**Introduction**

In today’s dynamic and interconnected global food market, predicting food prices accurately has become crucial for various stakeholders, including farmers, retailers, policymakers, and consumers. Fluctuations in food prices can impact food security, inflation rates, agricultural practices, and economic stability. Therefore, developing reliable predictive models to forecast food prices can provide valuable insights and aid decision-making processes in the food industry.

**Here are some reasons why predicting food prices is important**:

Economic Stability: Fluctuations in food prices can have significant implications for overall economic stability. Sudden spikes or drops in food prices can impact inflation rates, interest rates, and overall consumer spending.

Food Security: Predicting food prices helps governments and international organizations anticipate and mitigate potential food shortages or crises. By identifying potential price increases or supply chain disruptions in advance, policymakers can implement measures to ensure food security for vulnerable populations.

Business Planning: For businesses involved in food production, distribution, and retail, accurate price predictions are essential for strategic planning and decision-making. By forecasting future prices, businesses can optimize production levels, manage inventory effectively, and adjust pricing strategies to remain competitive in the market.

Consumer Welfare: For consumers, especially those with limited financial resources, fluctuations in food prices can have a direct impact on their purchasing power and overall well-being. Predictive models can help individuals and households plan their budgets more effectively.

Data Collection: The dataset used in this project was obtained from Statistics Canada, providing valuable insights into food prices across different regions and time periods (we analyzed data from 2007). In our basic CSV file, we have only three features GEO, Year, Month

Feature 1: GEO: Geographic location can significantly influence food prices due to variations in climate, soil fertility, transportation costs, and market demand. We encoded the geographical regions where the data was collected as categorical variables (Western Region, Eastern Region)

To enrich our analysis, incorporated additional datasets such as:

consumer\_price\_index\_energy.csv,

consumer\_price\_index\_mort\_int.csv',

consumer\_price\_index\_food.csv,

consumer\_price\_index\_gasoline.csv,

consumer\_price\_index\_all\_items.csv',

consumer\_price\_index\_shelter.csv'

As a result, we’ve got additional features:

Consumer Price Index (CPI) for various categories, including 'Energy', 'Mortgage and Interest', 'Food', 'Gasoline', 'Shelter', and 'All Items'.

**target variable** (dependent variable), we want to predict the 'value' column, which represents the commodity price.

**As a result, we’ve got additional features:**

Feature 2: Energy Prices (Energy) Energy prices may affect food production and transportation costs, thus influencing food prices.

Feature 3: Mortgage and Interest Rates (Mortgage\_and\_Interest): Mortgage and interest rates can impact consumer spending power and overall economic conditions, which may affect food demand and prices.

Feature 4: Gasoline Prices (Gasoline) Gasoline prices can influence transportation costs for food distribution, potentially affecting food prices.

Feature 5: Shelter Costs (Shelter) Shelter costs, including rent and housing prices, may affect household budgets and discretionary spending on food.

Feature 6: Overall Consumer Price Index (All\_Items) The overall Consumer Price Index reflects general inflation trends, which can impact food prices along with other goods and services.

Feature 7: Month Time-related variables such as year and month can capture seasonal or temporal patterns in food prices, such as fluctuations due to harvest seasons or holidays. We can encode the months as categorical variables (e.g., January, February, ..., December) or use numerical representations (e.g., 1 for January, 2 for February, etc.) to account for seasonal effects.

Feature 8: Yearly trends can reflect broader economic conditions, policy changes, technological advancements, and global market dynamics that affect food prices.

First, we observe our dataset to see how the price changes for all commodity groups over time. This plot is interactive at the bottom we can see the price for each month.

**Objectives**

The primary objective of this project is to develop and compare different regression models to predict food prices based on various influencing factors.

**First I implemented Linear regression models:**

Selection of target variable and features predictors;

Split the data into training and testing sets;

Fit the model with dataset of selected features;

Train a linear regression model on the training data

Use the model fitted by training data to validate on the testing dataset

Evaluate the model performance and, if acceptable,  use the model to make predictions on labels for new data.

Build the plot with actual and predicted values

**Than Identifying which features are important;**

Build bar chart with feature importance coefficients

The most important features are:

Food

Mortgage and interest

Shelter

The second Linear regression model I build with the most important features and got another plot that was similar to the first model

As a second model, I build Decision Tree Regression

I download Decision Tree into PNG file

Unfortunately, because it was huge on the PNG file I can see only lines and not a tree

Then, I compare the evaluation matrix for both models and

Overall, the second model seems to perform exceptionally well, especially in terms of R-squared (R²)R² and Root Mean Squared Error RMSE,

suggesting that it may be overfitting the data or there might be some other issues.