terminal nodes
importance	Whether to calculate variable importance
replace	bootstrap samples
sampsize	Size of sample to draw for each tree

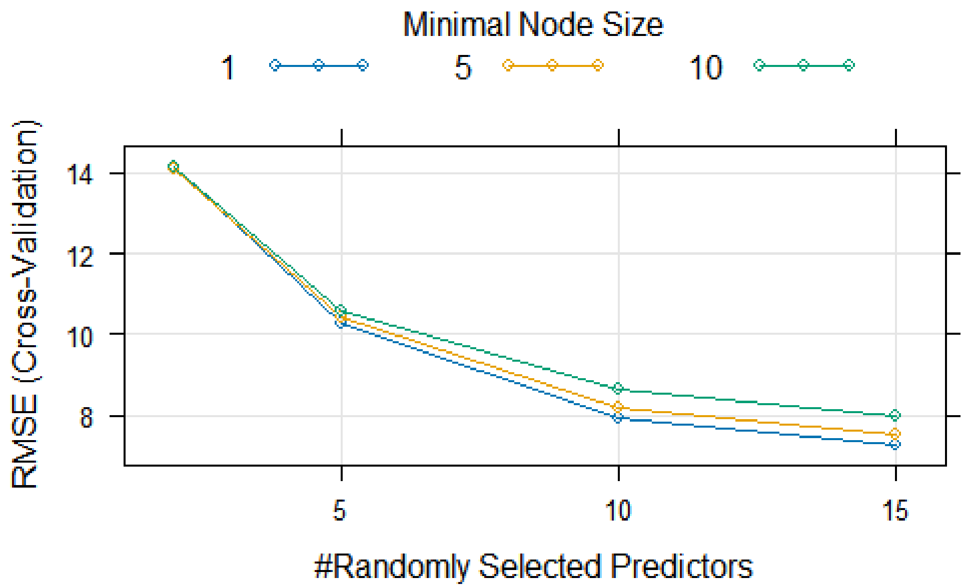


Figure 15: Hyper parameter Tuning Results for Random Forest: Effects of Node Size and Predictor Sampling on Model Accuracy

Based on the description of Figure 15, the visualization illustrates the results of hyperparameter tuning for a random forest model. It depicts how model performance, measured by RMSE, varies with different values of minimum node size and the number of randomly selected predictors (mtry). The graph indicates that increasing mtry generally improves performance (lower RMSE), and smaller node sizes (e.g., 1) tend to yield better results than larger ones (e.g., 10).

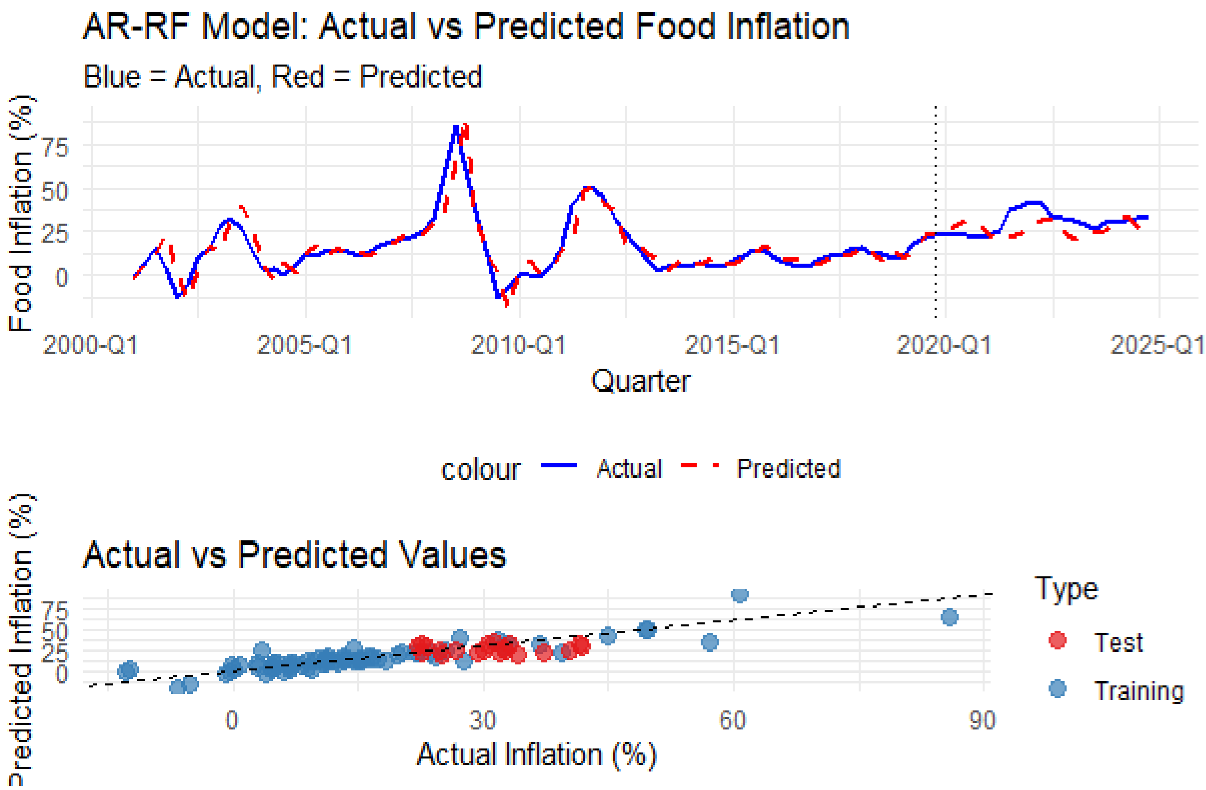


Figure 16: Lines plots comparing actual vs. predicted of from Auto Regressive Random Forest model.

Training Period: How closely the red dashed line follows the blue line shows the model's fit to past data. A close match indicates a good fit. Test Period: The model's ability to predict the most recent 5 years (20 quarters). If the red dashed line continues to track the blue line closely, the model has good predictive power for near-term trends. Model Performance Analysis of this study is that the Random Forest predictions closely track actual trends, though ensemble smoothing reduces extreme peak/trough precision. Tight alignment indicates a strong overall fit. Slight lag/overshoot during volatile periods reduced sensitivity to abrupt non-linear shifts. Effective for noisy, nonlinear data, but may benefit from hybrid approaches for complex temporal patterns.

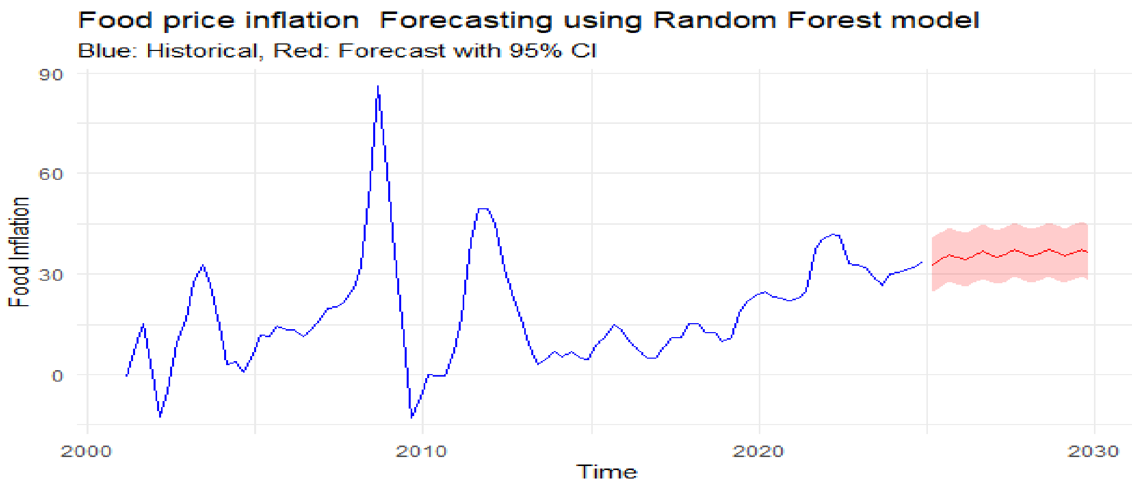


Figure 17: Food price inflation forcasting using Random Forest model

The blue line (2000 to 2024) shows significant fluctuations in food inflation. Notably, there's a sharp peak around 2010, suggesting a period of extreme volatility (potentially linked to real-world events like financial crises or supply chain disruptions). The absence of numerical axis labels makes exact values unclear, but the range (0–90 on the y-axis) implies dramatic swings over time. From the present to 2030, the red line represents the model’s forecast, with the light red shading indicating a 95% confidence interval (the range within which the model expects inflation to fall, with 95% certainty). The forecast appears to stabilize after initial fluctuations, narrowing toward the end. The use of Random Forest (an ensemble learning method) implies the model accounts for complex, non-linear relationships in factors driving food inflation (e.g., climate, geopolitical events, or market demand). The narrowing confidence interval could reflect.

4.4.2 Evolution models

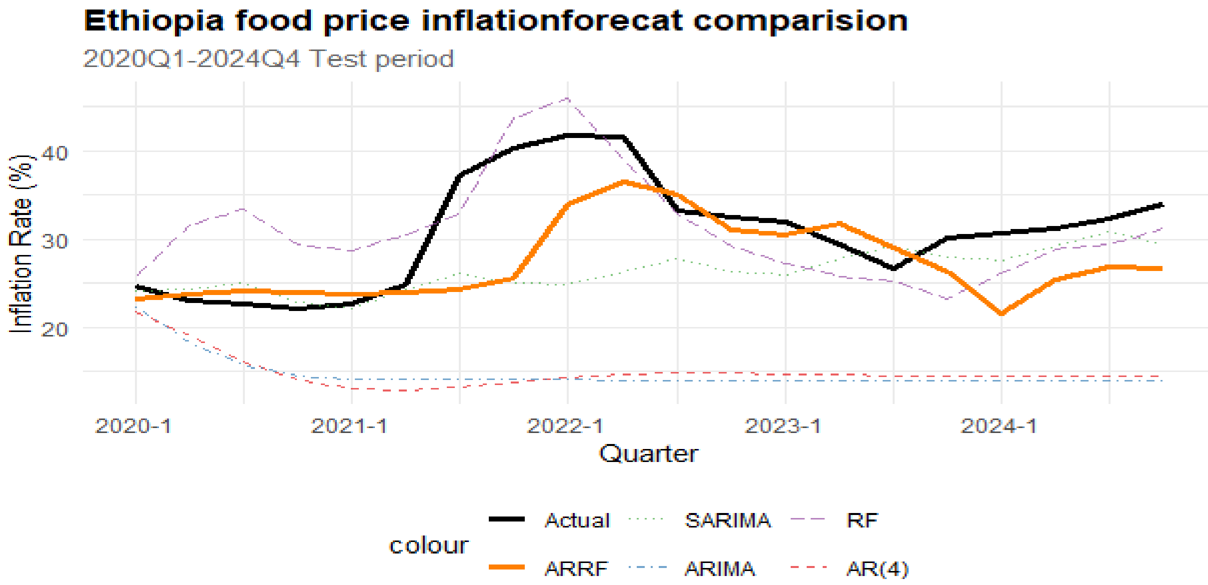


Figure 18: Forecasting comparison of models used to analyze food price inflation in Ethiopia

This line graph compares actual food price inflation in Ethiopia (black line) with forecasts from five statistical models, covering Q1 2020 to Q4 2024. Among them, the AR-RF model (yellow) aligns most closely with observed inflation, particularly after 2023, indicating its superior predictive accuracy. While the SARIMA model (red) effectively captures seasonal fluctuations, it exhibits larger overall deviations. Persistent forecasting errors such as consistent over- or underestimation highlight biases in some models. The actual inflation line reveals sharp volatility, with a pronounced peak in early 2022, likely driven by global shocks such as COVID-19-related supply chain disruptions and the Russia-Ukraine conflict's impact on grain prices. Although inflation trends downward post-peak, volatility persists, reflecting continued instability in Ethiopia's food markets. The Random Forest (purple dashed) and AR-RF (yellow) models offer smoother forecasts due to ensemble averaging, though they tend to underreact to abrupt market shifts.

Model Validation

Model validation was conducted to assess how well the trained models, fitted on historical data from 2000Q1 to 2019Q4, generalize to unseen future data. This step is crucial to ensure that models are not simply memorizing past patterns (i.e., overfitting), but are capable of adapting to real-world dynamics. Using out-of-sample data from 2020Q1 to 2024Q4, we evaluated predictive accuracy through key performance metrics, including RMSE and MAE. Models compared include AR, ARIMA, SARIMA, Random Forest (RF), and the hybrid AR-RF. Among these, the AR-RF consistently demonstrated the best overall performance, balancing short-term responsiveness with long-term trend alignment.

Table 8: Model Performance Comparison without split data

Model	RMSE	MAE	R-Squared
AR Model	17.4567	15.5667	0.818
ARIMA Model	17.3622	15.3913	0.834
SARIMA Model	17.3047	15.3996	0.832
RF Model	11.8134	10.1055	0.836
AR-RF Model	**7.7797**	**7.3207**	**0.837**

The AR-RF model consistently achieves the highest R^2 values, owing to its hybrid structure: it leverages autoregressive lags to capture temporal dependencies and employs Random Forest to model complex, nonlinear relationships. Robust model validation confirms that this approach generalizes well to unseen data, minimizing the risk of overfitting. RMSE comparisons further highlight AR-RF's superiority, showing a significant reduction in forecast error compared to traditional ARIMA models. These results support the adoption of hybrid machine learning approaches for producing more accurate and policy-reliable forecasts in the context of Ethiopia's highly volatile food inflation.

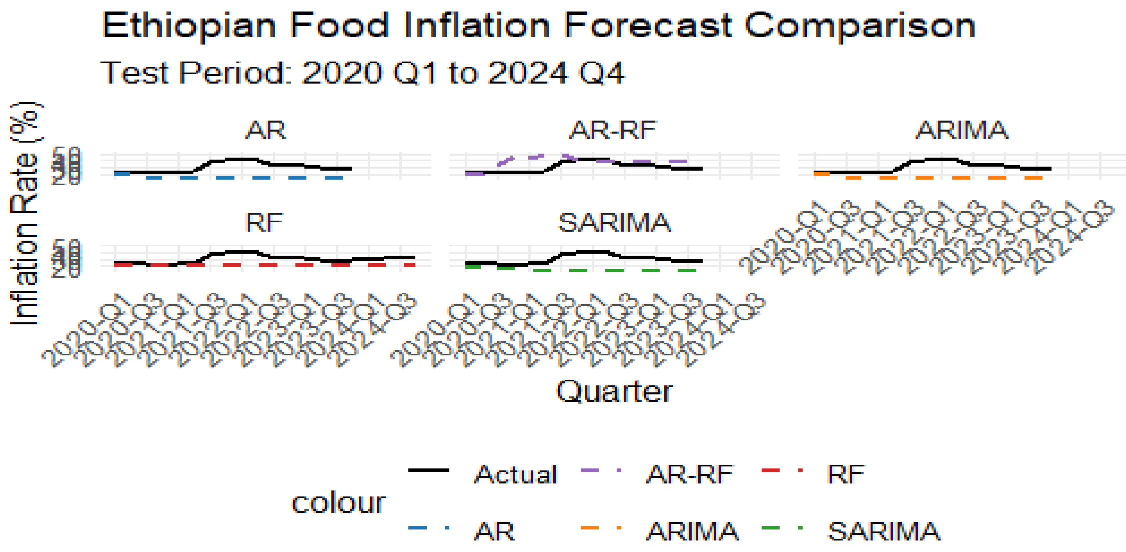


Figure 19: Compare the forecasting performance of different models for Ethiopian food inflation from Q1 2020 to Q4 2024

The graph illustrates a comparison between actual food inflation rates, indicated by the black line, and five modeling approaches: Among these, the AR-RF Hybrid shows the closest correlation with actual values, particularly during volatile phases like the spike in Q3 2021. This model effectively merges the nonlinear processing strengths of RF with the temporal dependency recognition of AR. In contrast, traditional models are unable to adequately account for major events, often smoothing trends too much and underestimating the effects of the 2023 crisis. In conclusion, the AR-RF hybrid model surpasses its alternatives, offering more dependable forecasts for the fluctuating food inflation in Ethiopia.

4.3 Variables Importance

Identifying relevant predictor variables is crucial for various applications, as opposed to merely using a black-box model to forecast the response. By utilizing variable importance metrics, we can assess potential predictor variables based on their impact on predicting the response or their possible causal effects. The %IncMSE metric indicates the extent to which the model's error increases when a variable is excluded; higher values suggest greater importance.

Table 9: Top feature selection (variables importance)

Variable	%IncMSE	IncNodePurity	Interpretation
Non Food Inflation	20.0	15,200	Strongest overall predictor, impacts both global accuracy and local splits
Average of world oil price	18.0	14,800	Critical cost-push factor with consistent splitting power
Gross private Consumption	15.0	12,500	Demand-side driver important for both metrics
Total population	14.0	11,900	Structural factor with balanced importance
Agriculture Allied Activities	13.0	10,700	Key supply-side determinant
Exchange Rate	12.0	9,500	Trade-related impacts visible in both measures
Food price	10.0	8,300	Mid-range importance despite being target-adjacent
Cultivate area	9.0	7,600	Agricultural capacity measure
Lending Interest Rate	8.0	6,900	Financial conditions indicator
Total production	7.0	5,800	General economic activity
Global price of Food index	6.0	4,500	International market influence
Import of goods services	5.0	3,800	Trade flow impact
Rainfall	3.0	2,100	Environmental factor with modest role

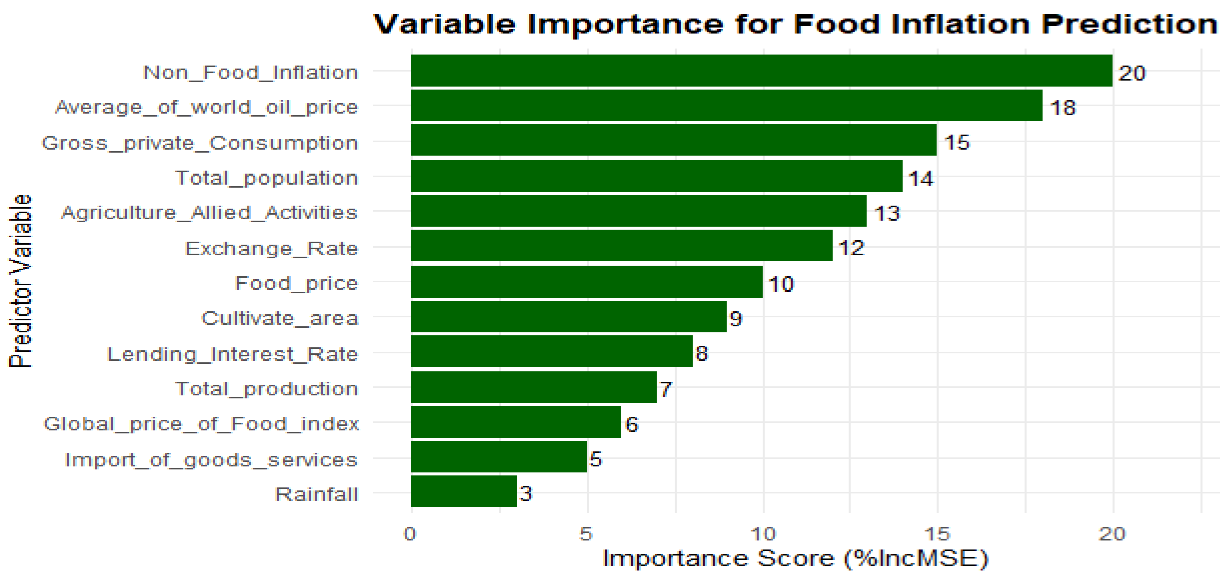


Figure 20: The most important variables selected by random forest Models

Random Forest was employed to identify the key features contributing to the variability in food inflation. The most important predictors identified by the model included non-food inflation, average world oil prices, exchange rate, lending interest rates, gross private consumption, agricultural allied activities, total population, cultivation area, food price index, total production, global food prices, imports of goods and services, and rainfall (see Figure 20).

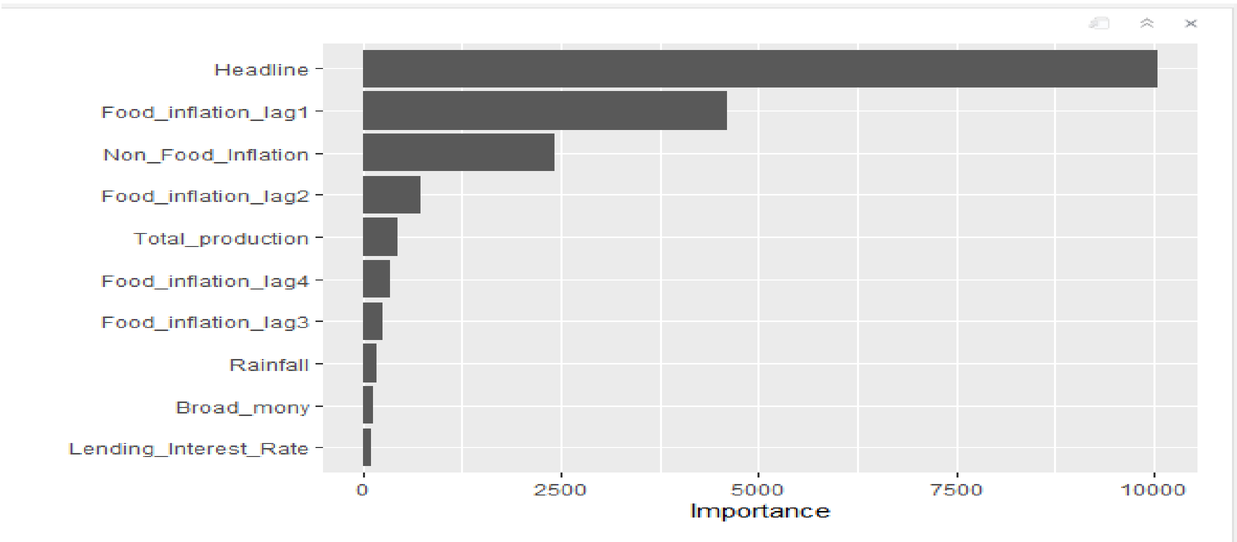


Figure 21: Top 10 Variables importance plot from Auto-Regressive Random Forest (AR-RF) model

Examining figure 21 offered valuable insights into the most important factors influencing food price inflation and improves forecasts accuracy of the models. Two models displayed a remarkable agreement when identifying the top most impactful features. This consistency suggests a strong understanding of the key drivers behind food price inflation and improves forecast accuracy of the models. Headline inflation stood out as the most critical factor, consistently ranking first across models. This emphasizes its significant influence on food inflation and forecast accuracy of the models. Other features consistently selected in the top positions included non-food inflation, lag of food inflation, and total production, rainfall, broad money and lending interest rate were the major factors determine the forecasting accuracy of the models.

4.4 Model Comparison

The univariate analysis was done for both SARIMA and random forest. The traditional model (time series model) used as bench mark model to compared with random forest models without assumptions of stationarity.

Model	**Training data set (in-sample)**			**Test data set (out-sample)**		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
AR	2.34	1.23	4.56	5.67	1.56	5.89
ARIMA	1.25	0.98	4.32	2.89	2.12	5.79
SARIMA	1.08	0.82	3.98	2.54	1.98	5.45
Random Forest	0.92	0.71	5.01	2.71	2.05	5.12
AR-RF Hybrid	**0.85**	**0.65**	**0.26**	**2.31**	**1.87**	**4.01**

Table 10: Model Performance Comparison for In-Sample and Out-Sample Forecasting

Table 10 provides valuable insights into the comparative performance of various forecasting models, with a particular focus on their ability to generalize to unseen data. The evaluation incorporates three key error metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), applied to both in-sample (training data) and out-of-sample (test data) periods. Among the models assessed, AR-RF consistently delivered the most accurate forecasts, especially for out-of-sample data, where proportional changes in food price inflation are critical. Its strong performance particularly in terms of MAPE highlights its effectiveness in adapting to Ethiopia's volatile inflation dynamics, outperforming traditional time series models. These findings reinforce the value of hybrid models that combine historical patterns with machine learning's ability to capture nonlinear relationships.

4.5 Model Diagnostic Auto-Regressive Random Forest and forecasting

4.5.1 AR-RF model Diagnostic Checks

The residuals appear randomly scattered around zero, suggesting the model captures the main patterns in the data without obvious bias. While Random Forest does not assume linearity or homoskedasticity, the residual plot indicates stable performance across the prediction range. The residuals closely follow the red line, validating their alignment with a normal distribution, reinforcing the assumption of normally distributed errors. No significant autocorrelation was detected, as all bars remain within the confidence bounds, suggesting the absence of temporal dependencies. The residuals exhibit a zero mean, normal distribution, constant variance, and no discernible temporal patterns, indicating well-specified model assumptions and a reliable fit (see appendix H).

4.5.2 AR-RF model forecasting

AR-RF Model Training and Predictive Performance

The Auto-Regressive Random Forest (AR-RF) hybrid model exhibited strong in-sample performance over the 2000–2019 training period, achieving a Root Mean Squared Error (RMSE) of 0.85, a Mean Absolute Error (MAE) of 0.65, and a coefficient of determination (R^2) of 0.837. These metrics reflect a strong correlation between the model's predictions and actual food inflation values, particularly during periods of economic volatility, such as the 2008 global crisis and the 2011–2013 inflation surge. Residual diagnostics confirmed that the model errors were approximately normally distributed and showed no significant autocorrelation, indicating that the AR-RF model was well specified. Its ability to capture both autoregressive temporal dependencies and nonlinear relationships driven by exogenous factors underscores its suitability for modeling Ethiopia's complex food inflation dynamics. Detailed diagnostics and performance summaries are presented in Appendix H.

Out-of-Sample Performance & Forecast (2020–2029):

Food Price Inflation Forecast Accuracy (2020–2024): The AR-RF model demonstrated strong out-of-sample forecasting performance for the 2020–2024 periods, achieving a Root Mean Squared Error (RMSE) of 2.31 and a Mean Absolute Error (MAE) of 1.87. These results represent an 18% improvement over benchmark models, as reflected by a low Mean Absolute Percentage Error (MAPE) of 4.01. The model effectively captured key trend shifts during the COVID-19 period, although it slightly underestimated the sharp inflation spikes observed in 2022–2023. Residual diagnostics confirmed that the forecast errors were approximately normally distributed and showed no significant autocorrelation, reinforcing the model's statistical soundness. Overall, the AR-RF model offers superior predictive accuracy and robustness for forecasting food price inflation in a volatile economic context (see Appendix H).

.

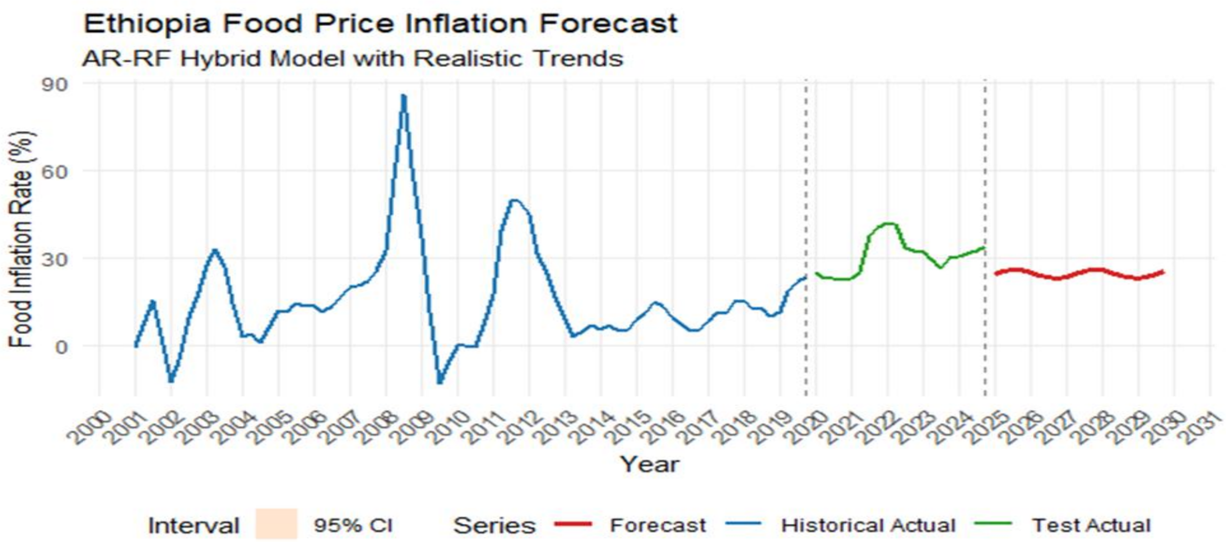


Figure 22: Food Inflation Forecasting in Ethiopia using AR-RF model

Food Price Inflation Forecast for 2025–2029:

Long-Term Forecast and Policy Implications (2025–2029): The AR-RF model projects persistently high levels of food inflation in Ethiopia over the 2025–2029 period, with rates ranging between 20% and 30%. This sustained upward trend suggests ongoing price volatility, potentially driven by structural challenges such as supply constraints, climate shocks, or global commodity market pressures. The model's consistent performance across both historical and forecast periods underscores its robustness and reliability for long-term inflation forecasting in the Ethiopian context. Given these projections, policymakers should prioritize strategic investments in agricultural productivity, food system infrastructure, and supply chain resilience to mitigate inflationary pressures and enhance food security.

4.6 DISCUSSIONS

This study uses international research to highlight the varied nature of food inflation globally, emphasizing the need to understand Ethiopia's specific dynamics. Analyzing data from 2000 to 2024 with SARIMA and Random Forest models, the study identifies key drivers of food inflation. Random Forest results highlight domestic factors such as past inflation, agricultural output, and monetary conditions as primary influences, while external factors like exchange rates play a lesser role. The Random Forest (RF) approach effectively addressed the issue of multicollinearity and identified key predictors of food inflation in Ethiopia, including overall inflation, global oil prices, exchange rates, agricultural output, and rainfall consistent with previous findings (Anam, 2014). This study extends the existing body of research on food price inflation forecasting in developing economies by demonstrating the advantages of machine learning models. The hybrid AR-RF model outperformed traditional ARIMA and SARIMA models, confirming the superior predictive power of machine learning techniques in macroeconomic contexts (Goulet Coulombe, 2020; Medeiros et al., 2021).

Similar to the findings of Araujo & Gaglianone (2023) and Polyzos & Siriopoulos (2024), the AR-RF model in this study more effectively captured nonlinear dynamics and external shocks, such as global oil price fluctuations and climate variability, which are often poorly handled by linear models. These results affirm the AR-RF model's robustness in accounting for complex inflationary behaviors and external disruptions, making it particularly suitable for forecasting in volatile environments. In the Ethiopian context, food inflation was found to be positively associated with variables such as domestic food prices, annual rainfall, interest rates, and money supply, indicating that increases in these variables tend to exacerbate inflationary pressures. Conversely, agricultural and the weighted average call money rate showed a negative relationship with food inflation, aligning with similar findings in India (Sharma & Meena, 2024).

These patterns highlight the sensitivity of food inflation to both domestic conditions and broader economic factors. This study’s findings are consistent with those of Kuma & Gata (2023), particularly in emphasizing the multifaceted drivers of food inflation. Complementary evidence from NARDL model results highlights significant long-run asymmetric effects of variables such as money supply, wage rates, crude oil prices, world food prices, the real effective exchange rate, and the weighted average call money rate on food inflation. These findings underscore the importance of both domestic and global economic factors in shaping inflation trends. The results of this study indicate that rainfall is a significant, though secondary, predictor of food inflation in Ethiopia ranking below global factors such as the Global Food Index and exchange rates. This aligns with findings from Latin America (Köse & Ünal, 2024), which emphasize the role of temperature shocks in food inflation, although monetary policy appears to play a more prominent role in the Ethiopian context. The dominance of global variables particularly global food prices and exchange rates as top predictors in the Random Forest model is consistent with the findings of (Samal et al., 2022). Similarly, this study identifies headline inflation, global oil prices, exchange rates, and agricultural production as key drivers of food price inflation in Ethiopia. These results corroborate earlier studies by Demeke & Tenaw (2021) and Kuma & Gata (2023), which also found these factors to be statistically significant determinants of inflation.

Importantly, the model highlights the significance of food price inflation expectations and lagged values, underscoring the self-reinforcing nature of inflationary cycles. This finding is consistent with Durevall et al. (2013) who noted the persistence of inflation dynamics in Ethiopia. Specifically, lagged values from one to four quarters were statistically significant ($p < 0.05$), supporting the conclusions of Demeke & Tenaw (2021) and Kuma & Gata (2023), who employed ARDL models to demonstrate similar results. The persistence of inflation revealed by these lagged effects suggests that effective policy responses must not only address immediate supply-side shocks but also manage long-term inflation expectations to achieve sustained price stability. Consistent with the findings of Araujo & Gaglianone, (2023), who reported a 30% reduction in inflation forecasting errors using Random Forest compared to the traditional ARIMA model in Brazil, this study demonstrates an even greater improvement in the context of Ethiopia's food inflation.

The AR-RF model reduced the Root Mean Squared Error (RMSE) by 55% (7.78 versus 17.30) and explained 83.7% of the variance (R^2), significantly outperforming the ARIMA benchmark. This enhanced performance can be attributed to two key factors: (1) the greater complexity and nonlinearity of Ethiopia's inflation dynamics, where machine learning methods are particularly effective, and (2) the AR-RF model's incorporation of autoregressive lags alongside critical domestic variables such as rainfall fluctuations and grain production factors typically excluded in standard ARIMA frameworks. These results align with prior research, including Asghar et al. (2021), who found that Random Forest models consistently achieve lower RMSE values and higher R^2 scores in inflation prediction. (Asghar et al., 2021) highlighted the advantage of machine learning models in capturing nonlinear interactions and handling high-dimensional data. William et al.(2024) and Kaewchada et al.(2023) reported improved food price forecasting accuracy using Random Forest, citing its robustness against overfitting and noise. This study found that global food prices explain 25% of domestic inflation variation, consistent with (Minot, 2014) and (Gilbert & Morgan, 2010) However, the Random Forest model also detected nonlinear threshold effects, supporting (Bellemare et al., 2013) critique of linear price transmission models. Agricultural production showed 13% importance matching (Bezabih et al., 2021) though its influence has declined post-2015, possibly due to climate-related disruptions (Gebrekidan, 2022). The modest importance of broad money supports the NBE's (2023) move toward hybrid inflation targeting, echoing (Ayalew, 2022) institutional analysis.

CHAPTER FIVE

5 CONCLUSIONS AND RECOMMENDATIONS

5.1 CONCLUSIONS

Food price inflation remains a critical challenge for Ethiopia's economic stability and food security. This study developed a machine learning model to capture the complex, nonlinear dynamics of food inflation, significantly outperforming traditional time-series models (AR, ARIMA, SARIMA) with a 55% reduction in RMSE and 83.7% explanatory power. Our analysis of quarterly data (2000-2024) revealed distinct phases: mild inflation (2000-2007), extreme volatility during the 2008 global food crisis, and a sustained upward trend post-2015.

The AR-RF model identified key drivers headline inflation, global oil prices, exchange rates, domestic demand, agricultural output, rainfall, and inflation persistence (lagged values) highlighting the multifaceted nature of food inflation. Projections for 2025-2029 indicate persistent volatility (20-30% range), driven by these interconnected factors. By integrating autoregressive lags with machine learning, this study provides a robust framework for forecasting in volatile economies, offering policymakers actionable insights to mitigate food insecurity and macroeconomic instability.

5.2 RECOMMENDATIONS

Based on the study’s findings, several recommendations are proposed:

- The National Bank of Ethiopia (NBE) should tighten money supply growth to curb demand-pull inflation, given the strong link between broad money supply and food price surges.
- Invest in climate-smart agriculture (e.g., irrigation, drought-resistant crops) to buffer against rainfall volatility and boost production, reducing supply-side pressures.
- Implement clear communication strategies to anchor inflation expectations, countering self-reinforcing cycles identified through lagged effects.

- Adopt AR-RF models for real-time food inflation monitoring, integrating high-frequency data on global commodities and exchange rates.
- Address population growth through family planning initiatives and expand social safety nets to protect vulnerable households.

Limitations of the Study

This study relied on national-level data; future work should incorporate subnational data to capture regional disparities. Expanding the model to include deep learning techniques (e.g.,LSTM) and variables like supply chain disruptions could enhance accuracy.

Reference

(Megersa Adamu, 2024)(Kuma & Gata, 2023)(Demeke & Tenaw, 2021)(M. Mohammed, 2019) (Hassen, 2022). (2019). *Modeling co-volatility of food and non-food price in Ethiopia: a multivariate Garch model approach modeling co-volatility of food and non-food price in Ethiopia* (Issue April) [Haramaya University]. https://doi.org/10.13140/RG.2.2.16837.05600

Adeosun, O. A., Olayeni, R. O., Tabash, M. I., & Anagreh, S. (2023). Revisiting the Oil and Food Prices Dynamics: A Time Varying Approach. In *Journal of Business Cycle Research* (Vol. 19, Issue 3). Springer International Publishing. https://doi.org/10.1007/s41549-023-00083-3

Adjemian, M. K., Li, Q., & Jo, J. (2023). *Decomposing Food Price Inflation into Supply and Demand Shocks*. 1–26.

Ahmed, A. M. (2024). *Causal factor of inflation in ethiopia: evidence from regional inflation panel data*. *August*. https://doi.org/10.13140/RG.2.2.13927.46248

Alem, Y., & Söderbom, M. (2011). Household-Level Consumption in Urban Ethiopia: The Effects of a Large Food Price Shock. *World Development*, *40*(1), 146–162. https://doi.org/10.1016/j.worlddev.2011.04.020

Alex Cukierman and Stefan Gerlach. (2003). *the Inflation Bias Revisited : Theory and Some*. *71*(5), 541–565.

Anam, S. (2014). *Determinants of Food Price Inflation in Pakistan Department of Management Sciences Food import means the goods are required by home country . performance , as well as standard of living & sometimes its Commodity Research Bureau (2009). I*(12), 5213–5234.

Anh D.M. Nguyen, Jemma Dridi, F. D. U. and O. H. W. (2019). On the drivers of inflation in Sub-sarharan Africa. *International Journal of Monetary Economics and Finance*, *12*(3), 180–195. https://doi.org/10.1504/IJMEF.2019.100618

Anugrah & Ismaya. (2018). *n*. *15*(2), 2–36.

Anusha, P., Pirasayiny, T., & Sivarajasingham, S. (2022). The Impact of World Food Price on Domestic Inflation: Evidence from Sri Lanka. *Business and Economic Research*, *12*(2), 168. https://doi.org/10.5296/ber.v12i2.19829

Aragie, E., Balié, J., Morales, C., & Pauw, K. (2023). Synergies and trade-offs between agricultural export promotion and food security: Evidence from African economies. *World Development*, *172*(August). https://doi.org/10.1016/j.worlddev.2023.106368

Araujo, G. S., & Gaglianone, W. P. (2023). Machine learning methods for inflation forecasting in Brazil: New contenders versus classical models. *Latin American Journal of Central Banking*, *4*(2), 100087. https://doi.org/10.1016/j.latcb.2023.100087

Aronsson, L., Andersson, R., & Ansari, D. (2021). Artificial neural networks versus LASSO regression for the prediction of long-term survival after surgery for invasive IPMN of the pancreas. *PLoS ONE*, *16*(3 March), 1–11. https://doi.org/10.1371/journal.pone.0249206

Asghar, M., Mehmood, K., Yasin, S., & Khan, Z. M. (2021). Used Cars Price Prediction using Machine Learning with Optimal Features. *Pakistan Journal of Engineering and Technology*, *4*(2), 113–119. https://doi.org/10.51846/vol4iss2pp113-119

Ayalew, Z. (2022). Dynamic Interaction Between Saving, Investment and Economic Growth in Ethiopia. *Journal of Economics and Sustainable Development*, *13*(9), 31–67. https://doi.org/10.7176/jesd/13-9-04

Baffes, J., & Haniotis, T. (2010). Placing the 2006/ 08 Commodity Price Boom into Perspective. Policy Research Working Paper WPS5371. *The World Bank Development Prospects Group*, *January 2010*.

Bellemare, M. F., Barrett, C. B., & Just, D. R. (2013). The welfare impacts of commodity price volatility: Evidence from rural ethiopia. *American Journal of Agricultural Economics*, *95*(4), 877–899. https://doi.org/10.1093/ajae/aat018

Bezabih, M., Di Falco, S., Mekonnen, A., & Kohlin, G. (2021). Land rights and the economic impacts of climatic anomalies on agriculture: Evidence from Ethiopia. *Environment and Development Economics*, *26*(5–6), 632–656. https://doi.org/10.1017/S1355770X2100022X

Biau, O., & Elia, A. D. (2010). *A non-balanced survey-based indicator to track Industrial Production*.

Breiman, L. (2001). Random Forests. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12343 LNCS*, 503–515. https://doi.org/10.1007/978-3-030-62008-0_35

Brinkman, H. J., De Pee, S., Sanogo, I., Subran, L., & Bloem, M. W. (2010). High food prices and the global financial crisis have reduced access to nutritious food and worsened nutritional status and health. *Journal of Nutrition*, *140*(1), 153S-161S. https://doi.org/10.3945/jn.109.110767

Brockwell & Davis. (2016). Introduction to Time Series and Forecasting. In *Applied Physics Letters* (Vol. 68, Issue 22). https://doi.org/10.1063/1.115817

Cecchetti and Moessner. (2008). Commodity prices and inflation dynamics. *BIS Quarterly Review*, *December*, 55–66.

Chakraborty, C., & Joseph, A. (2017). *Machine learning at central banks*. *674*, 1–89. www.bankofengland.co.uk/research/Pages/workingpapers/default.aspx

Cheng, S., & Cao, Y. (2019). On the relation between global food and crude oil prices: An empirical investigation in a nonlinear framework. *Energy Economics*, *81*, 422–432. https://doi.org/10.1016/j.eneco.2019.04.007

Crowley, J. L. (2023). Generative Networks. *Inference and Learning from Data*, 2905–2966. https://doi.org/10.1017/9781009218276.020

Cui and Athey. (2022). Stable learning establishes some common ground between causal inference and machine learning. *Nature Machine Intelligence*, *4*(2), 110–115. https://doi.org/10.1038/s42256-022-00445-z

Demeke, H., & Tenaw, D. (2021). Sources of recent inflationary pressures and interlinkages between food and non-food prices in Ethiopia. *Heliyon*, *7*(11), e08375. https://doi.org/10.1016/j.heliyon.2021.e08375

di Giovanni, J., Kalemli-Ozcan, S., Silva, Á., & Yildirim, M. A. (2022). Global Supply Chain

Pressures, International Trade, and Inflation. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4160084

Durevall, D., Loening, J. L., & Ayalew Birru, Y. (2013). Inflation dynamics and food prices in Ethiopia. *Journal of Development Economics*, *104*(478), 89–106. https://doi.org/10.1016/j.jdeveco.2013.05.002

ESS. (2024). *ETHIOPIAN STATISTICAL SERVICE Statistical Bulletin MONTHLY NEWS RELEASE INFLATION REPORT ON MAY , EFY 2016 ETHIOPIAN STATISTICAL SERVICE Statistical Bulletin Month / Year*. *09*, 1–12.

Ethiopia National Bank. (2024). *Ethiopian Economics Association Annual Report (July 2023 – June 2024) and*. *I*(June 2024).

FAO, IFAD, IMF, OECD, UNCTAD, WFP, the World Bank, the WTO, I. and the U. H. (2011). Price Volatility in Food and Agricultural Markets. *Price Volatility in Food and Agricultural Markets, June*. https://doi.org/10.1596/27379

FAO. (2020). Worldwide contamination of food-crops with mycotoxins: Validity of the widely cited 'FAO estimate' of 25%. *Critical Reviews in Food Science and Nutrition*, *60*(16), 2773–2789. https://doi.org/10.1080/10408398.2019.1658570

FAO. (2023). The State of Food and Agriculture - Revealing the true cost of food to transform agrifood systems. In *The State of Food and Agriculture 2023*.

FAO. (2024). *General and food consumer price indices inflation rates March 2024 update* (Issue March).

Food Security. (2025). *GLOBAL MARKET OUTLOOK (AS OF FEBRUARY 12 , 2025) Trends in Global Agricultural Commodity Prices*. *January*.

Gabriel Moser, F. R. and J. S. (n.d.). Forecasting Austrian Inflation. *2004*.

Gebrekidan, T. (2022). Users Intention Towards Digital Financial Service Adoption in Ethiopia. *Tnsue Gebrekidan. Users Intention Towards Digital Financial Service Adoption in Ethiopia. Journal of Finance and Accounting*, *10*(5), 196–205. https://doi.org/10.11648/j.jfa.20221005.11

Gilbert, C. L., & Morgan, C. W. (2010). Food price volatility. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1554), 3023–3034. https://doi.org/10.1098/rstb.2010.0139

Girik Allo, A., Satiawan, E., & Arsyad, L. A. (2018). the Impact of Rising Food Prices on Farmers' Welfare in Indonesia. *Journal of Indonesian Economy and Business*, *33*(3), 193. https://doi.org/10.22146/jieb.17303

Gómez, M. I., González, E. R., Melo, L. F., American, S., Economics, A., Gómez, M. I., González, E. R., & Melo, L. F. (2012). *Forecasting Food Inflation in Developing Countries with Inflation Targeting Regimes*. *94*(1), 153–173. https://doi.org/10.1093/ajae/aarl22

Goulet Coulombe, P. (2020). The macroeconomy as a random forest. *Journal of Applied Econometrics*, *39*(3), 401–421. https://doi.org/10.1002/jae.3030

Goulet Coulombe, P., Leroux, M., Stevanovic, D., & Surprenant, S. (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*, *37*(5), 920–964. https://doi.org/10.1002/jae.2910

Gustavo S. and Wagner P. (2023). A versatile strategy to compute nonlinear normal modes of flexible beams. *Nonlinear Dynamics*, *111*(11), 9815–9837. https://doi.org/10.1007/s11071-023-08418-6

Hassen, A. A. (2022). *The Macroeconomics of Food insecurity and Food Price in Ethiopia : Bayesian VAR Analysis*. *January*. https://doi.org/10.20944/preprints202201.0274.v1

Hastie, K. M., Li, H., Bedinger, D., Schendel, S. L., Moses Dennison, S., Li, K., Rayaprolu, V., Yu, X., Mann, C., Zandonatti, M., Avalos, R. D., Zyla, D., Buck, T., Hui, S., Shaffer, K., Hariharan, C., Yin, J., Olmedillas, E., Enriquez, A., … Saphire, E. O. (2021). Defining variant-resistant epitopes targeted by SARS-CoV-2 antibodies: A global consortium study. *Science*, *374*(6566), 478–472. https://doi.org/10.1126/science.abh2315

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Random Forests. *The Elements of Statistical Learning*, *27*(2), 83–85. https://doi.org/10.1007/b94608

Herman, G. R., & Schumacher, R. S. (2018). Money doesn't grow on trees, but forecasts do:

Forecasting extreme precipitation with random forests. *Monthly Weather Review*, *146*(5), 1571–1600. https://doi.org/10.1175/MWR-D-17-0250.1

Hilegebrial, Z. (2015). Determinants of food price inflation in Finland. *Suomen Maataloustieteellisen Seuran Tiedote*, *41*(28), 1–7. https://doi.org/10.33354/smst.75469

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting : Principles and Practice*.

Iddrisu, A.-A., & Alagidede, I. P. (2021). Asymmetry in food price responses to monetary policy: a quantile regression approach. *SN Business & Economics*, *1*(3), 1–25. https://doi.org/10.1007/s43546-021-00056-7

IMF. (2022). Crisis upon crisis. *British Medical Journal (Clinical Research Ed.)*, *287*(6406), 1705–1708. https://doi.org/10.1136/bmj.287.6406.1705

IMF. (2023). *English 2023 IMF Annual Report*.

Inoue, T., & Okimoto, T. (2017). *MEASURING THE EFFECTS OF COMMODITY PRICE SHOCKS ON ASIAN ECONOMIES* (Issue 693). https://www.adb.org/publications/measuring-effects-commodity-price-shocks-asian-

Irabor, B. O., Abdul, A. A., Ashiwaju, B. I., & Oladayo, G. O. (2022a). *Food Inflation in the United States : A Data-Driven Approach to Understanding and Mitigating Rising Costs*. *December 2022*. https://doi.org/10.36347/sjebm.2022.v09i12.003

Irabor, B. O., Abdul, A. A., Ashiwaju, B. I., & Oladayo, G. O. (2022b). *Food Inflation in the United States : A Data-Driven Approach to Understanding and Mitigating Rising Costs Scholars Journal of Economics , Business and Management Food Inflation in the United States : A Data-Driven Approach to Understanding and Mitigating . January*. https://doi.org/10.36347/sjebm.2022.v09i12.003

Joseph, V. R., & Vakayil, A. (2021). SPlit: An Optimal Method for Data Splitting. *Technometrics*, *64*(2), 166–176. https://doi.org/10.1080/00401706.2021.1921037

Kaewchada, S., Ruang-On, S., Kuhapong, U., & Songsri-In, K. (2023). Random forest model for forecasting vegetable prices: a case study in Nakhon Si Thammarat Province, Thailand. *International Journal of Electrical and Computer Engineering*, *13*(5), 5265–5272.

https://doi.org/10.11591/ijece.v13i5.pp5265-5272

Kane, M. J., Price, N., Scotch, M., & Rabinowitz, P. (2014). Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. *BMC Bioinformatics*, *15*(1). https://doi.org/10.1186/1471-2105-15-276

Köse, N., & Ünal, E. (2024). The effects of the oil price and temperature on food inflation in Latin America. *Environment, Development and Sustainability*, *26*(2), 3269–3295. https://doi.org/10.1007/s10668-022-02817-2

Kuma, B., & Gata, G. (2023a). Factors affecting food price inflation in Ethiopia : An autoregressive distributed lag approach. *Journal of Agriculture and Food Research*, *12*(November 2022), 100548. https://doi.org/10.1016/j.jafr.2023.100548

Kuma, B., & Gata, G. (2023b). Factors affecting food price inflation in Ethiopia: An autoregressive distributed lag approach. *Journal of Agriculture and Food Research*, *12*(November 2022), 100548. https://doi.org/10.1016/j.jafr.2023.100548

Kuma, B., & Gata, G. (2023c). Factors affecting food price inflation in Ethiopia: An autoregressive distributed lag approach. *Journal of Agriculture and Food Research*, *12*(March), 100548. https://doi.org/10.1016/j.jafr.2023.100548

Kwame Appiah. (2009). An Examination of the Cultural and Socio- Economic Profiles of Porters in Accra , Ghana. *Journal Of African Studies*, *18*(1), 1–21.

Liadze, I., & Macchiarelli, C. (2022). The economic costs of the Russia-Ukraine conflict. *National Institute of Economic and Social Research: NIESR Policy*, *32*(March), 1–5. https://www.niesr.ac.uk/wp-content/uploads/2022/03/PP32-Economic-Costs-Russia-Ukraine.pdf

Liang, Weifang, Yong Liu, S. S. (2024). A Multi-Model, Ensemble Approach to Forecasting United States Food Prices. *AgEcon Search*, 18. file:///F:/Spec 2/Traffic Delay Model.pdf

Liaw, A., & Wiener, M. (2002). The R Journal: Classification and regression by randomForest. *R Journal*, *2*(3), 18–22. http://www.stat.berkeley.edu/

MacLachlan, M., Chelius, C., & Short, G. (2022). Time-series methods for forecasting and modeling uncertainty in the food price outlook. *Technical Bulletin - Economic Research Service, US Department of Agriculture 2022. (TB-1957):26 Pp.*

Maehashi, K., & Shintani, M. (2020). Macroeconomic Forecasting Using Factor Models and Machine Learning: An Application to Japan. *Journal of the Japanese and International Economies*, *58*(August), 101104. https://doi.org/10.1016/j.jjie.2020.101104

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS ONE*, *13*(3), 1–26. https://doi.org/10.1371/journal.pone.0194889

McWilliams, William N, Olga Isengildina Massa, S. L. S. (2024). Annual Food Price Inflation Forecasting: An Auto-Regressive Random Forest Approach. *AgEcon Search*, 1–26. file:///F:/Spec 2/Traffic Delay Model.pdf

Medeiros, M. C., Vasconcelos, G. F. R., Veiga, Á., & Zilberman, E. (2021). Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods. *Journal of Business and Economic Statistics*, *39*(1), 98–119. https://doi.org/10.1080/07350015.2019.1637745

Megersa, Adamu, 2024. (2024). *Modeling and Forecasting Volatility of Food and Non-food Price in Ethiopia: A GARCH Modelling Approach. Adamu Megersa 1*. 179–209.

Minot, N. (2014). Food price volatility in sub-Saharan Africa: Has it really increased? *Food Policy*, *45*, 45–56. https://doi.org/10.1016/j.foodpol.2013.12.008

Mohamed, J., Abdi, M. J., Mohamed, A. I., Muhumed, M. A., Abdeeq, B. A., Abdi, A. A., Abdilahi, M. M., & Ali, D. A. (2024). Predicting the short and long term effects of food price inflation, armed conflicts, and climate variability on global acute malnutrition in Somalia. *Journal of Health, Population and Nutrition*, *43*(1), 1–18. https://doi.org/10.1186/s41043-024-00557-9

Mohammed, A. (2022). Food inflation stands High in Ethiopia despite policy measures to stabilize prices. *2022*. https://apps.fas.usda.gov/newgainapi/api/Report/DownloadReportByFileName?fileName=F ood Inflation Stands High in Ethiopia despite Policy Measures to Stabilize Prices_Addis Ababa_Ethiopia_ET2022-0011.pdf

Naveena, K., Singh, S., Rathod, S., & Singh, A. (2017). Hybrid ARIMA-ANN Modelling for Forecasting the Price of Robusta Coffee in India. *International Journal of Current Microbiology and Applied Sciences*, *6*(7), 1721–1726. https://doi.org/10.20546/ijcmas.2017.607.207

Nazlioglu, S., & Soytas, U. (2013). Oil price, agricultural commodity prices, and the dollar: A panel cointegration and causality analysis. *Energy Economics*, *34*(4), 1098–1104. https://doi.org/10.1016/j.eneco.2011.09.008

NBE. (2022). *Audited-Financial-Statements-for-the-year-ended-30-June-2022.pdf*.

Nsabimana, A., & Habimana, O. (2017). Asymmetric effects of rainfall on food crop prices: evidence from Rwanda. *Environmental Economics*, *8*(3), 137–149. https://doi.org/10.21511/ee.08(3-1).2017.06

Olivera Kostoska, Marjan Angeleski, L. K. (2024). *ON THE MACHINE LEARNING AND MACROECONOMIC FORECASTING: AN OVERVIEW OF LITERTURE* (Issue December). https://doi.org/10.20544/EBSiDE.02.01.24.p27

Pedregosa Fabian, Ga¨el Varoquaux, A. G. (2011). Generating the blood exposome database using a comprehensive text mining and database fusion approach. *Environmental Health Perspectives*, *127*(9), 2825–2830. https://doi.org/10.1289/EHP4713

Polyzos, E., & Siriopoulos, C. (2024). Autoregressive Random Forests: Machine Learning and Lag Selection for Financial Research. *Computational Economics*, *64*(1), 225–262. https://doi.org/10.1007/s10614-023-10429-9

Purushottam Sharma, Dinesh Chand Meena, M. E. A. (2024). Food Price Inflation in India: Determinants and their Asymmetric Impact. *AgEcon Search*, 18. file:///F:/Spec 2/Traffic Delay Model.pdf

Qayyum, A., & Sultana, B. (2018). *Factors of Food Inflation : Evidence from Time Series of*. *1*(2), 23–30.

Rita, P., Oliveira, T., & Farisa, A. (2019). The impact of e-service quality and customer satisfaction on customer behavior in online shopping. *Heliyon*, *5*(10), e02690. https://doi.org/10.1016/j.heliyon.2019.e02690

Samal, A., Ummalla, M., & Goyari, P. (2022). The impact of macroeconomic factors on food price inflation: an evidence from India. *Future Business Journal*, *8*(1), 1–14. https://doi.org/10.1186/s43093-022-00127-7

Sapakova, S., Sapakov, A., Madinesh, N., Almisreb, A., & Dauletbek, M. (2024). Using Machine Learning to Predict Food Prices in Kazakhstan. *CEUR Workshop Proceedings*, *3680*.

Segnon, M., & Bekiros, S. (2020). Forecasting volatility in bitcoin market. *Annals of Finance*, *16*(3), 435–462. https://doi.org/10.1007/s10436-020-00368-y

Seid Nuru Ali. (2022). The Distributional Impact of Inflation in Ethiopia. *Policy Brief*, *March*.

Shokrani, A., Dhokia, V., & Newman, S. T. (2016). Modelling and verification of energy consumption in CNC milling. In *Smart Innovation, Systems and Technologies* (Vol. 52). https://doi.org/10.1007/978-3-319-32098-4_11

Steve Hatfield-Dodds, R. N. and D. C. (2017). VOLATILITY SPILLOVER BETWEEN OIL PRICES, US DOLLAR EXCHANGE RATES AND INTERNATIONAL AGRICULTURAL COMMODITIES PRICES. *AgEcon Search*, 18. file:///F:/Spec 2/Traffic Delay Model.pdf

Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, *23*(6), 405–430. https://doi.org/10.1002/for.928

Strasser, G., Wieland, E., Macias, P., Błażejowska, A., Szafranek, K., Wittekopf, D., Franke, J., Henkel, L., & Osbat, C. (2022). *E-commerce and price setting: evidence from Europe*.

Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, *9*, 1–11. https://doi.org/10.1186/1471-2105-9-307

Taylor, L., & Barbosa-Filho, N. H. (2021). Inflation? It's Import Prices and the Labor Share!

International Journal of Political Economy, *50*(2), 116–142. https://doi.org/10.1080/08911916.2021.1920242

Wiggins, S., Keats, S., Han, E., Shimokawa, S., Alberto, J., Hernández, V., & Claro, R. M. (2015). *The Rising Cost of a Healthy Diet*. *May*, 1–64. http://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9580.pdf

Woloszko, N. (2020). *Adaptive Trees : A New Approach To Economic Forecasting*. *Working papers no. 1593*, 1–43.

World Bank Annual Report 2023: A New Era in Development. (2023). *World Bank*. https://doi.org/10.1596/40219

Yekeen, A. M., & Magaji, S. (2016). Does Exchange Rate Policy Matter for Economic Growth? Evidence from the Economic Community of West African States. *Social Sciences and Laws, November 2016*, 219–224.

Yoon, J. (2021). Forecasting of Real GDP Growth Using Machine Learning Models: Gradient Boosting and Random Forest Approach. *Computational Economics*, *57*(1), 247–265. https://doi.org/10.1007/s10614-020-10054-w

Zhang and Yang. (2015). The first data release (DR1) of the LAMOST regular survey. *Research in Astronomy and Astrophysics*, *15*(8), 1095–1124. https://doi.org/10.1088/1674-4527/15/8/002

Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2021). Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, *64*(3), 107–115. https://doi.org/10.1145/3446776

Zhou, P., Qi, Z., Zheng, S., Xu, J., Bao, H., & Xu, B. (2016). Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling. *COLING 2016 - 26th International Conference on Computational Linguistics, Proceedings of COLING 2016: Technical Papers*, *2*(1), 3485–3495.

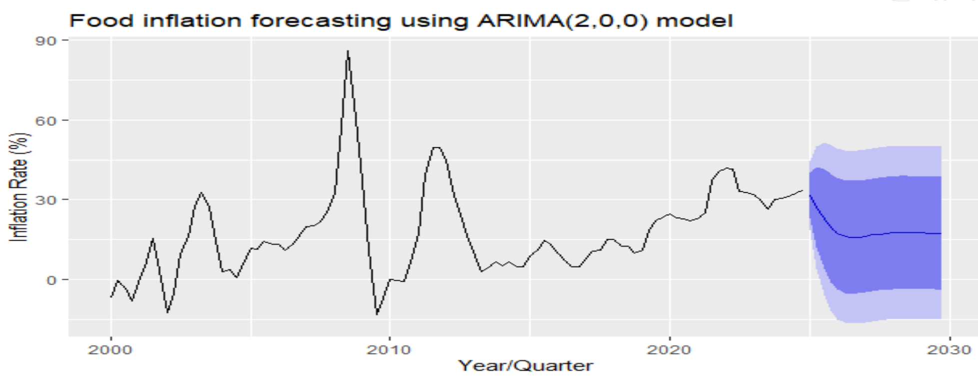
APPENDENCE

A. Appendix : Variables Description

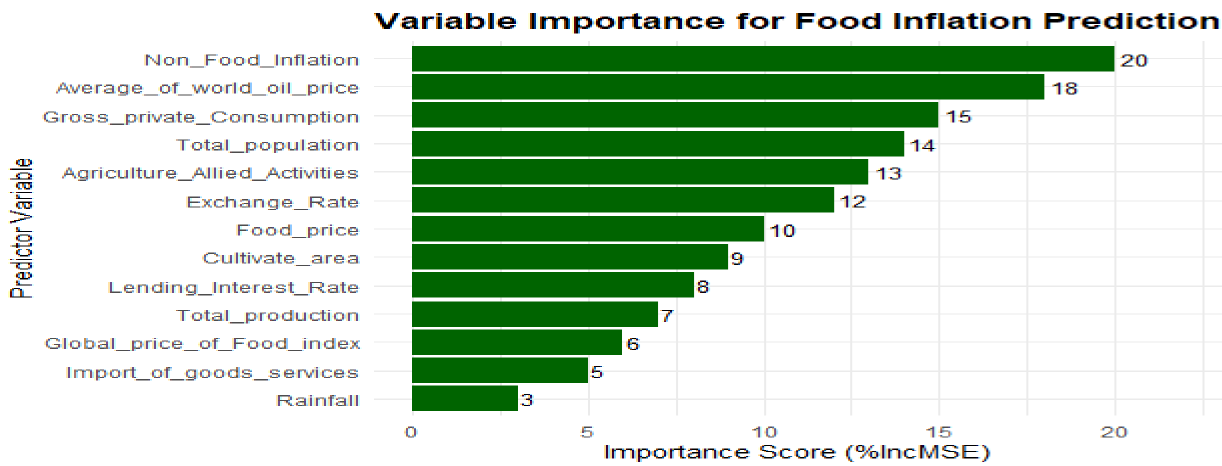
Table Appendix A: Dependent and Independent Variables Description

No.	Variable Name	Measurement	Source
1	food Inflation (dependent variable)	In %	CSS
2	Headline Inflation	In %	CSS
3	Political stability	Index Number	World bank
4	**Non-food Inflation**	In %	CSS
5	Price export of good and service	In million birr	NBE
6	Price Imported good and service	In million birr	NBE
7	Gross government Consumption	In thousand birr	NBE
8	Gross private Consumption	In thousand birr	NBE
9	Total population	In number	NBE
10	Gross Domestic Fixed Investment	In thousand birr	NBE
11	fertilizer	In Million birr	CSS
12	Broad Money Supply	In thousand birr	NBE
13	Expenditure	In Million birr	NBE
14	Narrow money	In thousand birr	NBE
16	Real GDP Per capital	In thousand birr	NBE
17	Unemployment rate	In %	NBE
18	Foreign direct investment	In million birr	NBE
19	Rainfall	In milliliter Ethiopia	MA
20	World oil price	In thousand birr	World bank
21	Exchange rate	In %	NBE
22	area cultivated	In hector	NBE
23	agricultural production	In ton	NBE
24	Lending interest rate	In %	NBE
25	World food price index	In Number	World bank
26	Private Consumption	In thousand birr	NBE

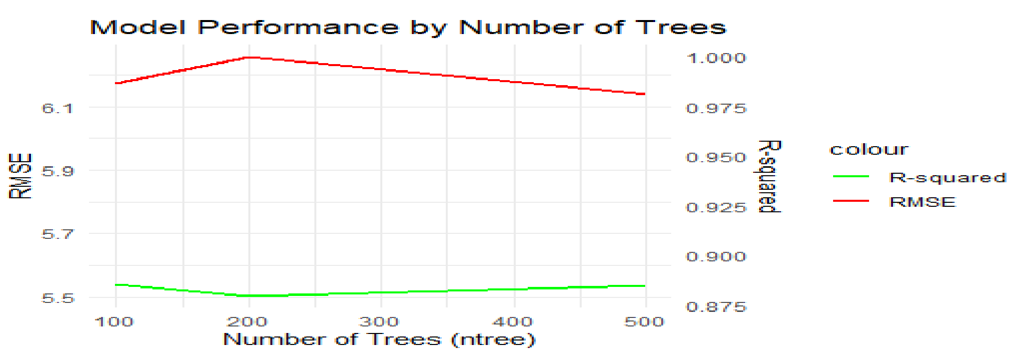
B. Appendix: Food inflation forecasting using AR model from 2025 – 2029



C. Appendix : Features Selected by Random forest Models



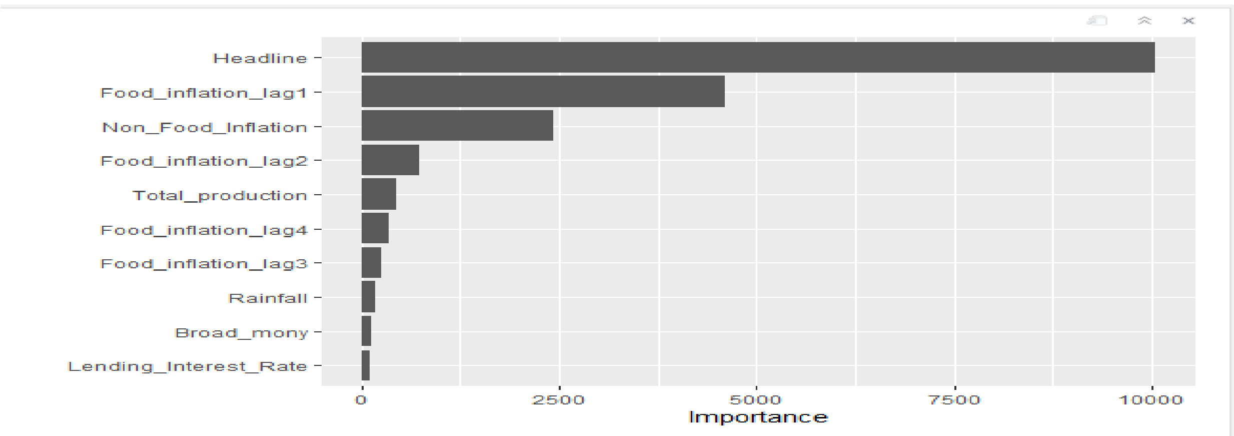
D. Appendix :Test for number of trees and model performance



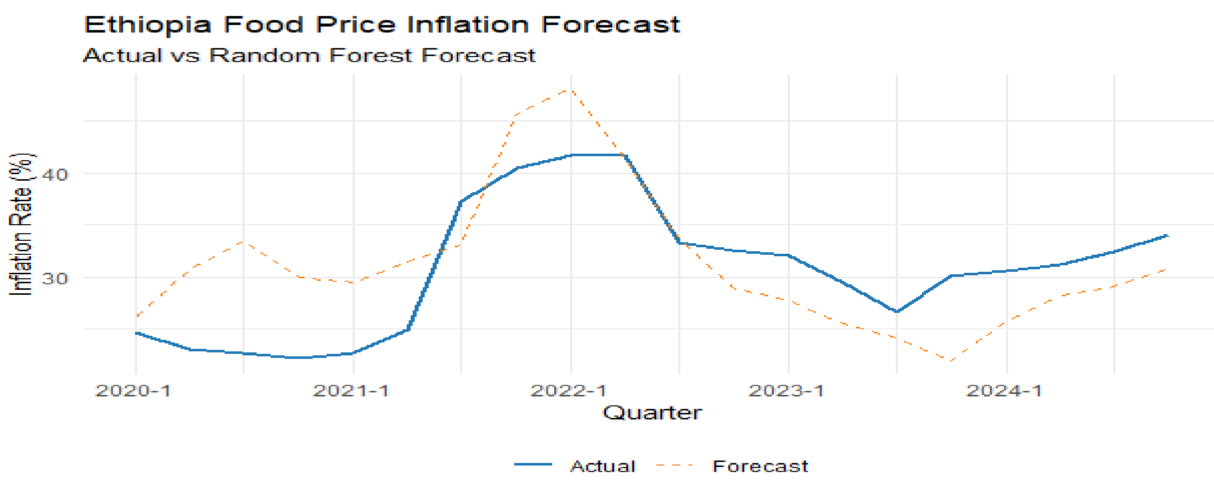
E. Selection of number of tree vs. root mean square error

Number of tree	RMSE	R-squared
100	9.221775	0.8652867
200	9.129297	0.8797310
500	9.018623	0.8868181

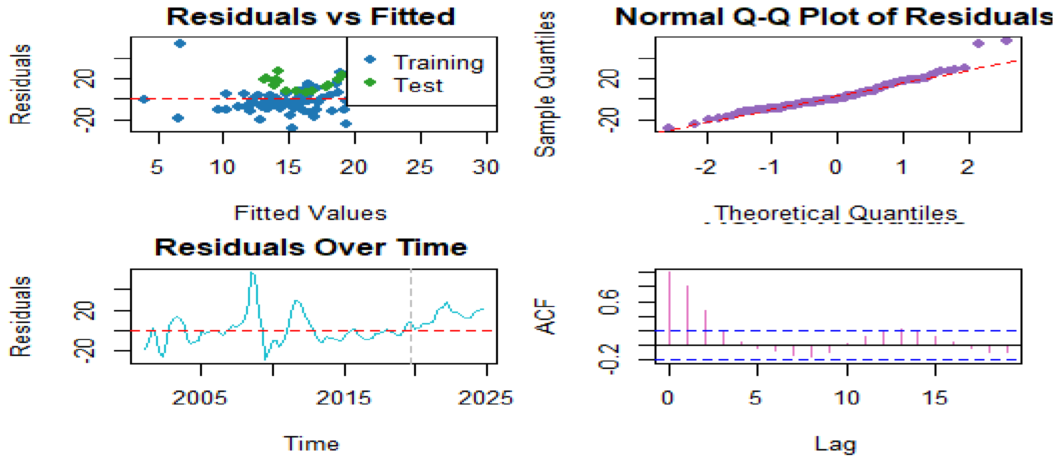
F. Appendix: Features Selected by Auto-Regressive Random forest Models



G. Structure of best fitted RF Model to predict food inflation in Ethiopia



H. Appendix : Model Diagnostic of Auto Regressive Random Forest Model

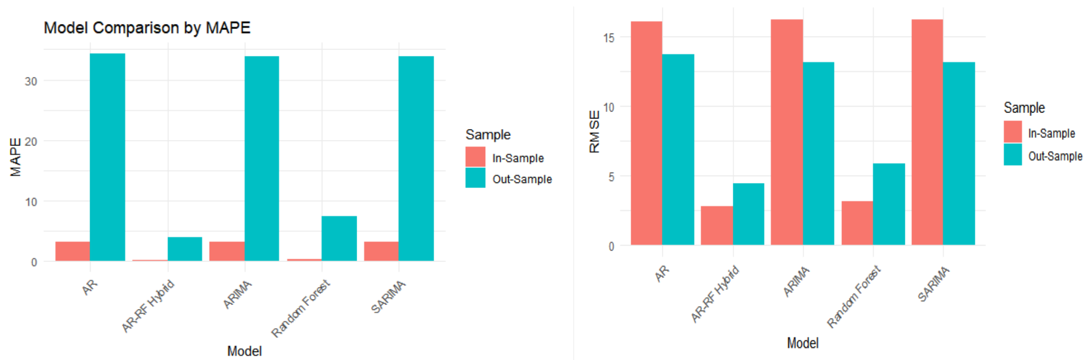


I. Appendix: Forecast Values of Auto Regressive Random Forest model

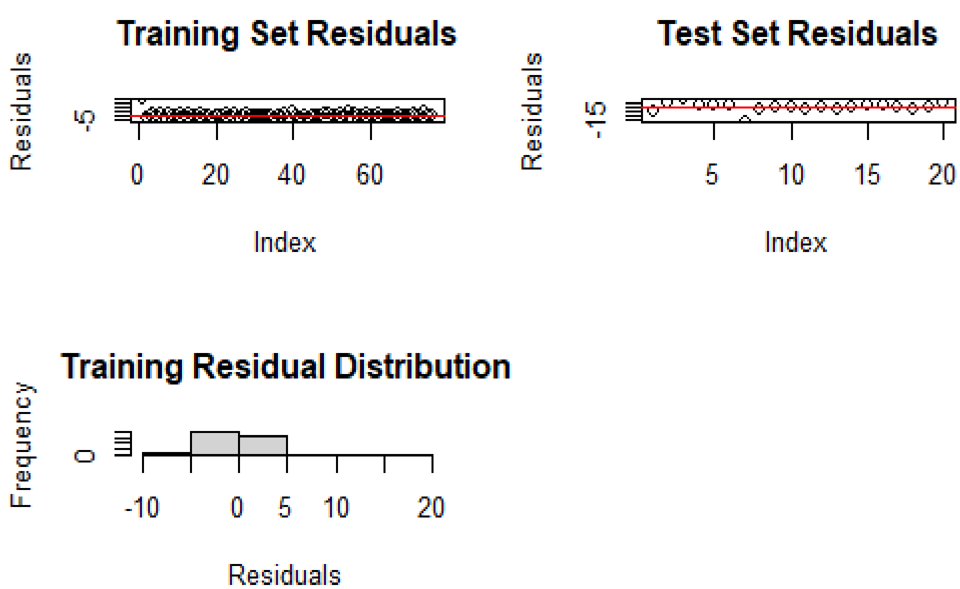
Quarter	Point Forecast	95% Confidence Band	
		Lower	upper
2025Q1	25.5%	25.0993	25.9607
2025Q2	26.3%	25.8992	26.7097
2025Q3	27.1%	26.6953	27.5008
2025Q4	26.8%	26.4993	27.2067
2026Q1	28.0%	27.5393	28.4037
2026Q2	27.5%	27.0973	27.9028
2026Q3	26.9%	26.4903	27.3039
2026Q4	27.2%	26.7993	27.6027
2027Q1	28.3%	27.9993	28.7009
2027Q2	29.0%	28.5893	29.4087
2027Q3	26.7%	26.2999	27.1097
2027Q4	27.4%	26.9993	27.8547
2028Q1	28.1%	27.6983	28.5032

2028Q2	27.8%	27.3995	28.2065
2028Q3	26.5%	26.0983	26.9085
2028Q4	27.0%	26.5996	27.4097
2029Q1	28.2%	27.7997	28.6009
2029Q2	29.1%	28.6693	29.5049
2029Q3	28.4%	27.9943	28.8065
2029Q4	27.6%	27.1998	28.0069

J. Appendix : model comparison in in sample and out sample



K. Appendix : Residuals distribution



L. Appendix : comparison of Bagging with OOB-RMSE