AI201 Report 1
Food Pricing Analysis of Food Commodities in the Regions of the Philippines Using Machine Learning Methods

Miguel Luis Martinez, Juan Carlos Roldan, Jeryl Salas

June 22, 2024
Master of Engineering in Artificial Intelligence
College of Engineering
University of the Philippines Diliman

1. INTRODUCTION

Food prices and their overall increase remain a major issue around the world. In the Philippines, the prices of common food items such as rice, meat, and vegetables have strained the budgets of Filipinos even before the major economic shocks of the 2020s that are often blamed for recent worsening prices, such as the COVID-19 pandemic, food cartels, supply chain disruptions from overseas wars, mass importation of goods, inflation, and the weakening of the peso with respect to the United States Dollar (USD).

There are several factors that contribute to the fluctuations in food prices. One is the direct relationship between supply and demand and food pricing—prices rise when supply fails to meet demand. ING, a Dutch multinational banking and financial services corporation, reported a 20% annual increase in rice prices in the Philippines due to tight supply constraints. This problem is exacerbated by added economic factors such as inflation, causing food inflation to reach 4.6% annually . Similarly, local news outlet The Philippine Star asserted that economic practices such as haggling and unregulated middleman networks lead to increased prices at the local round scad or galunggong markets. Seasonal variations in supply also have a significant impact on price volatility, with local red onions reaching an all-time high of 300–450 Philippine Pesos (PHP) per kilogram in 2023, which was further exacerbated by ineffective farming methods and climate vulnerability [13].

Geographic and meteorological factors further influence pricing dynamics due to the logistical constraints and weather patterns of each location or region. According to GMA Integrated News, the local GMA network's news division, transporting commodities like tomatoes from far-flung locations in Nueva Ecija to markets in the NCR, for example, is more expensive due to transportation costs, which are determined by trucking services, distance, and fuel prices. Furthermore, the Philippine seasons El Nino, which intensifies drought, and La Nina, which causes heavier rainfall, have a direct impact on overall agricultural productivity (El Niño waning, La Niña to develop in second half of 2024).[10]

The study leverages machine learning algorithms to analyze monthly food price data for different regions in the Philippines, identifying patterns and trends that provide significant results and insights, which allows for more informed food pricing decisions and policies. In addition, the study aims to discover which machine learning models are best suited for forecasting food prices as well as to uncover price anomalies across different regions.

1.1 Objectives

The main objectives of the study are to identify the following:

- 1a Are there price movements that occur simultaneously or inversely across different types of food (correlation)?
- 2a Which food categories had a more significant increase in price over the last 24 years?
- 2b Which food items in each category were the most volatile in terms of price?
- 3a Which regions experienced the most significant price increases in each category?
- 3b Are there recurring patterns in food prices that occur at specific times of the year (seasonal)?
- 4a How are different food categories related to each other within the region or across regions?
- 4b Are there anomalies in the dataset that indicate significant economic events?
- 4c Is the trend for the years 2020 to 2024 consistent with the trends from 2008 to 2019?

2. RELATED LITERATURE

Analyzing and forecasting food prices pose significant challenges due to the multifaceted nature of economic influences (Smith, 2020). Supply and demand, inflation, economic practices, seasonal variations, and geographic and meteorological factors all pose a considerable amount of uncertainty in price projections, which requires the use of advanced analytical techniques. Machine learning methods, with their ability to handle large and complex datasets and uncover non-linear relationships, offer promising avenues for improving the accuracy of food price forecasts in this volatile setting.

Traditionally, econometric models such as Vector Autoregressive (VAR) and Vector Error Correction Models (VECM) have been employed for this purpose [11, 8, 2]. However, with the advent of advanced computational techniques, machine learning models, including Gaussian Process Regression (GPR), have gained prominence for their ability to handle non-linear patterns in data [4, 5].

Econometric models have long been used in real estate price forecasting. McGough and Tsolacos [8] applied time-series models, including moving average and autoregressive models, for short-term forecasting of commercial rental prices. Brooks and Tsolacos [8] used VAR models to forecast retail rental prices, finding varying forecast accuracy depending on the horizon. These models provide a foundation but often struggle with non-linearity and complex interactions in the data.

Machine learning models offer robust alternatives to traditional methods by capturing complex, non-linear relationships. Neural networks, support vector regression (SVR), and random forests have been extensively studied for real estate price prediction. For instance, Khalafallah [6] demonstrated the effectiveness of neural networks in predicting housing market performance. Similarly, Park and Bae [11] used machine learning algorithms for housing price prediction in Fairfax County, Virginia, showcasing their superior performance over traditional models.

GPR has emerged as a potent tool for real estate price forecasting due to its flexibility and probabilistic nature. Unlike other machine learning models, GPR provides not only predictions but also uncertainty measures, which are valuable for risk assessment. In Dambon et al. [5], they applied GPR to large datasets, showing that it can handle significant amounts of data while providing accurate forecasts. Their model outperformed traditional econometric models, particularly in capturing non-linear trends. In a paper by Xu and Zhang [14], there was an application of GPR to forecast retail property prices in ten major Chinese cities. They employed Bayesian optimization and cross-validation to select optimal kernels and basis functions, achieving high forecast accuracy.

Studies comparing GPR with other machine learning models, such as neural networks and SVR, have generally found GPR to be competitive. For example, in a study by Chen et al, [3], GPR was found to outperform SVR and neural networks in forecasting housing prices due to its ability to model uncertainty. Similarly, Morano et al. [9] compared GPR with traditional hedonic price models and found that GPR provided more accurate and reliable predictions.

The researchers discovered similar concerns with food pricing, given the volatility of real estate values. Both markets are influenced by a number of unpredictable factors, such as the status of the economy and supply and demand dynamics; hence, this similarity encourages the use of GPR in food pricing as well. Food prices are a complex and diversified market, and GPR is a potential method for forecasting prices due to its ability to handle ambiguity and produce reliable predictions.

3. METHODOLOGY

All model training and evaluation was performed on a Google Colab server with a Nvidia T4 GPU.

3.1 Data acquisition and exploration

For the dataset, the researchers made use of the "Philippines - Food Prices" dataset taken from the World Food Program Price Database. This dataset contains monthly price data for different types of food, such as rice, meat, dairy, and vegetables, and is split by region and city, with prices in PHP and USD. The dataset was cleaned of non-numerical values, and the datatypes of each field were formatted so that the same pipelines could be applied to both datasets without further modification and renaming of fields.

Prior to the model training phase, the researchers obtained two sets of correlation matrices: one containing correlation between prices of categories of food per region, and another containing correlation between region prices per category of food. These will be used in identifying which food prices strongly correlate or strongly inversely correlate with one another.

3.2 Model Training and Evaluation

Two identically structured models were trained from the dataset, one for the PHP prices and one for the USD prices. All prices are taken directly from the dataset and are not adjusted for inflation.

3.2.1 Determining Price Volatility and Highest Price Categories

To identify the most volatile and most highly priced food categories in terms of historical pricing per region, K-Nearest Neighbors (KNN) clustering was used, followed by GPR to Process Regression, in order to build a model that can both identify which types of food products in the Philippines are most affected by price changes and inflation. Principal Component Analysis (PCA) is then used to map the clusters to a 2D space for visualization.

On the other hand, the performance of the KNN model was based on Mean Absolute Error (MAE) and R-squared (R2) values.

3.2.2 Anomaly Detection

Using the two components from the PCA, anomaly detection methods were used to find outliers in the dataset involving date and time, as well as the regions with their longitude and latitude. Three methods were used for this analysis, namely Kernel Density Estimation (KDE), Isolation Forest, and Local Outlier Factor (LOF). All of the methods were done through the Scikit-learn library.

KDE is a non-parametric method for estimating the probability density function of a random variable. In the case of anomaly detection, KDE is used to identify the outliers of the data based on the estimated density. KDE uses kernel functions to smooth the contributions of each data point. In this case, the Gaussian kernel function was used. As stated from Scikit-learn density estimation documentation, a kernel is a positive function $K(x; h)$ which is controlled by the bandwidth parameter h. Given this function, the density estimate at a point y within a group of points $x_i; i = 1...N$ is given by:

$$\rho_K(y) = \sum_{i=1}^{N} K\left(y - x_i; h\right)$$

Where:

- $\hat{f}(y)$: Estimated density function at point y.
- N: Number of data points.
- h: Bandwidth (smoothing parameter).
- $K(u)$: Kernel function.
- x_i: Data points ($i = 1, \ldots, N$).

The bandwidth acts as a smoothing parameter that controls the trade-off between bias and variance. GridSearchCV was performed to search for the optimal bandwidth. At the same time, Isolation Forest is a tree-based anomaly detection method that isolates observations by randomly selecting a feature and then randomly selecting a split value between the maximum and minimum values of the selected feature. As stated by Liu, F. T., Ting, K. M., & Zhou, Z.-H. [7], an Isolation Forest builds an ensemble of iTrees for a given data set, and anomalies are those instances that have short average path lengths on the iTrees.

The anomaly score for a point x is given by the average path length required to isolate x. The formal definition is:

$$s(x, n) = 2^{-\frac{E(h(x))}{c(n)}}$$

where:

- $s(x, n)$: Anomaly score for the point x in a dataset with n points.
- $E(h(x))$: Average path length to isolate x.
- $c(n)$: Average path length of unsuccessful searches in a Binary Search Tree, used to normalize the path length.

The LOF is a density-based anomaly detection method that computes a score called the local outlier factor, reflecting the degree of abnormality of the observations. As stated by Breunig et al.[1], it measures the local density deviation of a given data point with respect to its neighbors. The LOF score of an observation is equal to the ratio of the average local density of its k-nearest neighbors to its own local density.

The anomaly score for a point x is given by the LOF, which is defined as:

$$LOF(x) = \frac{1}{k} \sum_{i=1}^{k} \frac{lrd_k(x_i)}{lrd_k(x)}$$

where:

- $LOF(x)$: Local Outlier Factor for the point x.
- k: Number of neighbors considered.
- $lrd_k(x)$: Local reachability density of x with respect to its k neighbors.
- x_i: i-th neighbor of x.

After performing anomaly detection using the three methods stated, the researchers then performed a comparative analysis by visualizing their detected anomalies and computing their silhouette scores to evaluate the quality of the clustering. The agreed-upon anomalies that were detected by these three methods were then compiled.

3.2.3 Price Trend Analysis

For each food price dataset and currency used, the data is interpreted as a time series sorted by month, with price data for all categories and regions per month being treated as data points for the prices in that month. The dataset is then split into two parts: data until December 2019, and data from January 2020 to May 2024. The former dataset is used to train a GPR model with an 8020 train-test split, which is then used to predict the price distribution for the latter dataset. The model was trained using a sampled dataset of 1,000 samples due to the researchers' server being unable to train the model on the entire dataset. A significant deviation of the prediction from the real data from January 2020 to May 2024 will mark this period as anomalous with respect to previous years for the currency used.

In addition to GPR, two other models were used on the same sampled dataset: an SVR and a Holt-Winters time series forecasting model. The hyperparameters for all three methods were trained using Optuna with Tree-structured Parzen Eliminator (TPE) Sampler and Hyperband pruning to perform hyperparameter tuning.

GPR is a non-parametric, Bayesian approach to regression that is particularly useful for modeling complex, non-linear relationships. According to Rasmussen, C. E., and Williams, C. K. I. [12], GPR provides a probabilistic prediction, indicating that it not only predicts the mean of the target function but also quantifies the uncertainty of the prediction. It starts by initializing a prior distribution that would describe the relationship between x and $f(x)$. The prior distribution is then updated to form a posterior distribution by calculating the posterior mean and covariance functions. These are then used to make predictions based on the input data.

The prediction for a new input x_* is given by:

$$f_* \mid X, y, x_* \sim \mathcal{N}(\bar{f}_*, \mathrm{Var}(f_*))$$

where:

- $\bar{f}_* = k(x_*, X)[K(X,X) + \sigma_n^2 I]^{-1} y$: Mean prediction.

- $\text{Var}(f_*) = k(x_*, x_*) - k(x_*, X)[K(X,X) + \sigma_n^2 I]^{-1} k(X, x_*)$: Variance prediction.

- $k(x_*, X)$: Covariance vector between the test point x_* and the training points X.

- $K(X, X)$: Covariance matrix of the training points.

- σ_n^2: Noise variance.

- I: Identity matrix.

- y: Observed target values.

GPR was chosen since it is a non-parametric model and can adapt to a wide range of data distributions and capture complex patterns. A related study from Zhengqi et al.[] exploring uncertainty quantification using GPR in steel-bearing capacities provides insights into how this can be applied for food price prediction. In this field, the jagged quality of steel processing and stochastic corrosive environment make measurements a little unpredictable. Likewise, in economic studies, with non-linear relationships and the need to understand associated uncertainties, GPR can be used to provide projection intervals, which can help in assessing the risk and reliability of the model's predictions.

The performance on the GPR model will be based on the Root Mean Squared Error (RMSE) of the extrapolated 2020 to 2024 data from the 2008 to 2019 model with respect to the actual data from each dataset.

4. EXPERIMENTAL RESULTS AND DISCUSSION

This will be split into sections based on the questions that were stated in the objectives.

4.1 Correlation of Price Movements Between Regions for Each Food Category

Figure 1 shows the correlation matrix between food categories for Region VII, which has the most positive correlations between categories. This indicates that food categories are more likely to move simultaneously in this region.

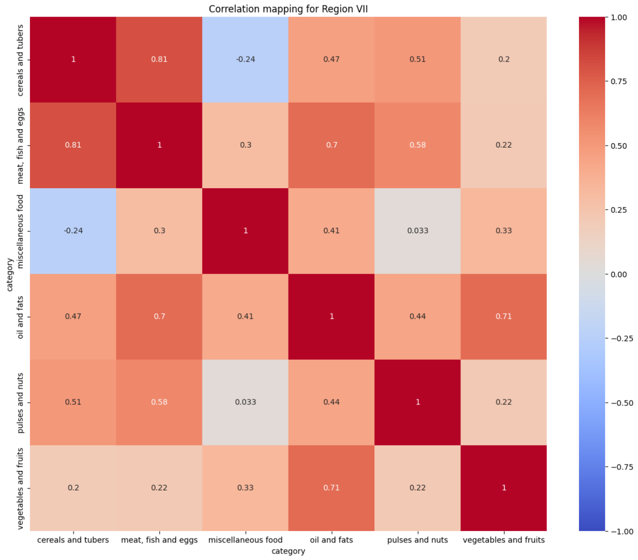

Figure 1: Matrix Between Food Categories for Region VII

Figure 2 shows the correlation matrix between food categories for Region VI. Interestingly, Region VI has the highest number of strong correlations between food items, both positive and negative correlations. The miscellaneous food category also has a high negative correlation with meat, fish, and eggs, as well as oil and fats. Additionally, there is a high positive correlation between pulses and nuts and oil and fats.

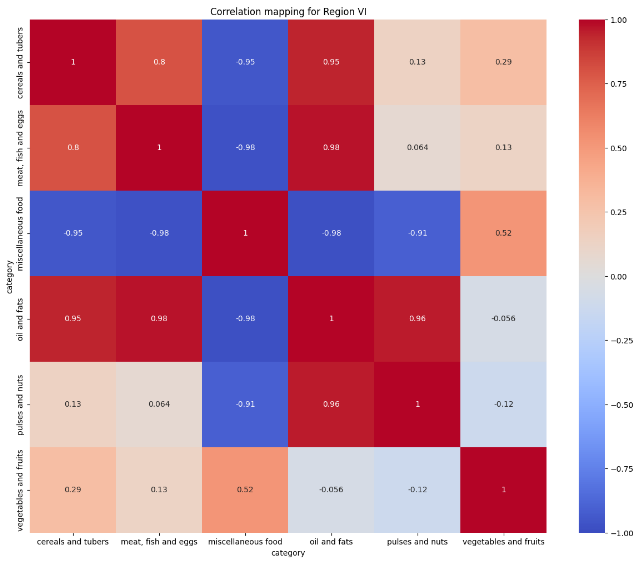

Figure 2: ***Correlation Matrix Between Food Categories for Region VI***

Figure 3 shows the correlation matrix between food categories for the Autonomous Region of Muslim Mindanao (ARMM). The region has the weakest correlations among food categories. The only strong correlation is between oil and fats and cereals and tubers, and the rest of the correlations have values below 0.5.

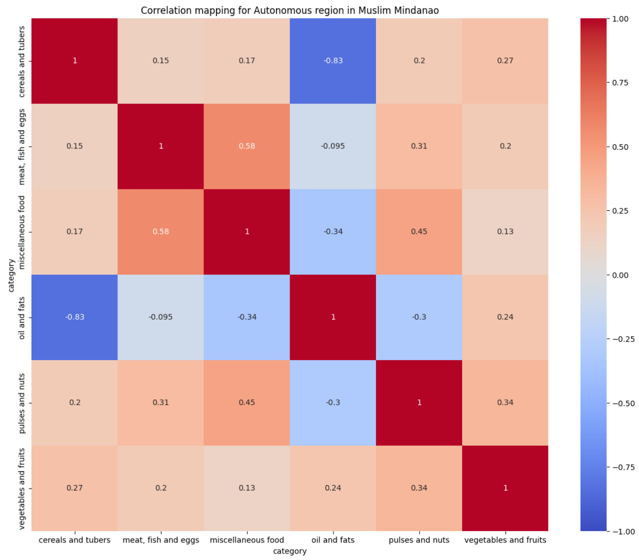

Figure 3: ***Correlation Matrix Between Food Categories for ARMM Region***

4.2 Price increase of food categories

Figure 4 shows a bar chart that display regions with the highest price increase per food category. As seen from the figure, the category meat, fish, and eggs had the most significant price increase with region 1 having the highest increase of 8934.65% followed by cereals and tubers where the ARMM region had the highest increase of 1031.11%.

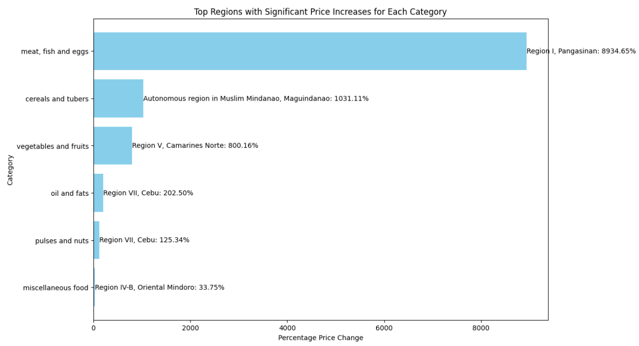

Figure 4: ***Food Categories with Significant Price Increase***

4.3 Volatile food items

Figure 5 shows the most volatile food items per food category. Volatility is calculated as the standard deviation of the monthly percentage changes. Mandarins have the highest volatility with 0.7992, followed by anchovies with 0.7332, and oil with 0.6715. The food category with the lowest volatility is miscellaneous food, with the food item sugar (brown) having the most volatility in its category with 0.13.

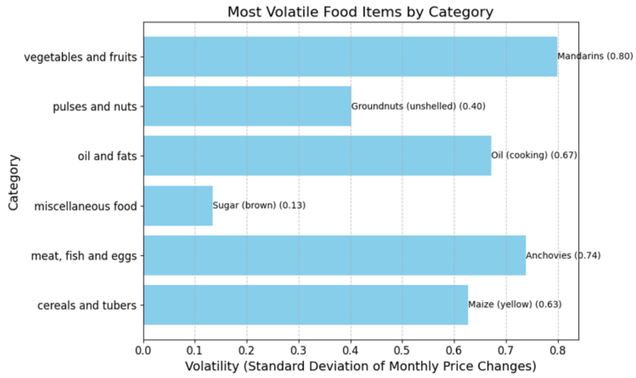

Figure 5: ***Most Volatile Food Categories***

4.4 Regions with significant price increase

Figure 6 shows the top five regions with significant price increases in terms of meat, fish, and eggs. Region I has the most significant price increase, followed by the Cordillera Administrative Region (CAR).

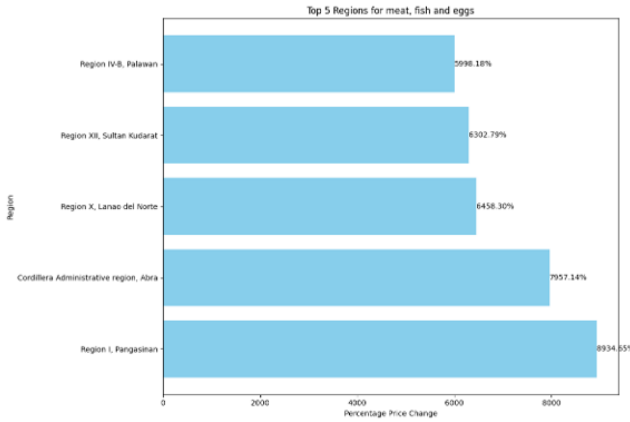

Figure 6: ***Top 5 Regions with Significant Price Increase for Meat, Fish, and Eggs***

Figure 7 shows the top five regions with a significant price increase in terms of the cereals and tubers category. The ARMM has the most significant price increase, followed by Region VII.

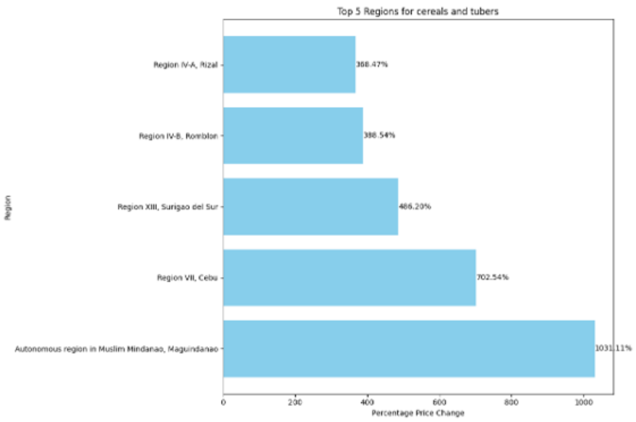

Figure 7: ***Top 5 Regions with Significant Price Increase for Meat, Fish, And Eggs***

Figure 8 shows the top five regions with a significant price increase in terms of vegetables and fruits. Region V has the most significant price increase, followed by Region IV-A.

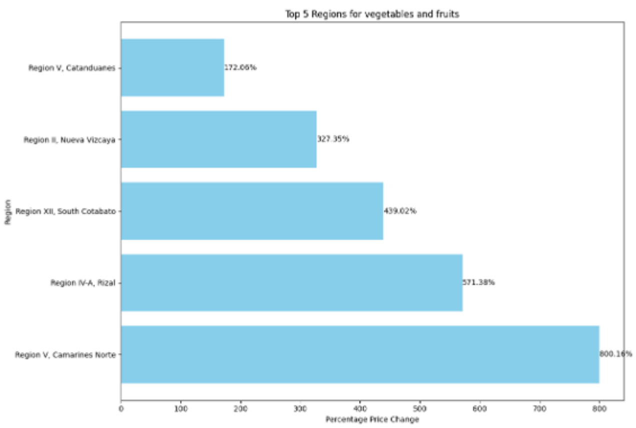

Figure 8: ***Top 5 Regions with Significant Price Increase for Vegetables and Fruits***

Figure 9 shows the top five regions with a significant price increase in terms of oils and fats. Region VII once again has the most significant price increase, followed by CAR.

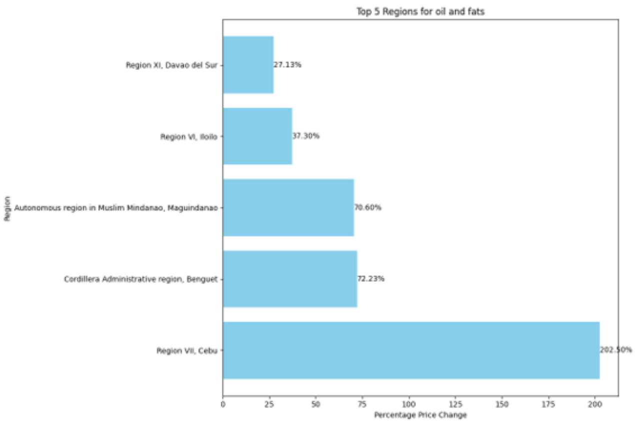

Figure 9: ***Top 5 Regions with Significant Price Increase for Oils and Fats***

Figure 10 shows the top five regions with a significant price increase in terms of pulses and nuts. Region VII, for the third time, has the most significant price increase, followed by Region VIII.

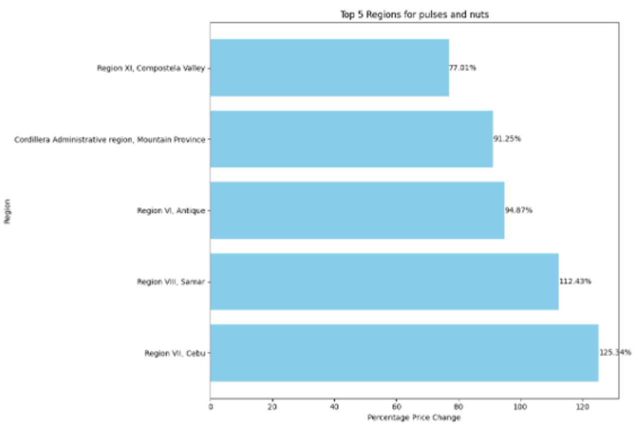

Figure 10: ***Top 5 Regions with Significant Price Increase for Pulses and Nuts***

Figure 11 shows the top five regions with a significant price increase in terms of miscellaneous foods. Region IV-B has the most significant price increase, followed by Region III.

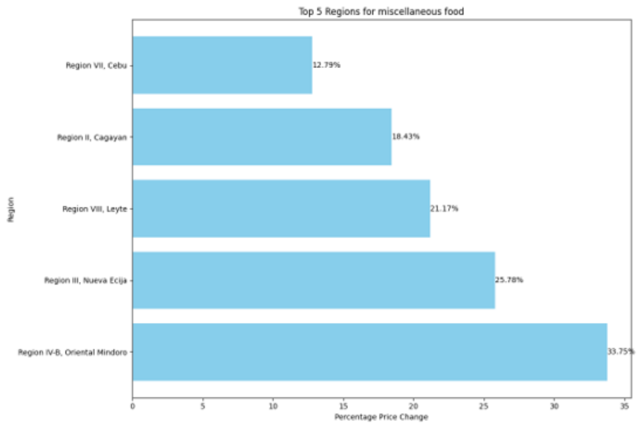

Figure 11: ***Top 5 Regions with Significant Price Increase for Miscellaneous Foods***

4.5 Seasonal patterns in food prices

The seasonal patterns in Figure 12 highlight a steep increase in prices for garlic, meat (pork), and onions during the months of May to July in 2016. Other food categories either remain stable or exhibit minor fluctuations in their prices throughout the year. Garlic shows a significant price increase starting from around May 2016, peaking around June to July, and then slightly dropping afterward. Meat (pork) had a moderate increase during the same period. Onions, specifically white onions, show a noticeable price increase from May to July, similar to garlic.

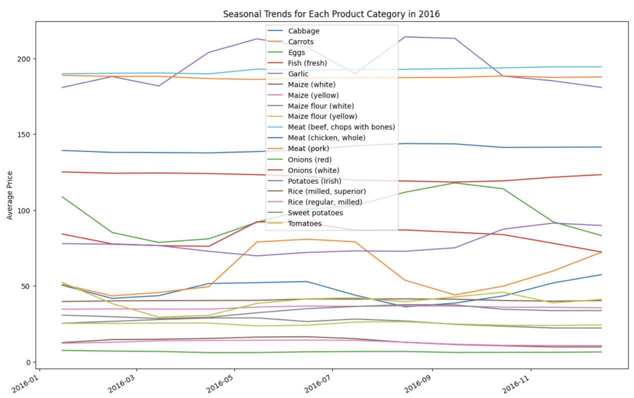

Figure 12: Steep Increase in Price of Garlic, Meat (Pork), and Onions During the Months of May to July of 2016

The scatter plot in Figure 13 then suggests that while the GPR model does capture some of the trends in the data, there are significant deviations, especially at higher price levels. There is a dense cluster of points at the lower end of the price spectrum, which are closer to 0, indicating that the majority of the data points have lower actual and predicted prices. There are outliers where actual prices are higher, but the model predicts lower prices, and vice versa. This indicates that the model may have limitations in accurately predicting higher price values. Further model tuning or alternative models may be necessary to improve the prediction accuracy, especially for higher price ranges.

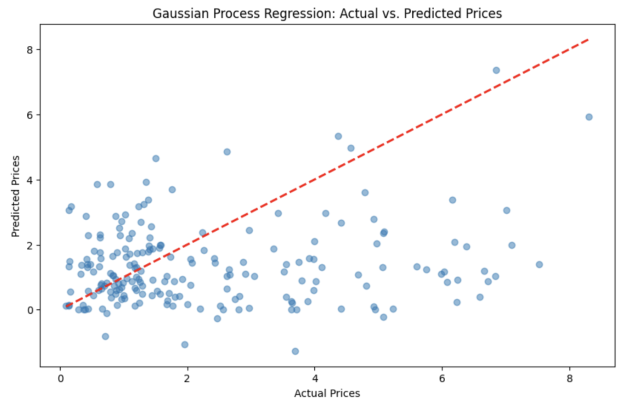

Figure 13: Gaussian Process Regression Results

4.6 Relation of food categories across regions

In Figure 14, the strongest correlation across regions is seen in the categories of cereals and tubers, as well as meat, fish, and eggs. Vegetables and fruits also project some of the weaker relationships across the country. This might be due to the fact that the top two categories are staple foods and have consistent demand across regions. This would lead to rather uniform pricing.

On the other hand, vegetables and fruits, being a little bit more susceptible to seasonal patterns, vary significantly from region to region. These goods are also a bit more perishable than the other categories and can induce more pronounced price fluctuations.

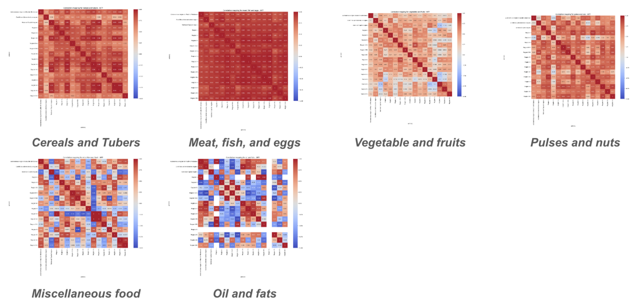

Figure 14: Correlation Matrices Across All 6 Food Categories

4.7 Anomalies in the dataset

In Figure 15, the use of three methods to perform anomaly analysis on the dataset is presented. KDE is best for datasets where a smooth density function can clearly separate normal and anomalous data, but it often struggles with high-dimensional and sparse data. KDE also shows very strict and straightforward boundaries around the dataset, whereas the Isolation Forest and LOF appear to be more flexible.

The majority of the data points are classified as normal observations by all three methods. Figure 15 shows that using multiple methods can enhance the robustness of anomaly detection by capturing different types of anomalies. The agreed anomalies in red provide high-confidence indicators of anomalies, which are detected by all three methods.

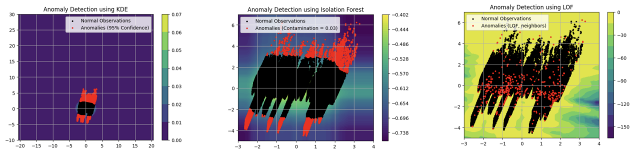

Figure 15: Anomaly Analysis with KDE, Isolation Forests, and LOF

4.8 Forecasting

Finally, in Figure 16, forecasting using the aforementioned methods is presented. The GPR model provides a smoother and more gradual forecast but seems to greatly miss the significant volatility and sudden changes seen in the test data. The SVR model predicts a decreasing trend, which performs a little better but does not match the actual data trends and suggests it might not be appropriate for this task. Exponential smoothing fails to capture any meaningful trend, indicating it is too simplistic for the complex nature of the data.

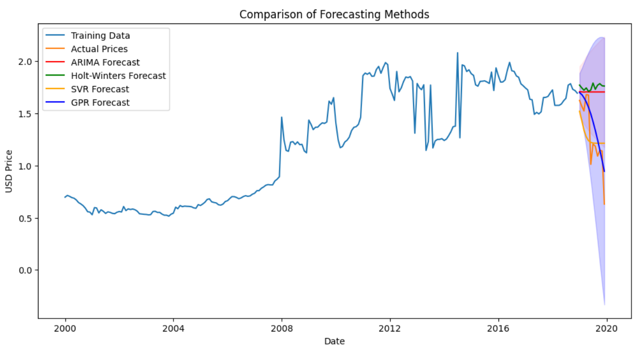

Figure 16: Forecasting with ARIMA, Exponential Smoothing, SVR, and GPR

5. CONCLUSION

The study successfully analyzed the pricing trends of food commodities across different regions of the Philippines using machine learning methods. By employing GPR, SVR, and Exponential Smoothing, the researchers were able to identify significant patterns and anomalies in the data, as well as forecast future trends.

The analyses revealed several key insights:

1. **Correlation of Price Movements**: The strongest correlations were found in the "Cereals and Tubers" and "Meat, Fish, and Eggs" categories, likely due to their consistent demand across regions. In contrast, "Vegetables and Fruits" showed weaker correlations, which can be attributed to their seasonal variability and perishability.

2. **Volatility and Significant Price Increases**: Certain food items, such as mandarins and anchovies, exhibited high volatility. Regions such as Region I and CAR experienced the most significant price increases in specific food categories.

3. **Anomaly Detection**: Through the use of KDE, Isolation Forest, and LOF, the researchers identified various anomalies in the dataset. Combining these methods enhanced the robustness of anomaly detection, highlighting the importance of using multiple approaches to capture different types of anomalies.

4. **Forecasting**: The GPR model provided smooth and gradual forecasts but failed to capture the significant volatility and sudden changes in the test data. The SVR model showed a decreasing trend, which did not align well with actual data trends, indicating that more sophisticated models might be needed to improve forecasting accuracy.

Overall, while machine learning models like GPR and SVR offer valuable insights into food price trends, their performance can be enhanced by incorporating more complex models and additional external factors. Future work should focus on improving model accuracy, especially for high-volatility food items and regions with significant price fluctuations. By leveraging advanced computational techniques, this study provides a foundation for developing more effective policies and interventions to stabilize food prices and ensure food security in the Philippines.

6. RECOMMENDATIONS

During the training of GPRs used in anomaly detection and time-series prediction, the computational complexity of the method prevented the use of the full dataset, as either dimensionality reduction or sampling was needed to successfully train the model without running out of memory. The performance of GPR may improve significantly if trained on the full dataset using better hardware. The use of deep learning models like Long Short-Term Memory (LSTM) Networks can be used to predict time series data with long-term dependencies, while dimensionality reduction methods like t-distributed Stochastic Neighbor Embedding (t-SNE) or Uniform Manifold Approximation and Projection (UMAP) can improve anomaly detection accuracy by better capturing distances between data points compared to PCA. Lastly, other clustering methods, such as density-based spatial clustering of applications with noise (DBSCAN) or Gaussian Mixture Models, can also be used to improve anomaly detection performance over KNN.

7. REFERENCES

[1] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander. Lof: Identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, pages 93–104. ACM, 2000.

[2] Chris Brooks and Sotiris Tsolacos. Forecasting models of retail rents. *Environment and Planning A*, 32(10):1825–1839, 2000.

[3] Yi Chen, Xin Liu, Xi Li, Yong Liu, and Xiang Xu. Mapping the fine-scale spatial pattern of housing rent in the metropolitan area by using online rental listings

and ensemble learning. *Applied Geography*, 75:200–212, 2016.

[4] Hamish Crosby, Peter Davis, and Theodoros Damoulas. A spatio-temporal, gaussian process regression, real-estate price predictor. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 1–4, 2016.

[5] Jonas A Dambon, Fabio Sigrist, and Reinhard Furrer. Maximum likelihood estimation of spatially varying coefficient models for large data with an application to real estate price prediction. *Spatial Statistics*, 41:100470, 2021.

[6] Amr Khalafallah. Neural network based model for predicting housing market performance. *Tsinghua Science and Technology*, 13(S1):325–328, 2008.

[7] F. T. Liu, K. M. Ting, and Z.-H. Zhou. Isolation forest. In *Proceedings of the Eighth IEEE International Conference on Data Mining (ICDM '08)*, pages 413–422, January 2009.

[8] Terry McGough and Sotiris Tsolacos. Forecasting commercial rental values using arima models. *Journal of Property Valuation and Investment*, 1995.

[9] Pierluigi Morano, Francesco Tajani, and Claudia Maria Torre. Artificial intelligence in property valuations: an application of artificial neural networks to housing appraisal. In *Advances in Environmental Sciences, Development and Chemistry*, pages 23–29, 2015.

[10] GMA News Online. Sibuyas sa pilipinas na mas mahal kaysa sa ibang bansa, dahil sa mahinang produksyon — ekonomista [video], 2023. Retrieved from `https://www.gmanetwork.com/news/video/24oras/613503/sibuyas-sa-pilipinas-na-mas-mahal-kaysa-sa-ibang-bansa-dahil-sa-mahinang-produksyon-ekonomista`.

[11] Byungkyu Park and Jeong Kyu Bae. Using machine learning algorithms for housing price prediction: The case of fairfax county, virginia housing data. *Expert Systems with Applications*, 42(6):2928–2934, 2015.

[12] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[13] J. Relativo. P200 kada kilo: Presyo ng galunggong, iba pang isda pinakokontrol, 2022. Retrieved from `https://www.philstar.com/headlines/2022/01/21/2155530/p200-kada-kilo-presyo-ng-galunggong-iba-pang-isda-pinakokontrol`.

[14] Xiaotong Xu and Yanbo Zhang. A gaussian process regression machine learning model for forecasting retail property prices with bayesian optimizations and cross-validation. *Decision Analytics Journal*, 8:100267, 2023.