Agricultural Product PRicing

James Bootsma, Aditya Manickam Arumugam, Krishang Karir, Abdul Aziz Essam, Mohd Daiyaan Nasir

## Abstract

Farm product prices are influenced by a multitude of factors, including climatic conditions and economic variables, making their prediction a complex but critical task for stakeholders in the agricultural sector. This study aims to develop a predictive model that forecasts future farm product prices by analyzing historical climate data and relevant economic indicators. Utilizing a multivariate time series approach, we explore the efficacy of both traditional time series models—specifically ARIMA (AutoRegressive Integrated Moving Average) and LSTM (Long Short-Term Memory) networks—as well as standard regression models, including Linear Regression and Random Forest Regression. By integrating data from Environment and Climate Change Canada and Statistics Canada, we examine the impact of climatic variables such as temperature and precipitation, along with economic indicators like GDP and oil prices, on the prediction of farm product prices. Our analysis focuses on feature engineering to create meaningful variables and ensemble models to enhance prediction accuracy. The goal is to identify the model that provides the most accurate forecasts and actionable insights, thereby enabling more informed decision-making for farmers, consumers, and policymakers. Preliminary results indicate that advanced machine learning techniques, particularly LSTM networks, offer significant improvements in predictive performance over traditional methods, highlighting the potential for these models to better capture the complex temporal patterns influencing farm product prices.

## Background

The agricultural sector faces significant volatility due to the complex interplay of climatic and economic factors affecting farm product prices. This unpredictability is exacerbated by climate change, which introduces extreme weather events impacting crop yields and prices. Traditional forecasting methods, such as ARIMA, often fall short in capturing the non-linear and lagged effects inherent in these dynamics. As a result, there is a growing need for more sophisticated models that can integrate diverse data sources, including climate and economic indicators, to enhance prediction accuracy. This need is critical for stakeholders like farmers, distributors, and policymakers, who rely on accurate forecasts for financial stability, resource allocation, and strategic planning.

Recent advancements in machine learning, particularly the use of Long Short-Term Memory (LSTM) networks, offer promising solutions to these challenges. LSTM models, a type of recurrent neural network, are adept at handling complex temporal patterns and long-term dependencies in time series data. They outperform traditional methods by capturing intricate interactions between variables, providing more reliable forecasts. This project leverages these advanced techniques, integrating data from sources like Statistics Canada and Environment and Climate Change Canada, to develop a comprehensive predictive model. By doing so, it aims to improve the accuracy and reliability of agricultural price forecasts, ultimately supporting more informed decision-making and contributing to the resilience and sustainability of the agricultural sector.

Our model strategically incorporates GDP, temperature, precipitation, gas, and oil prices as predictor variables due to their profound influence on agricultural markets. GDP reflects the overall economic health and consumer purchasing power, directly affecting the demand for agricultural products. Temperature and precipitation are crucial climate variables that impact crop yields, growing seasons, and harvest quality, thereby determining supply levels. Gas and oil prices are critical cost indicators, influencing production costs through farm equipment operation and fertilizer production, as well as transportation expenses across the agricultural supply chain. These energy prices have shown strong correlations with agricultural commodity prices, especially in modern farming systems reliant on mechanization and long-distance transportation. By integrating these variables, our model captures the intricate interactions affecting agricultural prices, providing a robust and accurate forecasting tool.

## Data Collection

To collect and prepare the necessary data, we begin by gathering agricultural, economic, and climate datasets from various sources such as Statistics Canada and Environment Canada. These datasets include monthly farm product price data, GDP data, oil prices, and weather conditions across different provinces and territories.

The farm product price data is cleaned by standardizing product names, creating a comprehensive date range, and ensuring that all relevant combinations of date, province, and product are included. Missing values are addressed using a function that imputes prices based on ratios to the mean price for each product. The data is then structured with key columns such as province, date, product, and price, ensuring that the dataset is ready for analysis.

For the GDP data, we focus on agricultural production values and filter for 'Chained (2017) dollars.' Missing provincial GDP values are estimated by calculating mean ratios and using these to fill gaps. The data is resampled monthly, ensuring that the monthly totals match annual GDP figures. Logarithmic changes in GDP are calculated to facilitate year-over-year comparisons.

Oil price data is processed by extracting relevant columns and grouping by province, year, and month. The data is transformed into a pivot table for different types of fuel, and missing diesel price data is filled using alternative values. The log change in oil prices is then calculated for each month, providing insights into fuel price trends.

Weather data is collected monthly, and unnecessary columns are removed. Missing data is interpolated for each province separately, creating a continuous dataset from 1980 to the present. This dataset is reindexed to include all combinations of date and province, ensuring no missing entries.

After merging all datasets, the log-transformed product price is calculated to normalize price fluctuations. Year-over-year (YoY) price changes are computed by comparing the log prices for a given month to the same month in the previous year. Lagged variables are created to account for the temporal relationship between product prices, GDP, oil prices, and weather conditions. The final dataset is filtered for specific years, irrelevant columns are dropped, and any rows with missing data are removed before saving the cleaned data for further analysis.

This process results in a comprehensive dataset that enables the analysis of agricultural product prices in relation to economic and environmental factors over time.

## ANALYSIS PLAN

In this analysis, the objective was to develop a predictive model for forecasting farm product prices by analyzing the effects of climatic and economic factors. Various modeling approaches were employed, including ARIMA (AutoRegressive Integrated Moving Average), Long Short-Term Memory (LSTM) networks, Linear Regression, and Random Forest Regression. Each of these models utilized the same set of predictors: temperature, precipitation, GDP, and oil prices, sourced from Environment and Climate Change Canada and Statistics Canada.

To evaluate each model’s effectiveness, ARIMA was selected for its capacity to handle time series data, allowing it to capture autocorrelation within price fluctuations. The LSTM network, as a deep learning model, was trained to leverage temporal dependencies, making it suitable for capturing long-term patterns in complex datasets. Random Forest Regression, an ensemble technique, was also used to explore potential non-linear relationships between variables. Linear Regression served as a straightforward baseline model for comparison.

Each model was evaluated based on its predictive performance on validation data. The final model selection process involved identifying the model that performed best overall. Additionally, we considered the possibility of applying different models for different categories of farm products, depending on their unique price dynamics.

Preliminary results indicate that advanced machine learning techniques, especially the LSTM model, may better capture the complex interactions between climatic and economic factors affecting farm product prices. This approach has the potential to generate actionable insights for farmers, consumers, and policymakers, supporting more informed decision-making in the agricultural sector.

VISUALIZATION PLAN:  
  
To illustrate model comparison, bar charts and line graphs will highlight the performance (e.g., RMSE, MAE) of each predictive model across different product categories. Bar charts provide a straightforward comparison of accuracy, while line graphs are suitable for tracking prediction errors over time. This will help users see which model performs best under different conditions.

For time series analysis, line charts will be used extensively to track historical price trends, seasonal fluctuations, and year-over-year price changes. These charts allow for a clear visualization of long-term patterns and cyclicality. Additionally, we will use multi-line charts to plot price trends alongside relevant economic and weather variables, showcasing how different factors move in tandem or diverge over time.

Forecasting insights will be presented using line or area charts with confidence intervals, indicating predicted future prices and their reliability range. Scenario simulations, where users can manipulate variables (like oil prices or GDP), can be displayed using interactive sliders that instantly update forecast visualizations. This enables stakeholders to simulate various economic and climatic scenarios.

To illustrate feature importance, bar charts will display the significance of each variable in influencing product prices, based on Random Forest model output. Heatmaps or scatter plots will be used for sensitivity analysis, where users can observe how variations in specific variables (e.g., oil prices, temperature) affect predicted prices. These techniques are effective for identifying critical price drivers and for exploring the sensitivity of the model’s predictions.

For geographic insights, choropleth maps will visualize regional variations in product prices, GDP, and weather conditions, allowing for easy identification of trends across provinces. Line charts for each region will further enable users to compare price patterns and economic variables on a regional basis over time.

Log-transformed line charts will be used to show volatility and changes in GDP and oil prices on a more comparable scale, offering a clear view of trends in economic indicators with potentially high variance.

The primary libraries we’ll use for these visualizations include Matplotlib and Seaborn for static visualizations, Plotly for interactive charts, and Mapbox or Plotly for geospatial data. In the dashboard environment, Power BI or Tableau will integrate these visualizations for dynamic interaction and filtering, enabling an accessible and visually rich experience for stakeholders.

This combination of libraries and visualization types provides a flexible and powerful approach to explore and communicate complex data relationships, allowing users to draw insights from multiple economic and climatic factors influencing farm product prices.

1. Baffes, J. (2013). A framework for analyzing the interplay among food, fuels, and biofuels. Global Food Security, 2(2), 110-116.
2. Baldi, L., Peri, M., & Vandone, D. (2016). Stock markets' bubbles burst and volatility spillovers in agricultural commodity markets. Research in International Business and Finance, 38, 277-285.
3. Chen, S. T., Kuo, H. I., & Chen, C. C. (2010). Modeling the relationship between the oil price and global food prices. Applied Energy, 87(8), 2517-2525.
4. Lobell, D. B., & Field, C. B. (2007). Global scale climate-crop yield relationships and the impacts of recent warming. Environmental Research Letters, 2(1), 014002.
5. Nazlioglu, S., & Soytas, U. (2012). Oil price, agricultural commodity prices, and the dollar: A panel cointegration and causality analysis. Energy Economics, 34(4), 1098-1104.
6. Ray, D. K., Gerber, J. S., MacDonald, G. K., & West, P. C. (2015). Climate variation explains a third of global crop yield variability. Nature Communications, 6(1), 1-9.
7. Rezitis, A. N., & Sassi, M. (2013). Commodity food prices: Review and empirics. Economics Research International, 2013.
8. Wang, Y., Wu, C., & Yang, L. (2014). Oil price shocks and agricultural commodity prices. Energy Economics, 44, 22-35.
9. Xiong, T., Li, C., & Bao, Y. (2017). Seasonal forecasting of agricultural commodity prices using a hybrid STL and ELM method. Neurocomputing, 275, 2548-2558.
10. Zhang, W., & Carter, C. A. (2018). State trading enterprises and price transmission: The case of Chinese agricultural commodity imports. American Journal of Agricultural Economics, 100(5), 1322-1344.