Project Report – CSCI 6612 Visual Analytics

Predictive Analysis of Seasonal Trends and Inflation in Food Prices (PASTIFP)

Team Members

Saransh Kotha sr384758@dal.ca

Malaika Pandit ml234831@dal.ca

Md Mukarrom Hossain md323976@dal.ca

Table of Contents

1. Introduction....3
1.1. Background Information....3
1.2. Problem Statement....3
1.3. Project Significance....3
1.4. Project Objective....4
1.5. Project Scope....4
1.5.1. Predictive Modeling....4
1.5.2. Policy Support....4
1.5.3. Out of Scope....4
2. Methodology....5
2.1. Data Collection and Preparation....5
2.1.1. Data Cleaning....6
2.1.2. Outlier Detection and Removal....6
2.1.3. Data Transformation....6
2.1.4. Exploratory Data Analysis (EDA)....8
2.1.5. Key Insights from EDA....10
2.2. Model Selection....11
2.2.1. Selection Approach....11
2.2.2. Model Training and Hyperparameter Tuning....11
2.3. Model Evaluation....11
2.3.1. Evaluation Metrics....11
2.3.2. Validation Strategy....11
2.3.3. Comparative Performance....12
3. Results....12
3.1. Model Performance Analysis....12
3.2. Performance Metrics and Comparisons....12
3.2.1. Ridge Regression....12
3.2.2. Random Forest (Fine-Tuned)....13
3.2.3. XGBoost....13
3.2.4. Lasso Regression....13
3.2.5. ARIMA....13
3.2.6. Gradient Boosting....13
3.3. Interpretation of Results....14
3.4. Implications....14

3.5. Limitations 15
4. Integration of Interactive Dashboard 15
4.1. Time-Series Visualization 15
4.2. Model Comparisons 15
4.3. Scenario Analysis 15
4.4. Key Takeaways from Dashboard Integration 15
5. Conclusion 16
5.1. Summary of Key Findings 16
5.2. Achievement of Objectives 16
6. References 18

1. Introduction

1.1. Background Information

The global food supply chain is a critical component of economic and social stability, influencing the affordability and accessibility of food for billions of people (Asadollah, Carmy, Hoque, & Yilmazkuday, 2024). However, food prices are highly sensitive to a wide range of factors, including seasonal variations, geopolitical events, supply chain disruptions, and global economic conditions. This volatility has far-reaching implications, particularly for developing economies where food expenditures form a substantial part of household budgets (Furceri, Loungani, Simon, & Wachter, 2023). Food price inflation affects economic growth, makes it harder for policymakers to stabilize markets, and makes food insecurity worse. Developing successful mitigation techniques requires an understanding of the underlying seasonal and structural patterns and trends in food price rise.

1.2. Problem Statement

Both consumers and politicians face serious difficulties as a result of the ongoing volatility in food prices worldwide. Unpredictable price behavior is caused by seasonal variations, economic upheavals, and inflationary pressures, which disproportionately affect economies that depend on agricultural imports and disadvantaged populations. Although the stability of food prices is vital, little is known about the complex trends and factors that influence these swings.

By examining seasonal patterns and inflation in food prices, this project aims to close this gap by identifying the main causes of price volatility and creating predictive models that will help predict future trends. By doing this, the project hopes to assist those involved in public policy, economics, and agriculture in making better decisions.

1.3. Project Significance

The issue of food price volatility is profoundly significant due to its extensive economic, social, and policy implications. Rising and unpredictable food prices exacerbate food insecurity, especially in low- and middle-income countries where a large proportion of household income is allocated to food (Kalkuhl, von Braun, & Torero, 2016). These inflationary pressures disproportionately impact vulnerable populations, pushing millions into poverty and hunger. For governments, addressing such fluctuations is essential for maintaining economic stability, as volatile food markets hinder effective planning and equitable distribution. Policymakers require reliable insights to implement timely interventions that stabilize markets and mitigate the impact of inflation on low-income households (Arezki, El Aynaoui, Nyarko, & Teal, 2016).

Understanding the volatility of food prices not only helps with short-term problems but also facilitates long-term solutions in economic policy and agricultural planning. In order to maintain a balance between supply and demand, farmers can use predictive information to improve production plans and expedite supply chain operations. To stabilize prices and safeguard consumers, policymakers can create targeted subsidies, rules, and actions based on data-driven results.

Additionally, this project highlights the transformative role of technology in solving real-world economic challenges, showcasing how predictive analytics and data visualization can inform decisions that promote food security and sustainable development (Acara Climate, 2024).

1.4. Project Objective

The primary objective of this project is to "*Analyze and predict food price inflation trends using historical data, enabling stakeholders to make informed decisions that promote economic stability and food security*". By identifying seasonal patterns, economic disruptions, and inflationary drivers, the project seeks to uncover the key factors contributing to food price volatility.

Specific objectives include:

- To look into how seasonal changes affect the inflation of food prices and spot reoccurring trends.
- To create prediction models that can precisely predict trends in food price inflation, assisting stakeholders and policymakers in making proactive plans.
- To offer practical advice for streamlining supply chain operations and agricultural production schedules based on anticipated price patterns.
- To educate decision-makers by presenting evidence-based suggestions for focused subsidies, market interventions, and regulatory measures that maintain food price stability.
- To show how well data visualization and predictive analytics can be used to tackle difficult economic issues like rising food prices.

By achieving these objectives, the project aims to contribute to sustainable agricultural development, strengthen food security, and enhance the ability of governments and organizations to navigate the complexities of global food markets.

1.5. Project Scope

The scope of this project is centered on analyzing and predicting food price inflation using historical data to support decision-making for economic stability and food security. The project will focus on two key areas:

1.5.1. Predictive Modeling

Developing advanced forecasting models to predict food price inflation, incorporating factors such as seasonal patterns and lagged dependencies. These models aim to provide accurate and timely predictions that account for historical trends and potential future fluctuations. The analysis will include data preprocessing, exploratory data analysis (EDA), feature engineering, and the application of machine learning algorithms to enhance predictive accuracy.

1.5.2. Policy Support

producing useful insights to help stakeholders and policymakers deal with the volatility of food prices. In order to inform regulatory policies, targeted subsidies, and actions that stabilize food prices and lessen their impact on vulnerable groups, it is necessary to determine seasonal trends, inflationary drivers, and lagged impacts.

1.5.3. Out of Scope

In order to effectively explain findings, the project will encompass predictive modeling, historical data analysis, and result visualization. It does not, however, entail direct interaction with supply chain managers, farmers, or legislators. Although the analysis's conclusions can help guide decision-making, working with industry stakeholders to put these suggestions into practice is outside the project's purview. The inflation of food prices is the exclusive subject of this research; it will not examine more general macroeconomic variables such as trade policy, currency exchange rates, or overall economic growth. With the exception of the larger economic environment that affects inflation on a national or international scale, the analysis will focus on trends in food prices.

By defining these boundaries, the project aims to set clear expectations while providing valuable insights into food price inflation trends and their implications for economic stability and food security.

2. Methodology

In this project, the problem being addressed is a regression task. The goal is to predict continuous food price inflation values based on historical data. Unlike classification tasks, which predict discrete categories, regression aims to forecast a continuous outcome, in this case, food price inflation, a variable that can take any numeric value. By applying regression techniques, our aim is to model the relationship between food price indices, and other relevant features, enabling accurate predictions of future price trends.

2.1. Data Collection and Preparation

The data for this project was collected from two primary sources: the World Bank's World Development Indicators (https://databank.worldbank.org/source/world-development-indicators) and the Microdata Catalog (https://microdata.worldbank.org/index.php/catalog/4509/data-api). These sources provide comprehensive datasets on various economic indicators, including food price indices, inflation rates, and other relevant variables for multiple countries over several years. The data encompasses monthly observations of food price trends, inflation rates, and related economic indicators. The dataset also includes country-level details such as ISO codes, timestamps (year and month), and inflation-adjusted price indices

The two datasets were merged based on common columns: 'Country Name' and 'Year', using a left join. This resulted in a merged dataset with 22 features and 6,203 data instances, spanning 17 years (2007–2023) and covering 36 countries. The Analytics Base Table (ABT) is given below:

Column Name	Description	Type
Country Name (Common Variable)	Country names	String
Country Code	Country codes	String
Year (Common Variable)	Year of the data	Integer
GDP (current US$)	Gross Domestic Product	Float
Tax revenue (% of GDP)	Tax revenue as a percentage of GDP	Float
Total debt service	Debt service as a percentage of exports	Float

Column Name	Description	Type
o_food_price_index	Opening food price index.	Float
c_food_price_index	Closing food price index.	Float
inflation_food_price_index	Inflation-adjusted food price index.	Float
Date	Date of observation.	Datetime
Month	Month of observation (1-12).	Integer

Year (Common Variable)	Year of observation.	Integer
Country (Common Variable)	Name of the country	Categorical
ISO3	ISO-3 country code	Categorical
o_food_price_index_transformed	Log-transformed opening food price index.	Float
c_food_price_index_transformed	Log-transformed closing food price index.	Float
inflation_food_price_index_transformed (Target Variable)	**Log-transformed inflation-adjusted food price index.**	**Float**
inflation_lag_1	Inflation index lagged by 1 period.	Float
inflation_lag_3	Inflation index lagged by 3 period.	Float
inflation_lag_6	Inflation index lagged by 6 period.	Float
inflation_rolling_3	3-period rolling average of the inflation index.	Float
inflation_rolling_6	6-period rolling average of the inflation index.	Float
season	Season of the year	Categorical

2.1.1. Data Cleaning

Missing values in key columns, such as the inflation food price index, were identified and addressed by filling them with the median values for numerical columns. The median was chosen to avoid the influence of extreme outliers, which could significantly affect the mean (Soley-Bori, 2013). As for example: Missing values in the column `inflation_food_price_index` were imputed with the median value of the column to maintain consistency.

2.1.2. Outlier Detection and Removal

Outliers were detected using both Z-score and IQR methods and were handled through appropriate transformations or removal, ensuring the models were not skewed by extreme values (Borah & Baruah, 2013).

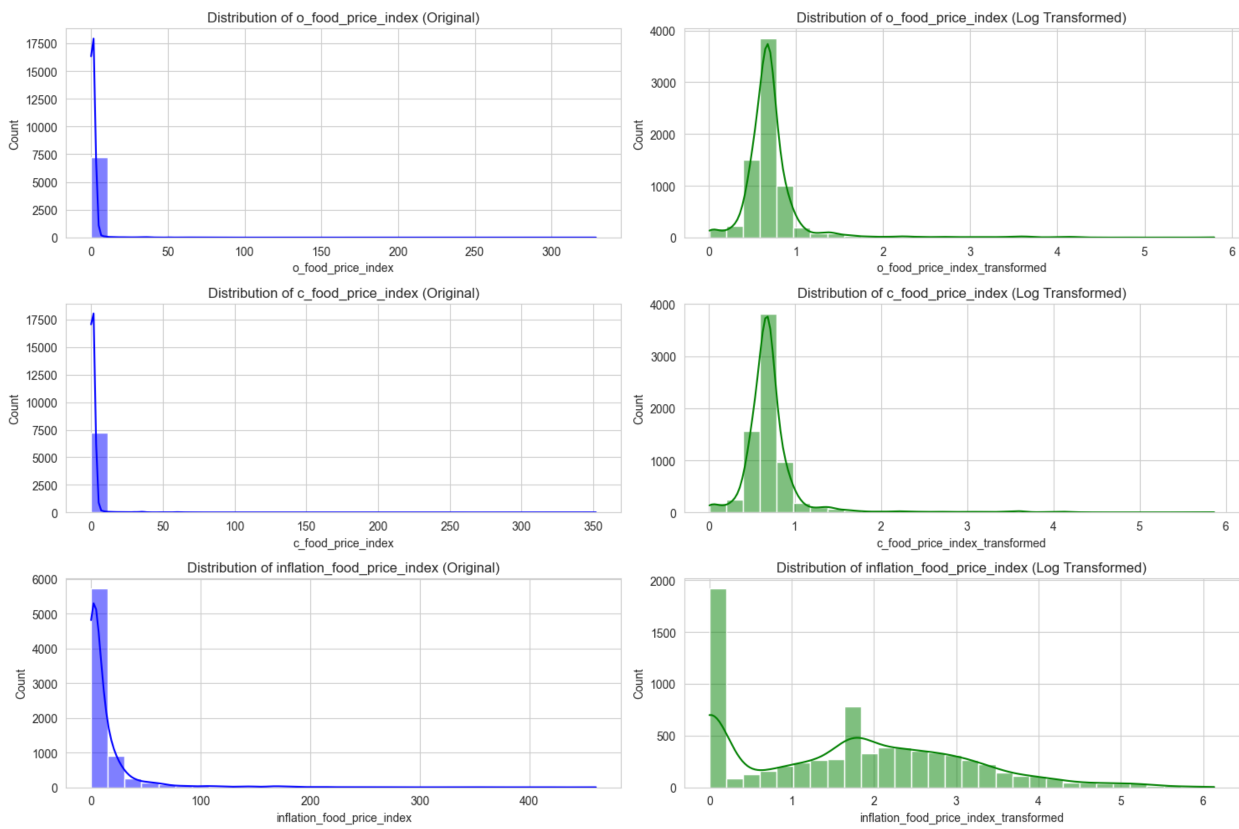
- **Z-Score Method:** Data points with Z-scores greater than 3 were flagged as outliers.
- **Interquartile Range (IQR):** Outliers were identified as data points outside 1.5 times the IQR from the first and third quartiles.

After identifying outliers in key columns like ‘o_food_price_index’, ‘c_food_price_index’, and ‘inflation_food_price_index’, transformations such as log scaling were applied to mitigate their impact while preserving the dataset's overall variability.

2.1.3. Data Transformation

The raw dataset underwent extensive transformations to ensure the data was suitable for advanced modeling and to enhance its predictive power.

- **Logarithmic Transformation for Normalization:** Key numerical columns, including ‘o_food_price_index’, ‘c_food_price_index’, and ‘inflation_food_price_index’, exhibited skewed distributions with long tails. To address this a log transformation was applied to these columns, which not only normalized the distributions but also retained interpretability for values near zero. The resulting columns were transformed into ‘o_food_price_index_transformed’, ‘c_food_price_index_transformed’ and ‘inflation_food_price_index_transformed’ displayed more symmetrical, bell-shaped distributions, improving the performance of statistical and machine learning models.

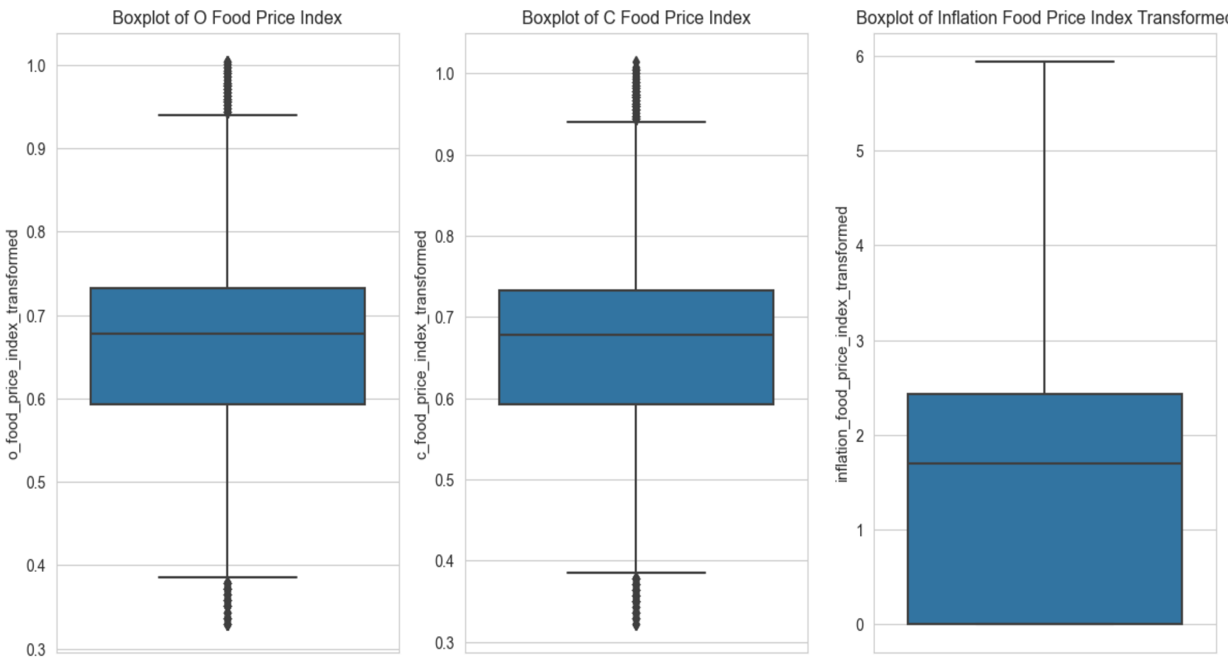


- **Feature Engineering:** To capture temporal and seasonal trends and enrich the dataset, several new features were introduced:
    - **Lag Features:** Lagged versions of the ‘inflation_food_price_index_transformed’ column were created to capture temporal dependencies and potential autocorrelation. Lags of 1, 3, and 6 months (‘inflation_lag_1’, ‘inflation_lag_3’, ‘inflation_lag_6’) were generated. Additionally, missing values for lagged features were handled by using the first valid non-null value for ‘inflation_lag_1’ and the mean of the column for ‘inflation_lag_3’ and ‘inflation_lag_6’.
    - **Rolling Averages:** Rolling averages for the transformed inflation index over 3-month and 6-month windows (‘inflation_rolling_3’, ‘inflation_rolling_6’) were computed to smooth out short-term fluctuations and highlight long-term trends. Furthermore, the missing values in these rolling averages were imputed with the mean of the respective columns.
    - **Seasonal Indicators:** A ‘season’ column was created by categorizing months into four seasons: Winter (December-February), Spring (March-May), Summer (June-August), and Fall (September-November). This feature provided a direct indicator of seasonal effects on food price inflation.
    - **Datetime Handling:** The ‘DATES’ column was converted into a proper datetime format to support time-series analysis and visualization. The dataset was enriched with derived features such as ‘year’ and ‘month’ extracted from the ‘DATES’ column.

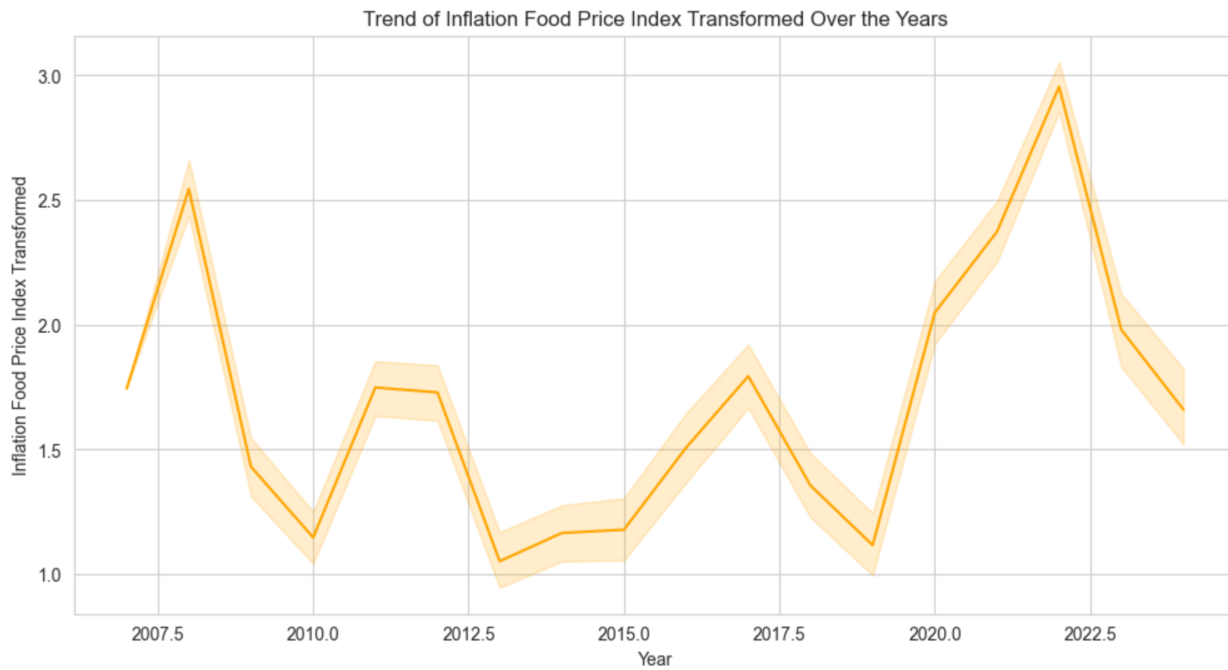
2.1.4. Exploratory Data Analysis (EDA)

A comprehensive EDA was performed to uncover patterns, relationships, and trends in the data.

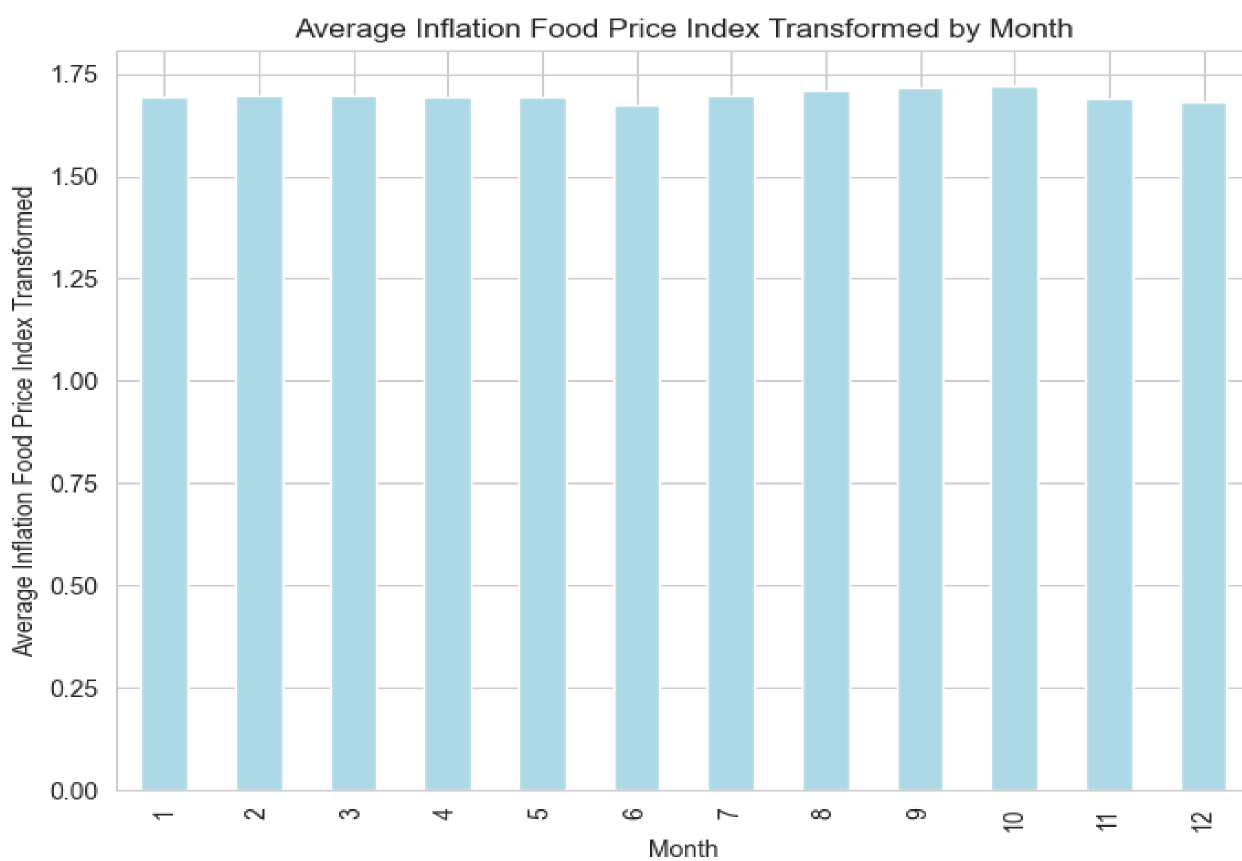
- **Distribution Analysis:** Histograms were plotted for the original and transformed columns ('o_food_price_index', 'c_food_price_index', and 'inflation_food_price_index'). Additionally, log-transformed distributions were more symmetrical, with reduced skewness and a better fit for statistical modeling.
  The original 'inflation_food_price_index' showed a bimodal distribution, indicating distinct low-inflation and high-inflation periods. The log transformation successfully addressed this, resulting in a unimodal distribution.
- **Outlier Analysis:** Boxplots revealed the presence of extreme outliers in the original data for all price indices and inflation rates. Subsequently, outliers were handled using the interquartile range (IQR) method, retaining only rows within 1.5 times the IQR of the first and third quartiles.
  Post-removal boxplots confirmed that the dataset was free from extreme outliers, reducing the risk of skewing model predictions.



- **Correlation Analysis:** A heatmap of the correlation matrix highlighted relationships among numerical variables. Moderate correlations were observed between the transformed 'o_food_price_index' and 'c_food_price_index' with the transformed inflation index, suggesting these variables might provide predictive value in a multivariate model. Temporal features, such as lagged and rolling averages, demonstrated meaningful correlations with the target variable, emphasizing the importance of incorporating time-dependent patterns.
- **Trend Analysis:** Line plots of 'inflation_food_price_index_transformed' over time revealed significant fluctuations in food price inflation, a sharp spike in 2008, corresponding to global economic disruptions. Subsequent periods of stability interspersed with high volatility, especially during global crises like the COVID-19 pandemic. Post-2020 trends indicated declining inflation, consistent with recovering supply chains and improved market conditions.



- **Monthly Patterns:** A bar chart of monthly average inflation indices demonstrated clear seasonal trends. Here, higher inflation rates during spring (March-May), possibly due to increased demand or reduced supply from agricultural cycles. And lower inflation rates during fall and winter months, reflecting seasonal harvesting and stable supply conditions.

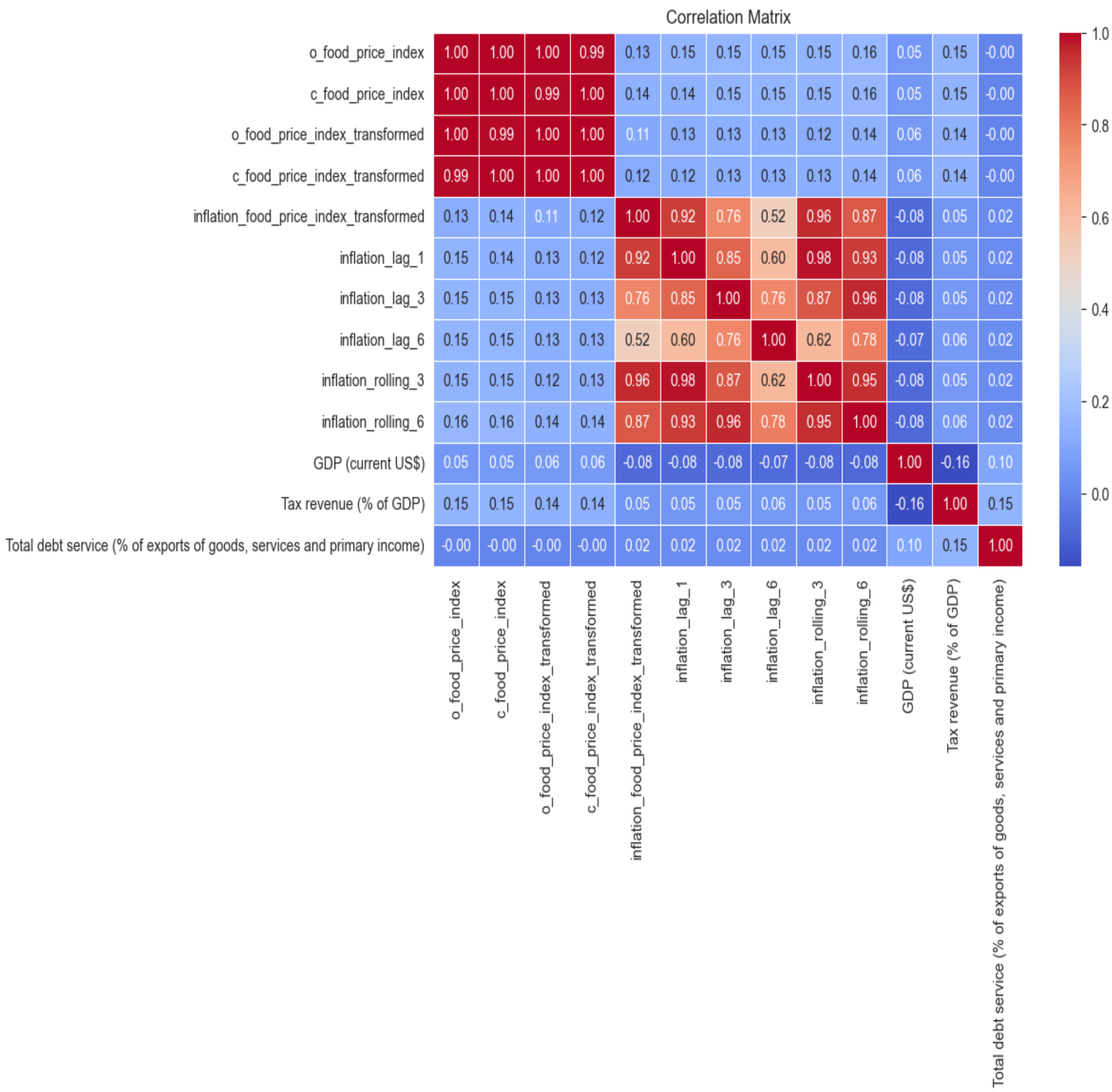


2.1.5. Key Insights from EDA

The rise of food prices showed clear seasonal and temporal trends, underscoring the importance of including seasonality and lagged values in prediction models. The data's temporal connections highlighted how crucial it is to comprehend historical patterns in order to accurately predict future inflation rates. Seasonality is an important consideration in modeling since seasonal changes, such as higher inflation in the spring and lower inflation in the fall and winter, represented fundamental agricultural cycles and market dynamics.

By lowering skewness and leveling the data, the use of logarithmic adjustment greatly enhanced the distribution of important variables. By guaranteeing that the input features matched the presumptions of the majority of prediction techniques, this improvement improved the dataset's suitability for machine learning algorithms. Furthermore, the dataset was enhanced by the addition of rolling average characteristics and seasonal indicators, which allowed it to identify intricate and nuanced patterns.

These transformations and feature engineering steps created a comprehensive and robust dataset, laying the groundwork for precise and insightful forecasting models (OpenAI, 2024).



2.2. Model Selection

2.2.1. Selection Approach

The development of predictive models in this study involved a rigorous selection process based on the characteristics of the dataset and the objectives of the analysis. Models were chosen to capture both the linear and non-linear relationships inherent in the data, as well as the temporal and seasonal patterns observed in food price inflation.

Initial experiments began with linear regression as a baseline, providing a straightforward and interpretable benchmark. Advanced ensemble methods, such as Random Forest Regressor and Gradient Boosting Regressor, were employed to capture complex, non-linear interactions between variables. Additionally, XGBoost was utilized for its computational efficiency and ability to handle large datasets. For time-series forecasting, ARIMA (Auto Regressive Integrated Moving Average) was selected to model temporal dependencies and seasonal patterns. Regularized regression techniques, including Ridge Regression and Lasso Regression, were incorporated to address potential multicollinearity and enforce feature selection.

2.2.2. Model Training and Hyperparameter Tuning

For machine learning models, the dataset was partitioned into training, validation and testing subsets. The training data was used to fit the models, while the testing data was reserved for evaluating generalization performance. The data splitting in the project was implemented as follows:

- **Temporal Splitting:** The dataset was divided into training, validation, and testing sets based on chronological order to preserve the time-series nature of the data. The training data includes observations from 2007 to 2019. Validation data includes observations from 2020 to 2021. Finally, testing data includes observations from 2022 to 2023
- **Rationale for Temporal Splitting:** This approach ensures that the models are trained on past data and evaluated on future observations, aligning with real-world forecasting scenarios. It prevents information leakage from future data into the training process, thereby maintaining the integrity of the predictive framework.
- **Feature Scaling:** Before model training, features in the training and testing datasets were scaled using StandardScaler to standardize the input variables. This step was particularly important for models like Ridge Regression and Gradient Boosting, which are sensitive to the scale of input features

Hyperparameter tuning was conducted using grid search combined with cross-validation to optimize model configurations. Parameters such as the number of estimators, learning rate, and maximum depth (for ensemble methods) and regularization strength (for Ridge and Lasso Regression) were systematically adjusted to balance bias and variance. For ARIMA, stationarity was ensured through differencing and log transformation, and seasonal components were incorporated to improve accuracy.

2.3. Model Evaluation

2.3.1. Evaluation Metrics

The effectiveness of the models was assessed using a combination of error-based metrics and goodness-of-fit measures. Mean Squared Error (MSE) and Mean Absolute Error (MAE) quantified the average prediction errors, with MSE penalizing larger deviations more heavily.

R-squared (R^2) measured the proportion of variance explained by the models, providing a sense of their explanatory power. These metrics were selected to evaluate not only the accuracy but also the reliability of the predictions.

2.3.2. Validation Strategy

To ensure robust evaluation, the dataset was split into training and testing subsets, with the latter providing a measure of model generalization. For machine learning models, k-fold cross-validation was employed, where the dataset was divided into k folds, and the model was trained and tested iteratively on different subsets. This technique minimized the influence of specific train-test splits on the evaluation results. For time-series models like ARIMA, a chronological split was used to maintain the temporal order, training on earlier data and validating on subsequent periods.

2.3.3. Comparative Performance

Model performance was compared across all candidates to identify the most suitable approach for forecasting food price inflation. Ensemble methods such as Random Forest and Gradient Boosting consistently outperformed simpler linear models, demonstrating their ability to capture complex relationships and interactions among features.

XGBoost provided further improvements in predictive accuracy due to its efficiency and scalability. Time-series models like ARIMA effectively captured temporal trends and seasonality but required significant parameter tuning to achieve optimal results. Regularized regression techniques (Ridge and Lasso) were valuable in highlighting key predictors while reducing overfitting.

3. Results

3.1. Model Performance Analysis

The modeling process culminated in the evaluation of multiple models, each assessed for its ability to predict food price inflation accurately and reliably.

Ridge Regression emerged as the top-performing model due to its consistent performance across both training and testing datasets, with a balanced approach to handling multicollinearity and regularization. Fine-tuned Random Forest showed significant improvements over its base version, achieving competitive results with Ridge Regression, particularly in terms of Mean Absolute Error (MAE).

Other models, such as XGBoost, ARIMA, and Lasso Regression, demonstrated varying levels of effectiveness but were ultimately outperformed by Ridge Regression and Random Forest.

Model	Training RMSE	Test RMSE	Training R^2	Test R^2	Training MAE	Test MAE
Ridge Regression	0.4993	0.4420	0.8308	0.8755	0.3451	0.3202
Lasso Regression	0.5130	0.4682	0.8214	0.8603	0.3576	0.3422
XGBoost	0.1756	0.6215	0.9791	0.7539	0.1187	0.4560
Random Forest	0.1701	0.4952	0.9804	0.8438	0.1043	0.3273
ARIMA	2.0745	2.0745	-1.7418	-1.7418	1.6536	1.6536
Gradient Boosting	0.6259	0.6259	0.7504	0.7504	0.4426	0.4426

3.2. Performance Metrics and Comparisons

3.2.1. Ridge Regression

Ridge Regression consistently demonstrated low error metrics and high R^2 scores on both training and testing datasets, making it the most reliable and interpretable model. The regularization effect of Ridge Regression ensured robust handling of multicollinearity, and fine-tuning of the regularization parameter (alpha) confirmed that the model was already optimized.

Its simplicity and efficiency make it a practical choice for real-world applications where computational cost and interpretability are critical.

3.2.2. Random Forest (Fine-Tuned)

Fine-tuning significantly improved the performance of Random Forest, particularly reducing MAE from 0.3273 to 0.2967. With a test RMSE of 0.4952 and an R^2 score of 0.8438, the model demonstrated excellent generalization capabilities, outperforming Ridge Regression in terms of test accuracy.

The inclusion of hyperparameter tuning (n_estimators, max_depth, and min_samples_split) was instrumental in balancing model complexity and avoiding overfitting. However, the model's computational complexity and reduced interpretability make it less practical for scenarios requiring simplicity and speed.

3.2.3. XGBoost

XGBoost showed strong performance on the training set with a very high R^2 (97.91%) and low RMSE (0.1756). However, it struggled to generalize, as evidenced by a sharp increase in test RMSE (0.6215). This overfitting suggests that XGBoost may not be well-suited for this dataset, likely due to its sensitivity to overparameterization and its reliance on a rich feature space.

3.2.4. Lasso Regression

Lasso Regression provided reasonable performance but was consistently outperformed by Ridge Regression. While its L1 regularization facilitated automatic feature selection, the resulting higher errors on the test set indicate that it was less robust in capturing the multivariate dependencies of the data.

3.2.5. ARIMA

ARIMA, as a univariate time-series model, struggled to handle the multivariate nature of the dataset and the complexity of the relationships between variables. Its negative R^2 score and high error metrics highlighted its limitations, especially for datasets that extend beyond traditional time-series patterns.

3.2.6. Gradient Boosting

Gradient Boosting performed moderately well, with a test RMSE of 0.6259 and an R^2 score of 0.7504. While it demonstrated better generalization than XGBoost and ARIMA, it still lagged behind Random Forest and Ridge Regression, making it a less optimal choice for this task.

Model	Test RMSE	Test R^2 Score	Test MAE	Observations
Ridge Regression	0.4420	0.8755	0.3202	Ridge remains one of the most consistent performers, showing excellent balance between simplicity and performance.
Fine-Tuned Ridge	0.4420	0.8755	0.3202	No significant improvement after fine-tuning, indicating the default alpha was optimal.
Lasso Regression	0.4682	0.8603	0.3422	Lasso performed slightly worse than Ridge, likely due to the stronger regularization.
Random Forest	0.4952	0.8438	0.3273	Original Random Forest showed strong performance but room for improvement.
Fine-Tuned Random Forest	0.4441	0.8744	0.2967	Fine-tuning improved Random Forest's performance significantly, particularly in MAE.
XGBoost	0.6215	0.7539	0.4560	XGBoost struggled with generalization despite strong training performance.
ARIMA	2.0745	-1.7418	1.6536	ARIMA performed poorly, indicating it is unsuitable for this problem.

3.3. Interpretation of Results

The findings of this study demonstrate that Ridge Regression is the most effective model for predicting food price inflation, offering consistent and reliable performance across both training and testing datasets.

This aligns with the objectives of the project, which aimed to develop a robust, interpretable, and computationally efficient model for forecasting. Ridge Regression's ability to handle multicollinearity while maintaining a low error rate makes it a valuable tool for understanding the complex relationships between food price indices, inflation, and seasonal trends. The inclusion of engineered features, such as lagged variables and rolling averages, significantly improved predictive accuracy, underscoring the importance of capturing temporal dependencies and seasonal patterns in modeling inflation dynamics.

Additionally, fine-tuned Random Forest demonstrated competitive performance, highlighting the potential of ensemble methods to address non-linearities in the data, albeit with higher computational demands. Models such as ARIMA and XGBoost, while less effective in this context, provided insights into the challenges of generalizing time-series and highly complex models to multivariate and dynamic datasets like food price inflation.

3.4. Implications

The results of this study have significant implications for industries and policymakers dealing with food price volatility. Reliable forecasts of food price inflation can inform agricultural planning, supply chain optimization, and pricing strategies in the food industry.

For policymakers, these models can provide valuable insights into inflationary trends, enabling better-targeted interventions to stabilize prices and mitigate the impact on consumers, particularly in regions vulnerable to food insecurity.

Seasonal insights gleaned from the analysis can guide production schedules and inventory planning, while the identified temporal dependencies highlight the need for continuous monitoring of past trends to predict future inflation accurately.

Furthermore, the robustness and interpretability of Ridge Regression make it an accessible tool for stakeholders without deep expertise in data science, bridging the gap between advanced analytics and practical decision-making.

On a societal level, accurate predictions of food price inflation can help mitigate the adverse effects of price surges, ensuring affordability and stability for consumers, particularly in low-income communities.

3.5. Limitations

Despite the promising results, we have found that this study has several limitations that should be acknowledged. First, the dataset's reliance on historical data assumes that past trends will persist, which may not hold true in the face of unprecedented events such as pandemics or geopolitical disruptions. This reliance may limit the generalizability of the findings to future conditions that deviate significantly from historical patterns.

Second, while Ridge Regression offers strong performance and interpretability, it may not fully capture non-linear dynamics that could be critical in highly volatile or complex market scenarios. The underperformance of ARIMA and XGBoost suggests that the dataset’s multivariate and dynamic nature challenges traditional time-series models and overly complex machine learning models, respectively.

Additionally, the study's scope was limited to specific modeling techniques and feature engineering approaches, other advanced techniques, such as deep learning or hybrid models, may yield improved results.

Finally, the study focused primarily on predictive accuracy, with less emphasis on real-time adaptability or integration into operational systems, which are critical for practical deployment. Addressing these limitations in future research could enhance the robustness and applicability of the findings.

4. Integration of Interactive Dashboard

An interactive dashboard was developed as a key output of this study, providing an intuitive platform for stakeholders to visualize, analyze, and interact with the results of the predictive models. Built using tools such as Plotly Express and Dash, the dashboard enables users to explore food price inflation trends, model predictions, and feature contributions in real time.

4.1. Time-Series Visualization

Dynamic plots display historical and forecasted food price inflation over time, allowing users to identify trends, seasonal patterns, and temporal dependencies directly. This complements the machine learning models by visually validating their outputs.

4.2. Model Comparisons

Comparative performance metrics (e.g., R^2, RMSE, MAE) for different models are presented through interactive bar charts and tables, helping users understand the trade-offs between accuracy, complexity, and interpretability.

4.3. Scenario Analysis

Users can simulate scenarios by adjusting input variables, such as timeline or country, to see their impact on future predictions. This feature is particularly useful for policymakers and industry stakeholders to test ‘what-if’ conditions and plan accordingly.

4.4. Key Takeaways from Dashboard Integration

The interactive dashboard demonstrated distinct seasonal patterns in the inflation of food prices, showing steady rises in certain months, such as spring, and decreases in fall and winter. Dynamic time-series visualizations were used to highlight these trends, confirming the importance of seasonal indicators as key components of the prediction models. Stakeholders, especially those involved in agricultural and supply chain management, can efficiently plan production schedules, pricing strategies, and interventions by identifying the months with the largest inflation risks.

By displaying parameters like RMSE, MAE, and R2, the dashboard also made it easier to compare the performance of different models. This feature helped users choose the best model for their needs and operational priorities by allowing them to weigh the trade-offs between predictability, interpretability, and simplicity.

Furthermore, heatmaps and bar charts were used to show feature importance information, making it clear which variables had the biggest impact on predictions. These insights are especially helpful for analysts and policymakers because they give them evidence-based tools to address the main causes of food price rise and put specific solutions in place.

5. Conclusion

5.1. Summary of Key Findings

This study successfully developed and evaluated predictive models for forecasting food price inflation, identifying Ridge Regression as the most reliable and interpretable approach. Ridge Regression demonstrated consistent performance across training and testing datasets, with low error rates and robustness against multicollinearity.

Fine-tuned Random Forest emerged as a strong alternative, achieving competitive accuracy, particularly in terms of MAE and R^2, though with higher computational demands. Temporal and seasonal dependencies, captured through engineered features such as lagged variables and rolling averages, significantly enhanced predictive performance across all models.

Other methods, including XGBoost and ARIMA, were less effective, highlighting the challenges of generalization for highly complex or traditional time-series models in multivariate settings. These results underscore the importance of balancing model accuracy, interpretability, and computational efficiency, providing valuable tools for understanding and managing food price volatility.

5.2. Achievement of Objectives

The project's main goals of creating and assessing prediction models to accurately and consistently predict food price inflation were accomplished. Ridge Regression and Random Forest were used to show that, with the right feature architecture, even basic and sophisticated models can handle the intricacy of food price movements. In order to achieve the goal of using historical and seasonal trends for forecasting, the models were made to capture the temporal and cyclical patterns present in the data through the use of lagged features, rolling averages, and seasonal indicators.

Furthermore, this whole work was greatly aided by the integration of machine learning models with visualization, which improved the data interpretability and offered useful insights. Although machine learning models like Random Forest and Ridge Regression produced precise forecasts of food price inflation, visualizations were crucial for identifying trends, confirming model results, and effectively informing stakeholders of discoveries.

On top of that, the results were made available to a wide audience by incorporating visualization into the output step. Using programs like Plotly or Dash, interactive dashboards were made to display seasonal fluctuations, temporal patterns, and prediction results in an easy-to-understand manner. Through these dashboards, stakeholders were able to comprehend the projections, examine the data in real-time, and evaluate the ramifications. In addition to improving the models' interpretability, this combination of machine learning and visualization enabled data-driven decision-making, bridging the gap between the technical and practical components of food price predictions.

However, there is room for critical reflection, even though the stated goals were mostly achieved. ARIMA and XGBoost's poor performance demonstrates the shortcomings of some methods when used on dynamic and multivariate datasets, indicating the need for more research into hybrid or ensemble approaches. Predictive accuracy was given top priority in the project, but real-time deployment and adaptation, both crucial for practical implementation, were not given as much attention. Notwithstanding these difficulties, the study offered useful tools for businesses and governments to predict and lessen food price volatility, including actionable insights and strong models that successfully address the problem statement.

6. References

1. *Acara Climate*. (2024). From Leveraging AI for Sustainable Agriculture: A Deep Dive into Predictive Analytics: https://acara-climate.com/leveraging-ai-for-sustainable-agriculture-a-deep-dive-into-predictive-analytics/
2. Arezki, R., El Aynaoui, K., Nyarko, Y., & Teal, F. (2016). Food price volatility and its consequences: Introduction. *Oxford Economic Papers*, 655-664.
3. Asadollah, O., Carmy, L. S., Hoque, M. R., & Yilmazkuday, H. (2024). Geopolitical risk, supply chains, and global inflation. *Journal of International Economics*.
4. Borah, M., & Baruah, H. K. (2013). A survey of Outlier Detection Methods in Network Anomaly Identification. *The International Journal of Computer Applications*.
5. Furceri, D., Loungani, P., Simon, J., & Wachter, S. M. (2023). Global food prices and domestic inflation: Some cross-country evidence. *Food Policy*.
6. Kalkuhl, M., von Braun, J., & Torero, M. (2016). Food Price Volatility and Its Implications for Food Security and Policy. *Springer International Publishing*.
7. OpenAI. (2024). *ChatGPT (Mar 14 version) [Large language model]*. From https://chat.openai.com/chat
8. Soley-Bori, M. (2013). *Dealing with missing data: Key assumptions and methods for applied analysis.* Boston University.