

Time Series Model to Forecast the Pricing of Dairy Products

Samir Rana

Asst. Professor, Department of Comp. Sc. & Info. Tech., Graphic Era Hill University,
Dehradun, Uttarakhand India 248002

Article Info

Page Number: 1670 - 1677

Publication Issue:

Vol 70 No. 2 (2021)

Abstract

Seasonal patterns and patterns of food production are linked with each other, which contributes to have a significant impact on the economy of a country. Seasons and food production patterns are linked with each other. Because of this, it is of the utmost importance to establish projections about the production of food patterns that are sensitive to variations in the climate, which will ultimately result in satisfied customers and successful production. Because of this, it is vital to establish trustworthy techniques of dairy production forecasting in order to prevent a shortage of dairy production in an industry. As a result, the purpose of this study is to construct a model that is capable of reliably predicting the price of dairy products in relation to changes in the climate. In the course of the research for the thesis, Ireland will serve as the case study, and a mix of time series and machine learning models will be utilized in order to make price predictions. A time series of data relating to dairy production from the year 1990 to 2019 is extracted and utilized as a data source for dairy products. This data source will be used until the year 2019. The work that was actually done for the proposed paper makes use of the entire dataset for training purposes and makes use of the pricing list for the previous year for the testing phase. These are then utilized as variables in order to assess product yields as well as product losses. The implementation, on the other hand, is based on the principles of time series as well as four machine learning algorithms, specifically ARIMA, ARIMA Garch, SEM, and SARIMA. In compared to the results of other models, it was shown that the SARIMA model produced superior results. In addition, the findings were computed on the basis of an assessment matrix that took into account the root mean square error.

Article History

Article Received: 05 September 2021

Revised: 09 October 2021

Accepted: 22 November 2021

Publication: 26 December 2021

Keywords: dairy prediction, machine learning, ARIMA, Garch, rmse

Introduction

Over the course of many years, researchers have noted that both the agricultural industry and the level of civilization achieved by mankind have been steadily advancing over time. The agriculture industry is sometimes referred to as the "backbone" of an economy because of its central role in the production of goods. According to the data, India is the country that produces the most dairy products, and the country supplies its goods to 37 percent of the world's population. It is followed by the United States of America and China [1]. However, it is important to note that dairy products have evolved into a commodity that must be stocked in every home. Since it was once considered a luxurious good that could only be discovered among the most affluent people, its consumption has

constantly expanded, and it is today one of the food products that is consumed the most frequently. Dairy products include a wide variety of proteins derived from milk, such as milk, cheese, butter, raw milk, milk powders, cheddar, and other varieties. These products have attracted the attention of government organizations, which may implement regulations in a variety of contexts in order to maximize the financial advantage. On the other side, their makers have also been consistently increasing the prices on watching the consumption by end customers. This is the case regardless of whether or not the price increase is justified. It is anticipated that the consumption of dairy products will rise by 10 percent, which will result in an increase in business of 283,000 crore for industrial producers [2]. According to the findings of another study, the world's population might reach 9 billion by the year 2050, which would result in widespread starvation and uncertainty over the availability of food [3]. This type of data analysis, which is conducted and surveyed, takes into consideration the production of dairy products, which in turn enables businesses to comprehend the pattern in the cyclical demand for dairy goods. This results in a rise in revenue for a country as a result of an increase in the price of things that are consumed. Therefore, it has been observed by forecast experts that consumption, the elements of demand and supply, and the rise in revenue all contribute to a substantial share of the economy. This is something that has been observed [4]. As a consequence of these factors, there is currently a growing number of threats to the food security that is taking place all over the world. It is to be anticipated that the processes and models used in the manufacture of food will undergo transformations in the not-too-distant future. This is due to the fact that the manufacturing of dairy supply is dependent on the scale of production and demand. In addition, policymakers can utilize this data on demand and supply through manufacturers to compile their yearly report on food and set prices for it. This data can be used in both of these ways. As was indicated before, an increase in the consumption of dairy products leads to an increase in the margin profit that producers make, which in turn causes government organizations to change their policies, which ultimately leads to an increase in the price of the product. In a situation like this, the local people are the ones who bear the brunt of such price changes, which can also occasionally put them in a position where they are poor. Also, knowing that milk is a fundamental requirement for each household, the cost of purchasing it tends to escalate to an extremely high level for them. These contributors also face a seasonal tendency in which their prices go up because of their increasing consumption and might go down because of decreased demand as a result of climatic changes. This trend occurs every year. These kinds of pricing swings have an impact on the overall operating costs and financial expenditures incurred by producers, manufacturers, and customers. People who are concerned about the price of dairy products are forced to buy the product at a rate that is twice as expensive when the price tends to go up. Sometimes, as a result of this, they end up purchasing potentially harmful but more affordable food products at a much-reduced price.

The "affecting nature" of price increases on consumers is identified as the primary issue that arises from the observations made in the thesis. They have a tendency, as prices rise, to replace the food products they buy with normal, low-cost products, which may not only put their health at risk but also bring the economy to a standstill. As a result, software developers have come up with two solutions to a problem of this nature: the first answer includes having a significant amount of dairy production available, and the second solution involves boosting agricultural production in countries other than the one in which the software developer resides.

Related Works

In the study described in [5], the authors present an architecture that, when given raw data consisting of multivariate labels as input to the system, is capable of making accurate predictions regarding food production. This model was able to extract events and clusters were built based on the time series. In addition to this, it merged the extracted events and clusters in order to generate calculated features, which were then added to the training dataset so that it could be used for predicting. The ARIMA model and the KNN method were utilized in the author's implementation of this work. All of the inputs that were provided to the model were computed based on the time series intervals.

A times series-based model was developed in the work that was proposed by the authors in [6]. This model was used to anticipate livestock production. His work was carried out with the assistance of the machine learning technique known as AdaBoost. AdaBoost's primary function is to build a single robust classifier that is capable of giving results in a linear form by combining and generating predictions from numerous existing classifiers. Because time series were utilized, an ARIMA model that was given a span of historical input values, which resulted in substantially improved outcomes, was applied. The same authors implemented the work using SVM as a classifier, and they observed that the model delivered superior efficiency when compared to the AdaBoost technique.

Deep learning models like ANN and CNN were utilized in the authors' other piece of research, which is referred to as [7], in order to carry out an application of forecasting. This model has the potential to forecast an increase in the cost of milk production due to natural variables and alterations in the climate. This model's inputs were historical occurrences in which the production of milk was influenced by the conditions of the environment at the time. The input consisted of a time series spanning the last five years. The aforementioned data was then put to use for the purposes of testing and training. The results of the investigation led the researchers to the conclusion that the ANN model performed significantly better than the CNN model.

Implementaion Details

The development of a model that is able to accurately predict changes in the cost of dairy products (butter) in Ireland is the primary objective of the research project that is being carried out. The principles of time series, in conjunction with machine learning, are what we propose using as an approach in order to successfully execute the model. The first step of the implementation process is carried out by means of the ARIMA model. During this phase, the data is retrieved and collected from repositories, and it then goes through the phases of training and testing.

A Dataset Used

The data was obtained from the repository that is maintained by the European Union. This dataset, on the other hand, includes the prices of various dairy goods such milk, raw milk, milk powders, butter, cheese, and other similar items. The price changes of Irish dairy products are going to be analysed, which will then lead to a price prediction for the next three years given that the study paper focuses on Irish items. However, in order to meet the requirements of the implementation, the model must be trained using the complete dataset. The primary goal of making use of this data is to determine whether or not there is a correlation between the changing of the seasons in Ireland and

shifts in the cost of dairy products. The purpose of the research paper is depicted in the following diagram:

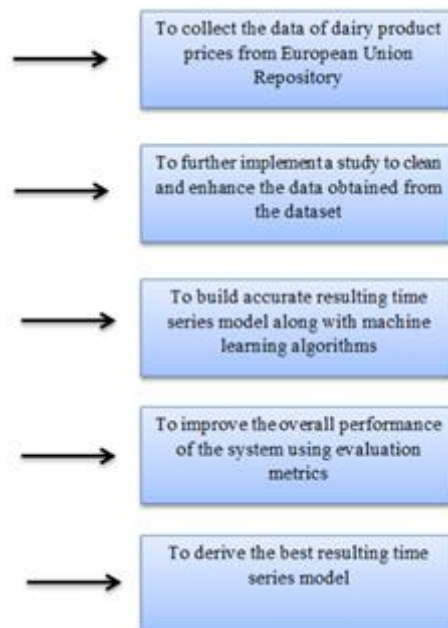


Figure 1: Workflow of the proposed system

B Data Pre-Processing

The process of cleaning up data typically involves removing records from the data that are irrelevant and then filtering the remaining records in such a way that the removal of these records does not impact the operation of the model as a whole. In a model for time series data, the process of cleaning the data occurs for both univariate and multivariate data at the same time. Some characteristics, including the nation and the sector code, are consistent across the board in the dataset pertaining to Irish dairy products, including those that fall under the category of food production. Therefore, such duplicated data must also need to be cleaned up so that it does not have an impact on the model's final predicting. This can be accomplished by removing duplicates.

C Feature Engineering

The concept of learning from past data and then developing and adding fresh data in order to enhance the dataset is at the centre of feature engineering. The expanded dataset is eventually put to use for the training purpose after it has been further processed. Because of this, it helps to improve machine learning algorithms because the features of these algorithms are now dependent on the domain knowledge of historical data. As a result, the notion of feature engineering can be utilised by the proposed thesis, and then expanded upon, in order to generate data based on climatic conditions and temperature.

D Proposed Workflow

The development of a model that is able to detect and forecast changes in the price of dairy products produced in Ireland is the primary objective of this thesis. In order to accomplish this goal,

a concept of time series was developed with CV, which could be used to the dataset that was collected in this manner. The data collection, which was obtained from the Kaggle repository, contains a substantial quantity of information. As a result, these data required some preliminary processing and filtering before being sent into the training and testing phase of the project. Processing the data turned out to be a significant contribution to the whole operation because removing extraneous data led to an increase in the model's degree of precision after being processed. Because of this, the data was taken from the dataset that had been obtained and supplied into the pre-processing model. The data set consisted of hourly pricing changes that were observed in Ireland throughout the time period in question. This information was gathered and inputted with consideration given to a certain period of time and on the basis of historical occurrences that took place in the past. When this data was finally handed on to the processing phase, the data processor was responsible for removing and filtering any data that was deemed to be irrelevant before sending it on to the building model. In this section of the programme is where all of the work that has to be done will be carried out. An accurate procedure was trained and tested utilising techniques that involved cross validation, and the data was split up during those processes. After the data was divided into a ratio of 80:20, it was decided that the training phase would receive 80% of the data, while the testing phase would receive 20%. During this phase of the model's construction, a correct execution of the data took place. In the end, the data that had been learned were input into the forecasting module. During this stage, the machine learning algorithms were put to use so that predictions could be made. All of the ML forecasting algorithms were implemented by following their standard procedures throughout the execution phase. These machine learning algorithms also incorporate the ideas of time series and CV into their processes. The architecture of the work that is being suggested is shown in the diagram below.

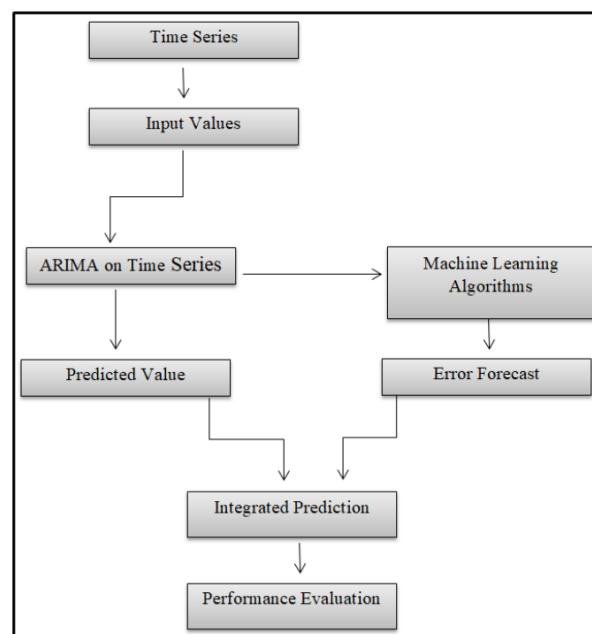


Figure 2: Architecture of the proposed system

In order to carry out the implementation, the dataset was acquired from the repository run by the European Union. This storage facility held a wide range of dairy products in its inventory. The time

period covered by the dataset was from 1990 to 2019, and it includes monthly prices for several dairy products. As a result of the fact that the case study for this thesis is carried out on the dataset of Ireland, additional price forecasting is carried out on Irish dairy goods. The time series of events were provided as input to the model so that it could carry out univariate forecasting after the model was built using the time series idea. This allowed the model to perform forecasting with a single variable. For the training set, the whole dataset was used, while for the testing set, just the data coming from the last columns were used.

A model goes through the phases of training and testing. During the training phase, a dataset is received from a repository, and the machine is taught using that dataset initially. In most cases, the testing phase occurs during the later stages of development, during which the accuracy of the model is examined. On the other hand, this data is divided into training and testing categories based on its temporal nature. The model employs distinct records that are already existing in the dataset whenever there is a split that takes place. This dataset might be a sequence of movies or still images that are subsequently segmented further during the processing stage. Given that the thesis makes use of a time stamp model, the data that is given in as input is historical data, which includes values from the past.

On a time axis, the model is given these historical values to feed off of. As a result, the procedure is carried out using a model called a time stamp. However, this data splitting happens in a specific order and follows an 80:20 ratio, which means that 80 percent of the data is utilised for training the model, while the remaining 20 percent of the data is used to test the model. After the model has been put through its paces in the testing phase, the evaluation phase continues.

Results

Training and testing were the two stages that made up the execution of the model. In order to train the time series model, the full dataset was utilised and provided as an input into it. After that, we applied some filters to the dataset, and then we used the final columns for testing. During the testing phase, the prices of the previous twelve months' worth of butter were utilised, and lastly, forecasting was performed for the subsequent three years.

Table 1: Evaluation Metrics of ARIMA Model

EVALUATION METRICS -1 : ARIMA MODEL	
Root Mean Square Error	19.60
Mean Squared Error	384.43

Testing on last 12 months in dataset using ARIMA

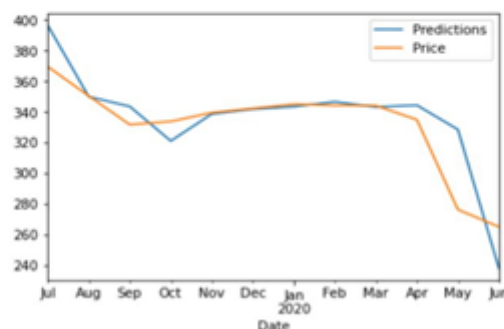


Figure 3: A graph displaying predictions of butter prices based on the ARIMA model, with the months listed along the x-axis and the price of butter along the y-axis.

The actual price values are shown in orange on the graph that was constructed above, and the values that were predicted by the model are shown in blue on the graph.

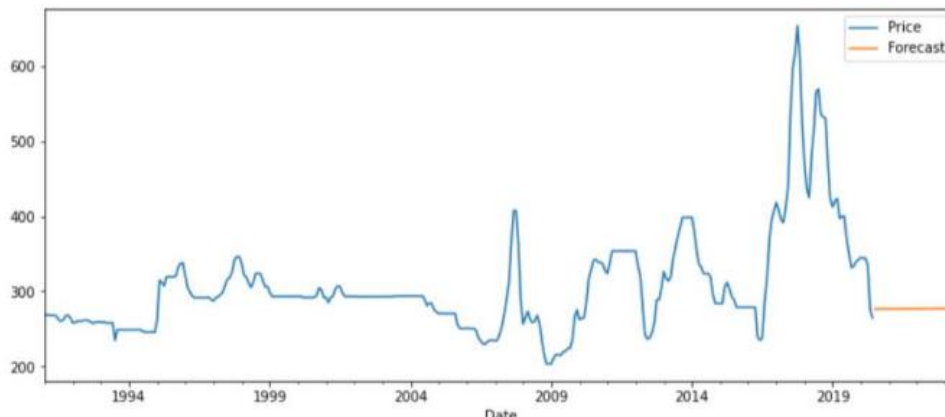


Figure 4: A graph illustrating the ARIMA model.

The price of butter is depicted in the blue line of the graph above, which extends from 1994 to 2019, while the orange line projects the price for the subsequent three years.

Conclusions

The implementation of a time series model for the purpose of forecasting the prices of Irish dairy products was the primary purpose of the thesis. The ARIMA model and other machine learning models were utilised in the execution of the thesis. Obtaining the dataset from the European repository and then carrying out the model's training and testing phases using that dataset comprised the entirety of the workflow for the model. During the training phase, the entire dataset was used, however for the testing phase, just the data from the preceding months were utilised. Following this, a time series idea was utilised, and the model was provided with the price data of butter for the preceding year's worth of time as an input.

References

- [1] Wunderlich, S.M.; Martinez, N.M. Conserving natural resources through food loss reduction: Production and consumption stages of the food supply chain. *Int. Soil Water Conserv. Res.* 2018, 6, 331–339
- [2] Beck N, Katz JN (2011) modeling dynamics in time-series–cross-section political economy data. *Ann Rev Polit Sci* 14:331–352
- [3] Box G, Jenkins E (2015) *Time series analysis: forecasting and control*. Wiley, Hoboken
- [4] Singh, M., Singh, R., & Shinde, V. (2011). Application of software packages for monthly stream flow forecasting of Kangsabati River in India. *International Journal of Computer Applications*, 20(3), 7-14
- [5] Jacobs JA, Siegford J (2012) Invited review: the impact of automatic milking systems on dairy cow management, behavior, health, and welfare. *J Dairy Sci* 95:2227–2247
- [6] Analysis and Study on the Classifier Based Data Mining Methods SN Popat, YP Singh *Journal of Advances in Science and Technology| Science & Technology* 14 (2)

- [7] Efficient Research on the Relationship Standard Mining Calculations in Data Mining SN Popat, YP Singh Journal of Advances in Science and Technology| Science & Technology 14 (2)
- [8] Pardhi, Shruti, and K. H. Wanjale. "Survey on techniques involved in image segmentation." Int. J. Comput. Sci. Trends Technol.(IJCST) 4.3 (2016).
- [9] Sayed, Asim, M. Sardeshmukh, and Suresh Limkar. "Improved Iris Recognition Using Eigen Values for Feature Extraction for Off Gaze Images." ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India-Vol II: Hosted by CSI Vishakapatnam Chapter. Springer International Publishing, 2014.
- [10] Shivadekar, Samit, S. R. Abraham, and S. Khalid. "Document validation and verification system." Int. J. Adv. Res. Comput. Eng. Technol.(IJARCET) 5.3 (2016).