

A MACHINE LEARNING APPROACH TO FORECASTING CONSUMER FOOD PRICES

Jay (Jabez) Harris
Dalhousie University
6299 South St.,
Halifax, NS Canada
(902) 495-7224
Jay.harris@dal.ca

ABSTRACT

Building on the success of the Canada Food Price Report 2017 and its inclusion of a machine learning methodology, this research paper poses and attempts to answer the following question, “What is the best way to predict food prices for the average Canadian consumer?” The research incorporates a popular benchmark model, a financial futures-market data model and an advanced feature selection process into the methodology utilized in the Canada Food Price Report 2017. The hope is to create a more robust model for future Canada Food Price Reports.

As hypothesized, the Financial Futures-Market based model outperformed the Food Price Report model with 1.6% and 2.4% average error rates respectively. Similarly, both models outperformed the Holt-Winters baseline method. Except for the CPI Seafood category in the case of the Food Price Report model, both models easily bested the popular baseline model.

The CPI Restaurant category which is regarded as the most difficult to forecast produced the lowest error rate for both the Food Price Report and Financial Futures-Market models. (Joutz F. L., 1997). Given the superb performance by both the Food Price Report and Financial Futures-Market models it is likely that even better results can be achieved by combining the datasets to share their uncommon information.

Keywords

Machine Learning, Consumer Price Index, Food Price, Inflation, Forecast, Financial Futures Market.

1. INTRODUCTION

1.1 Why forecast food prices?

As of January 2015, the average Canadian household spends 62.3% of its total income on food, shelter and transportation costs (Statistics Canada, 2015). Though this estimate may seem outrageous, in our hearts Canadians know that it is indeed true. When we consider our current socio-economic situation, we realize that Canadians from all provinces and all walks of life have been affected by rising food costs, runaway housing prices and increasing transportation expenses.

The Hunger Count 2016 report from Food Banks Canada revealed that almost 75% of recipients of foodbank aid in Canada in 2016 had secured market-rate housing through rent or an established path to home ownership, however to accomplish this, many Canadians were unfortunately forced to choose between shelter and food. This phenomenon, which has become prevalent in many developed countries, has contributed to a 28% increase in food bank visits in Canada since 2008 (Food Banks Canada, 2016).

The National Bank of Canada’s House Price Index estimates an increase of almost 13% in home prices in Canada’s six largest metropolitan areas during the 12-month period from November 2015 to November 2016. In Vancouver, a single-family home priced below \$1 million is now a rarity, as home prices continue to skyrocket amid an influx of foreign investors who view Canada as a gateway to previously inaccessible western markets.

Accordingly, Statistics Canada also reported in 2015, public transportation costs had soared to 36.5% higher than they were in 2002 (Statistics Canada, 2016b). To put this into context, adult bus fare in Toronto in 2002 could be purchased for \$2 cash but in 2016 that same bus fare is now \$3.25 cash for an increase of 62% in bus fares alone. This equates to an additional \$50 in monthly expenses or \$600 annually for a single person commuting to work daily since 2002. Similarly, for privately used transportation methods, gasoline prices in Toronto have risen by 50% since 2002 (Statistics Canada, 2016a). This represents an increasing burden on the Ontario consumer, who though having experienced some wage increase since 2002 with a growth rate in average weekly wages of 40%, is still struggling to stay ahead of inflationary prices (Statistics Canada, 2016c).

In Canada, food prices represent a cost variable which can still be manipulated by consumers to circumvent rising costs in shelter and transportation. The cauliflower crisis of 2016 illustrates how the Canada marketplace is still very susceptible to the demands of its consumers. In January 2016, a New York Times article titled, In Canada, the 8-Dollar cauliflower shows the pain of falling oil prices, highlighted how low commodity prices coupled with a devalued Canadian dollar and deficiencies in the supply chain had culminated in exorbitant food prices for Canadian shoppers. However, unlike similar events in the United States, Canadian shoppers through social media-styled protests managed to convince the major food retailers to absorb the excess costs being perpetuated by an unusually dry season in the farm intensive areas of California (Charlebois, et al., 2016).

Food prices are a moving target but with careful planning and consideration it should be possible to endure the price hikes in other consumer good categories by adjusting our food consumption habits. However, to confidently take advantage of volatile food prices would require not only planning but also the ability to accurately forecast food price inflation and initiate appropriate stop-gap measures, such as increasing social aid, adjusting agricultural trade policies, pursuing new vendors, introducing price ceilings or addressing inefficiencies in distribution. We must also consider the global marketplace and Canada’s vulnerability to the impact of rising global food prices given that the country’s food imports as a percentage of its merchandise imports, is at the highest it has been since the 1970s (The World Bank, 2015). At the same time Canada’s food

exports, as a percentage of its merchandise imports, are only now recovering from a 30-year low and near stagnant growth from 2010 to 2014 (Statistics Canada, 2016).

Inflation is seen by many economists as a necessary evil in maintaining a thriving and efficient economy but too low of an inflation rate could spell disaster for local businesses and economic policy. The proverbial sweet spot for inflation rates is believed to be between 1% and 2% annually and not 0% or even the occasional decline as many consumers would advocate for. The reasons for a positive and controlled inflation rate are (Billi & Kahn, 2008):

1. The available measures of inflation are not perfect and can overstate the “true” inflation rate.
2. A small amount of inflation may make it easier for firms to reduce real wages when necessary to maintaining employment levels.
3. A negative inflation rate or deflation, could be even more costly than a similar rate of inflation due to lost incomes, which suggests that a low rate of inflation might be more advantageous for protecting against the residual effects of falling prices.
4. At very low levels of inflation, nominal interest rates may be close to zero, limiting a central bank’s ability to ease monetary policy in response to economic weakness.

All in all, a small amount of inflation is believed to be necessary to allow manufacturers and service providers to improve the quality of their wares annually without putting any undue burden on the consumer or the economy.

So, what exactly is inflation? Per the International Monetary Fund’s quarterly periodical, *Financial Development*, it is the measure of how much more expensive a set of goods and services have become over a certain period. It is usually calculated with a broad scope such as the overall increases in consumer prices or cost of living increases for a specific sovereign nation (Öner, 2012). Alternatively, its scope can also be narrowed to focus in on individual price changes like food, services, education or fuel. In Canada, the Consumer Price Index (CPI) is the most recognized measure of inflation. The CPI can be thought of as a measure of the percentage change, over time, of the average cost of a large basket of goods and services purchased by an average Canadian consumer. The basket contains good and services, that are kept constant over time in quantity and quality to ensure consistency across all applicable years and generations of Canadians. Thus, changes to the cost of the basket overtime are not due to changes in the quantity or quality of the goods and services observed but rather, reflect only pure price movements (Statistics Canada, 1996).

Of course, forecasting of any naturally occurring event is relatively difficult and even impossible in some instances, especially when we consider the limitations of our own minds when handling enormous amounts of quantitative data. For food prices, which are impacted by international and domestic factors, it is almost impossible for an individual to comprehend the full scope of attributes that can influence the cost of a single product. In addition to location based factors, food prices are also affected by countless other forces including supply, demand and the value of the trading currency.

The ability to accurately forecast price increases and circumvent inflation will become one of the most prioritized accomplishments for any country that continues to thrive in the 21st century. Governments, businesses and consumers engaged in forecasting

prices and interpreting price information are likely find support for their efforts in the domains of machine learning and statistical analysis as these methods of generating meaningful information have the potential to precisely predict price increases and inflation rates. Machine learning is a multifaceted discipline in which there are numerous proven techniques for identifying strategic information and solving complexed problems. Statistical analysis has existed for over a century and is routinely used in business and academia to derive meaningful information from stores of data. In the realm of food price predictions, the combination of both disciplines could not only generate knowledge, which could potentially forecast important increases in costs but also can frame newly discovered knowledge in a manner that can be comprehended with relative ease by all levels of society. These results could then be used to inform consumer budgets, maintain supermarket inventories or even influence federal trade policy. To finally overcome the struggles caused by the volatility of price inflation as it silently plagues the Earth, would require an extraordinary endeavor and this research paper aims to deliver an illustration of that effort in the form of machine learning models and statistical decision making.

1.2 Research Objectives

Traditionally, food forecasts have been performed by financial and economic reporting firms which speculate internally to form general assumptions and make predictions surrounding demand and supply. These assumptions and predictions of supply and demand are then used to as the basis for their forecasting tools. The results are as would be expected, widely varying in predicted value and error rate (Joutz F. L., 1997). The Economic Research Services (ERS), an arm of the United States Department of Agriculture (USDA) is the only United State Government entity which regularly examines food prices and produce forecasts. In its own 2000 report on its forecasting processes from 1984 to 1997, the ERS revealed that simple univariate models produced using the menu-based time series forecasting system available in SAS produced a lower Root Mean Square Error (RMSE) in 7 of 23 price indexes and comparable RMSE rates in another 13. The report indicates that the long standing ERS approach to forecasting through a process of predetermined links to commodity supply and demand was only able to outperform the baseline models in 3 of the 23 price indices (Joutz, Trost, Hallahan, Clauson, & Denbaly, 2000). To put it bluntly, the ERS with all its knowledge and assumptions was outperformed by baseline statistical models which can be easily accessed by any non-expert forecaster.

Machine learning forecasts have become popular and even common place in financial and commodities markets (Kim, 2003) (Cao & Tay, 2003) (Ticlavilca, Feuz, & McKee, 2010). This is likely due to the influx of physicists and mathematicians to the financial industry after the Cold War. Wall Street and other major financial markets took advantage of the availability of these highly skilled individuals and set them on the task of forecasting market prices. On the other hand, economics which is the field where food prices are usually examined, is widely regarded as a humanities subject, separate from business and finance, and has been less impacted by this trend. Economists, have for decades attempted to connect commodity prices and consumer prices but as Blomberg and Harris indicate, while this may have been true in the 1970’s and 1980’s, the same theory does not apply today. Commodities simply do not hold the same importance that they once did. The food processing industry as well as financial markets have increased in complexity to such an extent that input commodities prices have little impact on the cost of the final

output and any financial shocks at the commodity level have only a slight impact on the overall economy (Blomberg & Harris, 1995).

This research paper with its machine learning approach is aimed at breaking the mold which constraints food price forecasts. By incorporating machine learning techniques for data analysis over the conventional approach of expert generated assumptions, the hope is to create a global view of the marketplace to produce more accurate forecasts. In bypassing the commodity specific connections to food prices by using advanced feature selection processes to determine strong relationships, more complex views of the marketplace can be created and further analyzed by the machine learning models to produce far more robust forecasts.

1.3 A Machine Learning Approach

The primary objective of this research paper is to employ a machine learning approach to the creation of econometric models, which can then be used to forecast food prices in the major food categories outlined by the Canada Consumer Price Index as published by the Government of Canada. The research will compare a benchmark model currently being used by financial analysts and economists to identify simple trends in a time-series data, with two other models designed to use historical datasets as predictors.

1.3.1 Benchmark Model:

The Holt-Winters exponential smoothing method is quite popular in statistical and economic analysis for forecasting data points in a time-series. The Holt-Winters method incorporates baseline, trend and seasonality analysis to forecast data points of seasonal data, it will serve as the benchmark comparison for the other predictor-based models. To determine the presence of seasonality in the datasets, each CPI category will be tested for statistical stationarity which investigates whether properties such as mean and variance are constant over time.

1.3.2 Food Price Report Model:

A dataset identical to that used in the Food Price Report 2017 will serve as the data source of several machine learning models. The dataset consists of historical data features, identified as being correlated with the volatility of food prices and the Canadian consumer price index.

Input	Duration	Source
CAD vs USD	1999 - 2016	Canada Forex
Canada CPI Energy	1999 - 2016	Statistics Canada
Canada Deposable Income	1999 -2016	Statistics Canada
Canada Household Farm Income	1999 -2016	Statistics Canada
Canada Household Income	1999 -2016	Statistics Canada
Canada Household Net Saving	1999 -2016	Statistics Canada
Canada Immigrant Mean Income	1999 – 2014	Statistics Canada
Canada Immigrants with Income	1999 – 2014	Statistics Canada
Canada Income Distribution	2000 – 2014	Statistics Canada
Canada Net Lending or Borrowing	1999 – 2016	Statistics Canada

Input	Duration	Source
Unemployment Rate	2000 – 2016	Statistics Canada
Precipitation	1999 – 2015	U.S. EPA
Temperature	1999 – 2016	U.S. EPA
FAO Commodity Index	1999 - 2016	FAO
Gross Agricultural Production	1999 - 2013	FAO
International Aid Dollars	1999 - 2016	FAO
Population	1999 - 2015	FAO
Agriculture Fertilizers	1999 - 2014	FAO
Canada Credit to Agriculture	1999 – 2015	FAO
US Credit to Agriculture	1999 - 2015	FAO
Crude Oil Prices	1999 - 2015	U.S. Energy Information Administration
US Overnight Lending Rates	2000 - 2016	Federal Reserve
Commodity Market Food Price	1999 - 2016	World Bank
Diesel Prices	1999 - 2014	World Bank
Gasoline Prices	1999 - 2014	World Bank

Table 1: Sources of Food Price Report dataset features

1.3.3 Financial Futures-Market Based Model:

Owing to its purpose, the financial futures-market inherently attempts to predict commodity prices at the market level. This research paper speculates that these predicted market prices could themselves be predictors of consumer prices. The financial futures-market based model is built on historical closing prices of commodities in the financial futures-market.

Input	Duration	Source
Soybeans(CBOT)	2001 - 2015	Moore Research Center
Soybean Meal(CBOT)	2001 - 2015	Moore Research Center
Soybean Oil(CBOT)	2001 - 2015	Moore Research Center
Canola(WCE)	2001 - 2015	Moore Research Center
Corn(CBOT)	2001 - 2015	Moore Research Center
Wheat(CBOT)	2001 - 2015	Moore Research Center
Wheat(KCBT)	2001 - 2015	Moore Research Center
Wheat(MGE)	2001 - 2015	Moore Research Center
Oats(CBOT)	2001 - 2015	Moore Research Center
Rough Rice(CBOT)	2001 - 2015	Moore Research Center
Live Cattle(CME)	2001 - 2015	Moore Research Center
Feeder Cattle(CME)	2001 - 2015	Moore Research Center
Lean Hogs(CME)	2001 - 2015	Moore Research Center
Class III Milk(CME)	2001 - 2015	Moore Research Center
Cocoa(ICE)	2001 - 2015	Moore Research Center
Coffee "C"(ICE)	2001 - 2015	Moore Research Center

Orange Juice(ICE)	2001 - 2015	Moore Research Center
Sugar #11(ICE)	2001 - 2015	Moore Research Center
London Cocoa(LCE)	2001 - 2015	Moore Research Center
London Sugar(LCE)	2001 - 2015	Moore Research Center

Table 2: Sources of Financial Futures-Market dataset features

1.3.4 Model Evaluation:

The best performing model will be identified by its ability to achieve the lowest Mean Absolute Percentage Error (MAPE) below that of the Holt-Winters benchmark model. As a tool for hedging and speculation, futures trading provides exceptional convenience and economies of transactions allowing it to possess similar qualities to that of spot markets and is consequently ideal for forecasting prices (Working, 1976). It is hypothesized that the financial futures-market based model will achieve the lowest Mean Absolute Percentage Error (MAPE) as compared to the food price report model since accurate price predictions are a critical component of trading in the financial futures-market.

2. FOOD PRICES

2.1 Historical, Global Context

Food prices have been on the rise since the mid-2000s and have now become parallel with the prices of other top commodities. During this period, both developed and developing countries have witnessed the hardships experienced by consumers and producers as the pressure of higher food prices along with higher input costs, perpetuates a cycle of increasing prices. The most obvious driver of the price spike in the previous decade has been the worldwide increase in demand. Since World War II and after the introduction of the Baby Boomer generation the world's population has tripled in size and with more people on earth generally living longer lives, under much better conditions than their pre-World War II counterparts, we have been able to consume exponentially more food, products and services (European Union, 2015).

The increased demand for food from both developed and developing countries has caused food commodity prices to skyrocket on the international market. In the domestic markets, increases in income and improvements to the accessibility of foreign food markets have enhanced the demand for a variety of foods which in turn caused price increases which have trickled down from the commodities to the consumers. Between 2005 and 2007, just prior to the United States housing crisis and subsequent global recession, the commodity prices of wheat, coarse grains, rice and oilseed crops all nearly doubled. Meanwhile from 2005 to 2008 the Food and Agricultural Organization's food price index indicates an 83% increase in global food prices at the consumer level. This increase was almost three times (3x) higher than the previous 3 years, 2002 to 2005, and almost three times (3x) higher than the following 3 years, 2008 to 2011, (Food and Agriculture Organization, 2017).

The causes of this spike in prices are complex and due to a combination of mutually reinforcing factors (Organization of Economic Co-Operation and Development, 2008) including:

1. Droughts in key grain-producing regions that created negative yields
2. Low stocks for cereals and oilseeds caused by increased demand in developing countries
3. Increased feedstock use in the production of biofuels owing to policy-driven demand

4. Increased production costs brought on by rapidly rising oil prices and price jumps due to a continuing devaluation of the US dollar as the currency in which indicator prices for these commodities are typically quoted.

Production yields and macroeconomic conditions have also had a significant impact on food prices. The negative yields, such as those observed in 2015 – 2017 due to unusual weather conditions in California, may have been temporary but growing concerns for the impact of long-term climate change and pollution may prove to be well-founded as environmentalists continue to forecast more permanent effects which will undoubtedly also impact food prices.

Meanwhile, the macroeconomic conditions that favor economic growth, increases in purchasing power, and stronger demand for agricultural commodities are a permanent factor in the determination of food prices. Additionally, the demand for food is less responsive to changes in prices when the purchasing power of the consumer increases. So, with each country actively striving for economic growth it is not difficult to see how food prices can increase to inflationary rates in the absence of reactionary demand as brought on by growing income rates among consumers in many developing countries. In the context of the financial world, this type of non-participatory price increase in food would be equivalent to an impossible situation where stock prices could be manually manipulated by the listed companies without consequence and without the involvement of the market, who would then be expected to honor the new prices. Consumers of food products eventually lose some of their purchasing power when their income increases because higher prices will eventually become the standard for all products and services.

The thinness of markets, or the share of imports and exports relative to the size of global consumption or production can also have large scale effects on global prices because of protectionist federal policies and imperfections in domestic markets. In other words, increased demand for a particular food item in one country, such as corn for biofuels in the United States, could have the effect of price increases in the US domestic market which are not aligned with the price rates of the international trade. Thus, world market prices would adjust to accommodate the external shock to demand and traded quantities. In the case of biofuels, policy-driven demand for biofuel production is generally considered less responsive to prices than traditional food demand and is a strong factor in the increase of fuel and animal feed prices (Organization of Economic Co-Operation and Development, 2008).

2.2 Canadian Context

Originally named the Food Price Index (FPI), The Canada Food Price Report was created at University of Guelph by Dr. Sylvain Charlebois and Dr. Francis Tapon. Since 2010, the Food Price Report has been published as a tool which focuses on factors affecting the future of consumer food prices over a 12-month period (Wilson, 2015). The report is now published by Dalhousie University and includes a machine learning methodology which is further supplemented by the expert advice of its prominent authors.

The 2017 report features a range of models from Multivariate Linear Regression to Neural Networks and twenty independent variables, identified as potential inputs due to high correlations with certain categories of the consumer price index. Also, included in the report is an interpretation of economic drivers, fundamental to the movement of food prices in Canada. These

economic drivers are sub-divided into three main categories, macro drivers such as energy cost and climate change, sectorial drivers such as the food processing industries and consumer food awareness, and domestic drivers, such as consumer debt, consumer income and income distribution.

The authors' understanding of these drivers plays a critical role in the report's forecast and each group of drivers is updated yearly to highlight the predicted impact on food prices. Additionally, the individual drivers formed the basis for the twenty independent variables utilized in the 2017 report. It is helpful to consider how the size and impact of these groups of drivers has changed over the years to maintain alignment with the current state of affairs. While the number of macro and sectorial drivers has slightly increased in the past five years, the number of domestic drivers has decreased. Although, the authors offer no specific explanation for the removal of the domestic driver, it appears that the local influences on food prices had taken a back seat to international pressures. Given that Canada imports most its food, roughly 80% of it fresh fruit and vegetables, it is not difficult to see how domestic drivers could be eliminated and replaced by much more influential global drivers.

In this research project, the machine learning model plays a role like that of the report authors, in that, it attempts to determine the impact of each group of variables then uses this knowledge to build a predictive model.

It is not clear if the machine learning model will uphold the theories of the report authors regarding the importance of the macro-economic drivers. Although, an investigation of the attribute selection process incorporated into the research methodology could prove helpful in addressing this issue. However, such a comparison is outside the scope of the research paper and unfortunately will not be included in the analysis of the model. Nevertheless, the similarities between the feature selection process and the authors' views on the changing macro, sectorial and domestic drivers should not be ignored.

3. RESEARCH METHODOLOGY

3.1.1 Consumer Price Index – Dependent Variable:

The primary goal of the machine learning model is to forecast, for the coming year, the Canadian Consumer Price Index (CPI), which is published each month by Statistics Canada. The consumer price index was chosen for this task because it is the de facto measure of inflation in Canada and is in fact used extensively by the Bank of Canada to chart inflation and adjust interest rates. In this way, the Consumer Price Index becomes the initial dependent variable as well as the final output produced by the model. Eight categories of the consumer price index are examined during the process:

1. Meat: Fresh or frozen beef, pork, poultry and other processed meats such as ham and bacon, excluding seafood.
2. Seafood: Fresh, frozen, canned or otherwise preserved seafood including fish and other marine products.
3. Dairy: Fresh milk, butter, cheese, ice cream and other related processed products, also includes eggs.
4. Bakery: Breads, cookies, crackers and other bakery products along with rice, pasta, flour and cereal products excluding infant formulas.
5. Fruit: Fresh and preserved fruit including apples, oranges and bananas along with fruit juices and nuts.

6. Vegetables: Fresh, frozen, dried, canned or otherwise preserved vegetables including potatoes, tomatoes and lettuce.
7. Other: Sugar and confectionery products, edible fats and oils, coffee and tea, condiments, spices, vinegars, soups, infant formula, pre-cooked and frozen food preparations, non-alcoholic beverages.
8. Food Purchased from Restaurants: Table-service, fast food, take-out, cafeterias and other restaurants.

Ultimately, three models will be developed (benchmark, food price report, financial futures-market) to forecast each food category of the consumer price index by using various unique datasets of publicly available historical records. The Consumer Price Index is the actual value being predicted by the models and the accuracy of these predictions is used to evaluate the performance of each model to determine which is the best performing model per food category.

3.1.2 Holt-Winters – Benchmark Model:

The first step in the process of determining the best performing model per food category is to create a benchmark calculation, which will be used to judge the performance of the other computer-generated models. The Holt-Winter Triple Exponential Smoothing algorithm represents a popular form of time-series analysis used in statistics and econometrics that has been peer reviewed and proven to deliver satisfactory results and as such it has the honor of being the measuring stick for the more data intensive, food price report and financial futures-market based models.

The Holt-Winter Triple Exponential Smoothing algorithm is specifically designed to cope with trends and forecast seasonal data, i.e. data points that are repetitive over some period. The triple smoothing method is useless on non-seasonal data, so to determine the presence of seasonality in the consumer price index a stationarity test is performed on each food category of the index using a XLSTAT version 19.02.44125. Per XLSTAT, a time series is said to be stationary if its statistical properties such as mean, variance and autocorrelation do not vary with time. Simply, the stationarity tests will determine whether the properties of the consumer price index datasets depend on the time at which the time series is observed. If a dataset is stationary, then it lacks the trends and seasonality which would allow it to be effectively modelled by the Holt-Winters Triple Smoothing algorithm, (Hyndman & Athanasopoulos, 2013).

Per the XLSTAT website, the software is a suite of statistical add-ins for Microsoft Excel that has been developed since 1993 by Addinsoft to enhance the analytical capabilities of Microsoft Excel. Since 2003, Addinsoft has been a Microsoft partner and all the XLSTAT analytical add-ins are registered with the Office Marketplace. The XLSTAT software relies on Microsoft Excel for the input of data and to display results. This makes the software very convenient for exporting data and results. Computations are done using autonomous software components that are optimized for speed and efficiency. XLSTAT has been benchmarked against other popular statistical packages to ensure the integrity of its calculations.

The Holt-Winters Triple Exponential Smoothing method is accessed via the Waikato Environment Knowledge Analysis (WEKA), version 3.8.0. WEKA is the product of the University of Waikato in New Zealand and was first implemented in its modern form in 1999 and is licensed under the GNU General Public License (GPL). WEKA is a data mining workbench and consists of a collection of machine learning algorithms for data

mining tasks. The WEKA suite contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization as well as several add-on packages for specialized tasks in areas such time-series analysis. It incorporates several popular machine learning techniques into a streamlined user interface which is widely used in classroom environments and preliminary research. These advantages make WEKA well-suited for developing new machine learning schemes.

For this research paper the WEKA Time Series Package, part of a new release of dozens of data mining packages available since version 3.7.3, will be used to provide all the analytical processing required to determine the best performing model per food category.

3.1.3 Feature Selection – Correlation-based Feature Selection (CFS)

Feature selection is the process of identifying and removing as much irrelevant and redundant data from a dataset as possible. By reducing the dimensionality of the dataset, the size of the hypothesis space is also reduced, allowing algorithms to operate faster and more effectively; in some cases, the accuracy of forecasts can also be improved, (Hall M. A., 2000).

This research paper presents a relatively new approach to the feature selection in Correlation-based Feature Selection (CFS), which utilizes a correlation based heuristic to determine the merit and rank of each dataset's features. The hypothesis behind the heuristic is: good feature subsets contain features which are highly correlated with the dependent variable, yet uncorrelated with each other. To accomplish this CFS examines many feature subsets, if the dataset has n possible features in a dataset then there can be 2^n possible subsets to be examined and the only way to find the best subset will be to evaluate them all, (Hall M. A., 2000). The CFS method is accessed via the WEKA workbench.

3.1.4 Food Price Report – Canada Food Price Report 2017 Model

The food price report model is based on an identical dataset to that used during the creation of the Food Price Report 2017. The specifics of the data sources are outline in Appendix A. This model applies the Correlation-Based Features Selection (CFS) process which was not incorporated in the original Food Price Report 2017 methodology. The CFS procedure is a statistical method of determining which features are most likely to impact the individual food groups represented by the consumer price index.

The historical records are used to create a general assumption of how these features have impacted food prices in the past and what could be their likely impact in the future. The CFS process identifies features which are highly correlation with each consumer price index food category (the dependent variable) but are not correlated with each other. The historical records pertaining to each identified feature are grouped together and form a specialized dataset for each food category of the consumer price index. These specialized datasets will then serve as the basis for the computer-generated models created by several well-known algorithms. A range of regression algorithms were selected to offer various approaches to forecasting the consumer prices index. Four algorithms were selected from among the WEKA offerings:

- Multivariate Regression: Linear Regression
- Neural Network: Multilayer Perceptron
- Support Vector Machine: SMOreg
- Decision Tree: MP5

3.1.5 Futures – Financial futures-market based Model:

In the case of the Financial futures-market based model, the historical closing prices of several financial commodities are evaluated against each Consumer Price Index food category to identify a specialized dataset for each food group. This process uses the same CFS procedure as outlined in the food price report model above. As with the food price report model, the datasets are modelled by the same well-known algorithms in the WEKA workbench. The process is identical to the perform in the food price report method, with the only difference being the initial financial futures-market dataset used in the process.

3.1.6 The best performing models per food category:

With the Holt-Winters model as the benchmark calculation, the Food Price Report and Financial futures-market based models are compared to it to determine if either produces a lower error. This evaluation is performed by using the Mean Absolute Percentage Error (MAPE) metric, which is outputted by the WEKA workbench. The MAPE is calculate for each food category of the consumer price index for both the Food Price Report model and the Financial futures-market based model.

3.1.7 MAPE – Mean Absolute Percentage Error

The Mean Absolute Percentage Error (MAPE) is a popular measure of performance in forecasting. Two of the major reasons for its popularity are its ease of interpretation and its scale independency which allows it to be used across different datasets. In some forecasting problems, the MAPE may be a good alternative to the Mean Squared Error (MSE) (Lam, Mui, & Yuen, 2001). However, many authors have also argued against the use of the MAPE in certain situations. MAPE measures have the disadvantage of being infinite or undefined if actual values of the dependent variable are equal to zero for any period, and having an extremely skewed distribution when any value is even close to zero (Hyndman & Koehler, 2006). For this research, the disadvantages associated with the MAPE are minimal as the dependent variable, the Consumer Price Index (CPI) per food category has no values at zero. Despite the varying views on the MAPE metric it was selected as the measure of comparison for this research because of its popularity in finance and economics, as evidenced by its availability in the analytical software utilized in this research project (Lam, Mui, & Yuen, 2001).

No. 1 Choice – Best method for forecasting food prices:

The average MAPE error is then calculated for the food categories in each model (Holt-Winters, Food Price Report and Financial futures-market) and the model with the lowest average error is deemed the best method for predicting food prices.

4. RESULTS & CONCLUSION

4.1 Stationarity Tests

The results of the stationarity test (not presented here) indicate that each CPI food category is has an significant level of seasonality and could be analyzed with the Holt-Winters Triple Smoothing algorithm.

4.2 Model Evaluation

CPI Food Category	Holt-Winters	Food Price Report	Financial Futures-Market
Meat	11.6652	3.2889	*2.5615
Seafood	5.2747	6.2303	*2.4149

Dairy	7.5212	3.3426	*0.8778
Bakery	11.0135	3.0025	*1.001
Fruit	7.7517	*0.9187	1.2892
Vegetables	7.2469	*1.0963	2.0273
Other	5.6015	*1.1588	2.3415
Restaurants	8.0033	*0.5031	0.5451
Average MAPE	8.00975	2.44265	*1.63229

Table 3: Mean Absolute Percentage Error (MAPE) results (* - Lowest Error per Category)

As hypothesized, the Financial Futures-Market based model outperformed the Food Price Report model with 1,6% and 2.4% average error rates respectively. Similarly, both models outperformed the Holt-Winters baseline method. Except for the CPI Seafood in the case of the Food Price Report model, both models easily bested the popular baseline model.

The CPI Restaurant category which is regarded as the most difficult to forecast produced the lowest error rate for both the Food Price Report and Financial Futures-Market models. (Joutz F. L., 1997). Across all three models the CPI Other category was produced the least forecasts errors with just 3% while the CPI Meat category was the most difficult to predict with an average error rate of almost 6%

Given the superb performance by both the Food Price Report and Financial Futures-Market models it is likely that even better results can be achieved by combining the datasets to share their uncommon information. However, this is outside of the scope of this research paper but does leave room for further research in this area.

5. ACKNOWLEDGMENTS

Sincerest thanks and appreciation to my supervisory committee, Dr. Vlado Keselj, Dr. Sylvain Charlebois and Dr. Carolyn Waters.

6. REFERENCES

- [1] Billi, R., & Kahn, G. (2008). *What Is the Optimal Rate of Inflation?* Kansas City: Federal Reserve Bank of Kansas City.
- [2] Blomberg, S., & Harris, E. (1995). The Commodity–Consumer Price Connection: Fact or Fable? *Federal Reserve Board of New York Economic Policy Review*, pp. 21 - 38.
- [3] Cao, L., & Tay, F. (2003). Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Transactions on Neural Networks*, 1506 - 1518.
- [4] Charlebois, S., Harris, J., Tyedmers, P., Bailey, M., Keselj, V., Conrad, C., . . . Chamberlain, S. (2016). *Canada Food Price Report 2017*. Halifax: Dalhousie University.
- [5] European Union. (2015). *World food consumption patterns – trends and drivers*. Brussels: European Union.
- [6] European Union. (2015). *World food consumption patterns – trends and drivers*. DG Agriculture and Rural Development.
- [7] Food and Agriculture Organization. (2017). *Food Price Index*. FAO.
- [8] Food Banks Canada. (2016). *HungerCount 2016*. Toronto: Food Banks Canada.
- [9] Hall, M. A. (2000). *Correlation-based Features Selection for discrete and Numeric Class Machine Learning*. Hamilton: University of Waikato.
- [10] Hyndman, R. J., & Athanasopoulos, G. (2013). *Forecasting: principles and practice*. OTexts.
- [11] Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 679–688.
- [12] Joutz, F. L. (1997). Forecasting CPI Food Prices: An Assessment. *Agricultural & Applied Economics Association*, 1681-1685.
- [13] Joutz, F., Trost, R., Hallahan, C., Clauson, A., & Denbaly, M. (2000). *Retail Food Price Forecasting at ERS: The Process, Methodology and Performance from 1984 to 1997*. Washington, DC: The Economic Research Service, U.S. Department of Agriculture.
- [14] Kim, K.-j. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, 307–319.
- [15] Lam, K. F., Mui, H. W., & Yuen, H. K. (2001). A note on minimizing absolute percentage error in combined forecasts. *Computers & Operations Research*, 1141–1147.
- [16] Öner, C. (2012, 03). Inflation: Prices on the Rise. *Finance and Development*.
- [17] Organization of Economic Co-Operation and Development. (2008). *Rising food prices: Cause and consequences*. OECD.
- [18] Statistics Canada. (1996). *Your Guide to the Consumer Price Index*. Ottawa: Statistics Canada.
- [19] Statistics Canada. (2015). *Table 326-0031 - Basket Weights of the Consumer Price Index*. Statistics Canada.
- [20] Statistics Canada. (2016, 03 18). *Canadian Agri-Food and Seafood Exports by Country (by Value)*. Retrieved from Agriculture and Agri-Food Canada:
- [21] Statistics Canada. (2016, 03 18). *Canadian Agri-Food and Seafood Imports by Country (by Value)*. Retrieved from Agriculture and Agri-Food Canada:
- [22] Statistics Canada. (2016a). *Table 326-0009 - Average retail prices for gasoline and fuel oil*. Statistics Canada.
- [23] Statistics Canada. (2016b). *Table 326-0021 - Consumer Price Index*. Statistics Canada.
- [24] Statistics Canada. (2016c). *Table 282-0073 - Labour force survey estimates (LFS)*. Statistics Canada.
- [25] The World Bank. (2015). *World Development Indicators: Structure of merchandise imports*. The World Bank.
- [26] Tciavilca, A., Feuz, D., & McKee, M. (2010). Forecasting Agricultural Commodity Prices Using Multivariate Bayesian Machine Learning Regression. *NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*. St. Louis, MO .
- [27] Wilson, R. (2015). Food Price Report. *Research (University of Guelph)*, 17.

Working, H. (1976). Futures Trading and Hedging. In *The Economics of Futures Trading* (pp. 68-82). Palgrave Macmillan UK.

