Comparative analysis of Time Series Models to predict the seasonal prices of Ireland Dairy Products

MSc Research Project
Data Analytics

Shital Namdeo Raut

Student ID: x19243294

School of Computing
National College of Ireland

Supervisor: Dr. Paul Stynes, Dr. Pramod Pathak

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Shital Namdeo Raut
Student ID:	X19243294
Programme:	Data Analytics **Year:** 2021
Module:	MSc Research Project
Lecturer: **Submission Due Date:**	31/01/2022
Project Title:	Comparative Analysis of Time Series Models to predict the seasonal prices of Ireland Dairy Products.
Word Count:	5634 **Page Count:** 19

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Shital Namdeo Raut

Date: 31st January 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Comparative analysis of Time Series Models to predict the seasonal prices of Ireland Dairy Products

Shital Namdeo Raut

x19243294

Abstract

Seasonal patterns and food production patterns are however linked with each other and contribute to have a major impact on the economy of a country. This makes it an urgent need to forecast the production of food patterns that are reliant to climatic changes; further leading to customer satisfaction and industrial deliverability. Therefore, it becomes necessary to develop reliable forecasting methods to prevent shortage of dairy production in an industry. Hence, this thesis focuses to build a model that could accurately predict the price of dairy products in proportion to the climatic changes. Ireland is selected as the case study in the thesis and a combination of time series and machine learning models are adopted to forecast the price. A time series data related to dairy production from the year 1990 to 2021 is taken and is used as a data source for dairy products. The implemented work in the proposed thesis utilizes full dataset for training purpose and uses the price list of last 12 months for testing phase. These are further used as variables to evaluate product yields and losses. However, the implementation occurs on the fundamentals of time series and four ML algorithms, namely: ARIMA, ARIMA Garch, SEM and SARIMA. It was observed that the SARIMA model performed better in comparison to other models. Further, results were calculated based on evaluation matrix including root mean square error.

1 Introduction

1.1 Background

For years, it has been observed that the agricultural sector and the civilization of humans have been on evolving at an inclining rate. The agricultural sector is also known to be as the backbone of a country's economy. As per statistics, India happens to be the largest dairy producer, providing its products to 37 percent of world's population, followed by USA and China *(Dairy Production).* It's nonetheless to mention that dairy product has now become a necessity in every household. Its consumption has steadily increased from being observed as a luxurious product found only amongst the elite people to now – being one of the most commonly used food products. Dairy product comprises a wide variety of milk based proteins such as milk, cheese, butter, raw milk, milk powders, cheddar etc. that have grabbed the attention of government organisations who might imply policies in varying situations to increase the monetary benefit. On the other hand, their manufacturers have also been constantly elevating the prices on witnessing the consumption by end users. According to statistics the consumption of dairy products is believed to increase by 10 percent, generating 283,000 crore of business amongst industrial manufacturers *(Fao.org).* In another report, it has been estimated that total population in the world will have surpassed 9 billion by 2050.might exceed to 9 billion by 2050 that shall lead to global hunger and food insecurity *(Wunderlich).*

Figure below illustrates world's dairy production by different countries; wherein European countries are the second largest manufacturers of dairy products with Africa being the last.

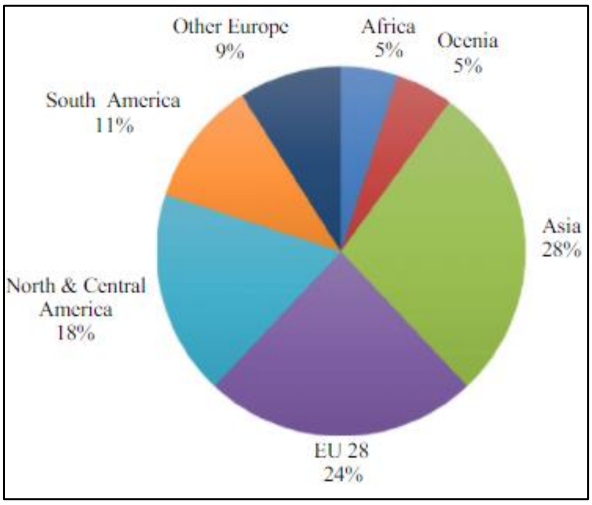

Figure 1: Global Dairy Production

Such conducted and surveyed analysis of data comprehends the production of dairy items which in turn allows the manufacturers to understand the trend in seasonal demand of dairy products. This leads to increase in revenue of a country by increasing the price on consumed products. Hence, it has been witnessed by forecast analysts that the consumption, the demand and supply factors and the increase in revenue; contributes towards a significant portion of the economy. According to analytical surveys so conducted it is expected that the dairy production segment shall grow due to the increase in the usage of technological applications. Hence, this results into providing a way for manufacturers to increase their margin and thereby increase the prices of products (*Soomro MF.2006)*.

Where there has been an increase in prices based on the demand and supply chain of customers, factors such as climatic changes and natural hazards also contribute to increase in prices. These further results into an increasing amount of threat with regards to food security that takes place worldwide *(Rahman, S.2010)*. Since the manufacturing of dairy supply depends on the production and demand scale, it is expected that the food production methods and models might change in the near future. Further, this data of demand and supply through manufacturers can also be used by policy makers to generate their food annual report and fix their prices.

1.2 Problem Statement

As mentioned earlier, that due to the rise in consumption of dairy products, the manufacturers increase their margin profit and government organisations alter their policies; which in turn leads to a rise in product prices. In such a scenario, it is the local people who suffer from such price fluctuations, also sometimes leading to poverty amongst them. Also considering milk to be the basic necessity of any household, tends to become very expensive for them to buy. Since the thesis has taken Ireland to be as the case study, the agri-food industry serves its purpose to 7 percent of the country's economy, with its main contributors in dairy products to be cereals, dairy and cattle. These contributors also experience a seasonal trend wherein its prices are at hike due to its increased consumption and might fall due to less demand based on climatic changes. Such price fluctuation affects the overall working and monetary expenses of producers, manufacturers and customers. When the price of dairy products tends to incline, its concerned people are forced to buy the product at twice the rate. At times this also leads them to buy hazardous and cheaper food products at a much lower rate.

The main problem observed in the thesis is the "affecting nature" of price increase on customers. As price increases, they tend to alter the food products with regular cheap products that might not only put their health at risk but also lowers down the economy. Hence, software developers evolved with two solutions to such a problem: the first solution involves having a considerable amount of dairy production at hand and the second consists of promoting agricultural production outside the country.

1.3 Motivation

As the problem stated above, it becomes necessary to design a model that could forecast the increase/fall in prices with regards to dairy products. Such a model shall also assist the customers and warn them to either stock their products in terms of increased prices or hold their consumption of dairy products, so that it might not affect them monetarily. This motivates us to create and develop a forecasting model that could serve the purpose of not falling into the trap of hiked prices.

Hence determining seasonality plays a significant role in stabilising the agricultural, food production and dairy production operations in a country. This motivates us to incline the focus of thesis in providing seasonal prices and later forecasting the future price that could encourage customers to buy according to climatic changes; thereby enhancing customer patterns and controlling the supply chain. Further, it is noteworthy to mention that the induced seasonality prevails throughout the year resulting in fluctuation towards customer preferences; which in turn affects the production price.

Therefore, the proposed thesis aims to implement a price forecasting model for dairy products using fundamentals of time series including the ARIMA model.

1.4 Organisation of Thesis

A detailed description on dairy production and its impact factors along with consequences on customers has been presented in chapter 1. A survey of research authors to forecast the same has been described in chapter 2 along with time series analysis. Chapter 3 briefs about the research methodologies that are used to implement the proposed work. The workflow of the proposed system along with data mining and feature engineering process is summarised in chapter 4 followed by implementation details in chapter 5. The observed analysis from the experiments so conducted in presented in chapter 6 proceeded by conclusion and references.

2 Related Work

The basic concept of time series is to forecast an application by giving a series of historical input values to the model. This model is responsible to record a series of events on regular intervals of time. Further, depending on the frequency, a time series can forecast yearly (annual budget), monthly (air traffic) and weekly (sales) data. Apart from forecasting in agricultural and food sectors, this concept can also be applied to businesses. The fundamental of time series are applied to the ARIMA model that works on the idea of providing past values of information to the time series. Further, these values assist analysts and developers to predict and forecast future values. In the proposed thesis, a seasonal model of ARIMA as – SARIMA is put forward to forecast dairy production prices in Ireland. A summarised detail of related works by other authors have been presented.

2.1 The Time Series Classification Approach

(M. W. Kadous) author describes an architecture in his work that could predict food production by giving raw data of multivariate labels as the input to the system. This model was capable of extracting events and formed clusters based on time series. Further, it combined the extracted events and clusters to generate computed features that could be further given to the training dataset for forecasting. The author implemented this work using ARIMA model and KNN algorithm. All the inputs given to the model were computed over intervals of time series.

In a work proposed by *(J. J. R. Guez)*, the author implemented a times series based model to forecast livestock production. His work was implemented using AdaBoost as machine learning algorithm. AdaBoost is primarily responsible to combine multiple classifiers and generate one strong classifier that could give the results in a linear fashion. Since time series was used, a span of historical input values was fed to the ARIMA model that further enhanced the results. The same authors implemented the work with SVM as a classifier and witnessed that the model gave higher efficiency in comparison to AdaBoost algorithm.

Another approach *(S. S. Skiena)* implemented a similar work using three machine learning algorithms to forecast stock prices. The authors used the SARIMA model to forecast the price hikes in stock market. Further classification was implemented using four machine learning techniques namely: gradient boosting , logistic regression and random forest, decision trees. Further the authors performed a comparative analysis of all the four models. it was however, observed that gradient boosting outperformed the rest of the models and provided better efficiency.

In another research by *(J. Schmidhuber)*, a forecasting implementation was carried out using deep learning models such as ANN and CNN. This model could predict the price hike of milk productions depending on natural factors and climatic changes. The inputs fed to this model were past events in which the production of milk was impacted due to weather conditions. A time series of five years was given as the input. This data was further used for testing and training purpose. The output of the work concluded that the ANN model performed better in comparison to the CNN model.

In another survey by *(L. Ye and E. Keogh)*, the authors implemented bitcoin forecasting model using LSTM as the base algorithm. A time series of ten years was fed as the input and ARIMA concepts were implemented on it. The model performed efficiently and a significant amount of error loss was observed.

Historical data has also been used to keep a track on agricultural stocks so produced, in order to forecast its price hike in future. The sales impacted the associated risk and minimized the price fluctuations for individuals. The authors *(Z. Xing, J. Pei)* of this model obtained the dataset from Kaggle repository and performed machine learning techniques such as logistic regression and SVM. The time series of past events was fed as the input and the data was further classified for training and testing purpose. This model assisted the consumers to make a smart choice and purchase their agricultural stocks accordingly.

The ARIMA model was also used to predict the price of tea production in various countries. The input so obtained was considered to be as the raw data of tea production. The experimentation was carried out on univariate model using SARIMA. This model was further, combined with the concepts of ML algorithms such as decision trees and logistic regression. The authors *(S. Y. Philip)* conducted the survey and evaluated results using mean square error.

(Guo et al) presented one of most fascinating real-world uses of time-series models for streaming web data. (2016). To anticipate streamed time series containing irregularities and transition points, RNN and adaptive gradient learning algorithms are utilized. This analysis was carried out to investigate the local characteristics of the time series in order to automatically weight the loss with the distributional qualities of the data that is currently available. Intensive research investigation was undertaken on both created and actual samples to examine the suggested technique's ultimate performance.

Various research has been undertaken to see how weather affects dairy pricing. *(Kurumatani et al. 2020)*, for example, utilised twin recurrent neural network (RNN)-based forecasting approaches, notably time-alignment of time point forecast (TATP) and direct future time series prediction (DFTS). As the fundamental forecaster, both approaches use a different type of RNN, RNN or LSTM, for example.

When the results of various models were examined, it was evident that now the LSTM model outperformed the others in regards of prediction performance. GRU and LSTM are outperformed by Simple RNN, which uses less epochs. LSTM, on the other hand, surpasses SRNN's expanded linear approaches given adequate training. Figure 2 depicts the whole forecasting paradigm employed in this research.

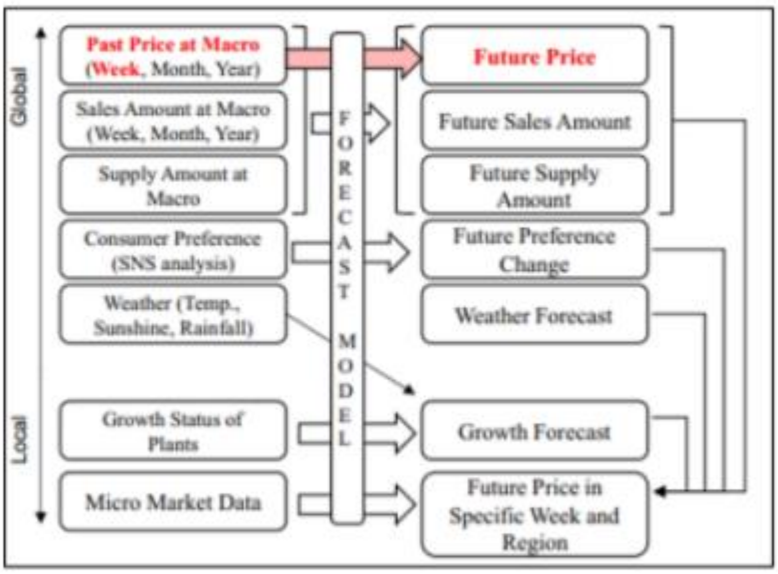

Figure 2: Seasonal price forecasting framework (adapted from (Kurumatani, 2020))

2.2 Time Series in Dairy Production

Several surveys have been conducted to analyse existing dairy production patterns in Ireland. The researchers of a few of the research analysed data from nine Irish milk processing units between 2007 and 2011 to assess somatic cell counts (SCC), laboratories pasteurisation counts (LPC), and bacterial cell counts (BCC) (BCL). SCC climbed considerably in October and declined in November, according to the authors, who noted seasonal tendencies. The influence of different seasons on milk intake was proven in the previous study. The presence of pathogens has an impact on the quality of the product. *(O'Connell et al., 2015)*.

Another research was planned to look at dairy sector in Pakistan during 1990 and 2000 employing time series dataset from the ARIMA model. MINITAB software was utilized by the researchers for time series data. Due to a major considerable rise at lag 1, and it was bigger than just the 95 percent confidence boundaries, it demonstrated that somehow the series was non-stationary *(Ahmed et al., 2011).* Model estimation, diagnostic testing, and forecasting were all part of the overall analysis.

In the Eastern United States, a latest survey report issued by Bolotova (2017) explores the pure dairy chain trends in a number of American dairy farms. The author looked at the periodic pricing trends over whole milk and agriculture to sale prices in major cities in the impacted areas. The whole milk pricing data was gathered using the USFDA Agriculture and Market studies. The trend of farm milk pricing, wholesale milk prices, and associated earnings varies throughout two eras, as per the experts' findings: during regulatory intervention *(Bolotova, 2017).* Besides from either the studies described above, several more investigators have used Time series technique and ARIMA models to anticipate monsoonal dairy price trends. Inside this wake of the recent of Italy's milk quota, there is a considerable body of work on the instability of dairy product cost predictions. The key topic of this paper is whether expiry of milk restrictions would influence economic stability and agent behaviour at various market levels. Two prior studies looked at several facets of the Italy dairy business, including market inequities and their consequences for price movements, and the presence of oligopolistic competitors and its consequences for pricing structure. The present study uses per week dairy prices of goods using time series analysis and looks at economic patterns in pre / post regimes. Using previous milk product's price well before quota implementation in 2015, one of its key purposes of this study is to investigate the political effect on market price volatility. The researcher estimates the fate of milk products more than a five-year period. Later the limit release in an uncontrolled market situation. Excluding a few sectors in which there was significant market meddling, volatility rose in practically all regions *(Rosa et al., 2016).*

To anticipate the pricing of its most widely consumed dairy items, the author applied five linear and nonlinear modelling techniques. Integrating linear and nonlinear models was effective in projecting dairy prices of goods, as per this research. From 2006 to 2007, prices reached a strategic low point, according to this research. The AR model looked as being the most challenging to use for projecting cheese prices; yet, after that year, the AR model worked admirably. When tested by evaluation parameters such as RMSE and MAPE, the standard AR model, on either hand, successfully identified butter price.
However, one important finding is that multiple models can generate outcomes that are almost identical. Although MAPE and RMSE might suggest that certain frameworks performed much better than others, the Diebold-Mariano (DM) test was shown to be beneficial in determining only one of the model variants is important. The idea that the findings of DM tests are really only relevant for first 6 months is undeniably a drawback. Moreover, 10 to 12 months in ahead, the cost of Milk Powder (SMP) is quite easy to estimate. More than a 6- to 8-month period, error terms are reasonable for other items *(Hansen, 2020).*

3 Methodology

The primary aim of the thesis is to develop a model that is capable to forecast the price fluctuations of dairy products (butter) in Ireland. In order to successfully implement the model, we propose to use the concepts of time series along with machine learning. The initial part of implementation takes place through the ARIMA model wherein the data is fetched and gathered from repositories and undergoes the training and testing phase. All the inputs being provided to the model contains times series events of price fluctuations of butter from the year 1990 to 2021. Further, the model undergoes through machine learning algorithms and forecasts the price hikes/falls of the coming three years.

3.1 Data Collection

The data has been collected from the European Union repository. This dataset however, comprises of price of dairy products ranging from milk, raw milk, milk powders, butter, cheese etc. Since the thesis is based on Irish products, and the price fluctuations of Irish dairy products shall be evaluated and, further result into forecasting its price for the next three years. However for implementation purpose the entire dataset shall be used to train the model. The fundamental objective to use this data is to establish a link between seasonal changes of Ireland and price fluctuations in dairy products. Below diagram illustrates the goal of the thesis:

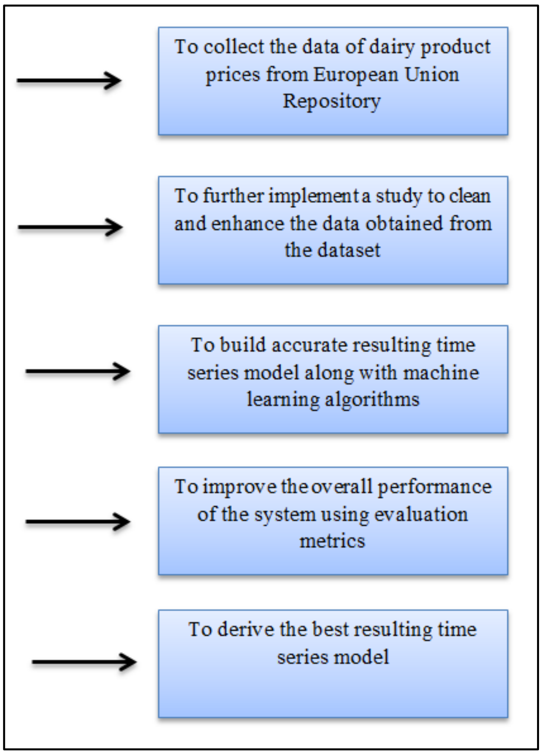

Figure 3: Primary goals of the proposed work

3.2 Data Cleaning

The process of cleaning a data generally involves identifying irrelevant records in a data and further filtering them in such a way that it does not affects the overall working of a model. The process of data cleaning in a time series model takes place for univariate and multi variate data. In the dataset of Irish dairy products some variables such as country and sector code remains the same for all the products listed under food production. Hence, such a redundant data must also need to be cleaned so that it does not have an impact on the final forecasting of the model.

3.3 Feature Engineering

Feature engineering revolves around the concept of learning from historical data and further creating and adding more of data so as to increase the dataset. This increased dataset is then later used for the training purpose. Hence, it helps to enhance the ML algorithms as its features are now based on the domain knowledge of past data. Therefore the proposed thesis can use the concept of feature engineering and further expand, to generate data based on climatic conditions and temperature.

3.4 Data Mining

The process of data mining involves extracting large amounts of data from the past in order to predict future trends and patterns. This results into the generation of more output. Further, to build a model with higher accuracy, some parts of the previous data needs to be arranged sequentially. This process is accomplished through cross validation technique that distributes the data into training and testing phase. Below figure depicts the workflow of the proposed study:

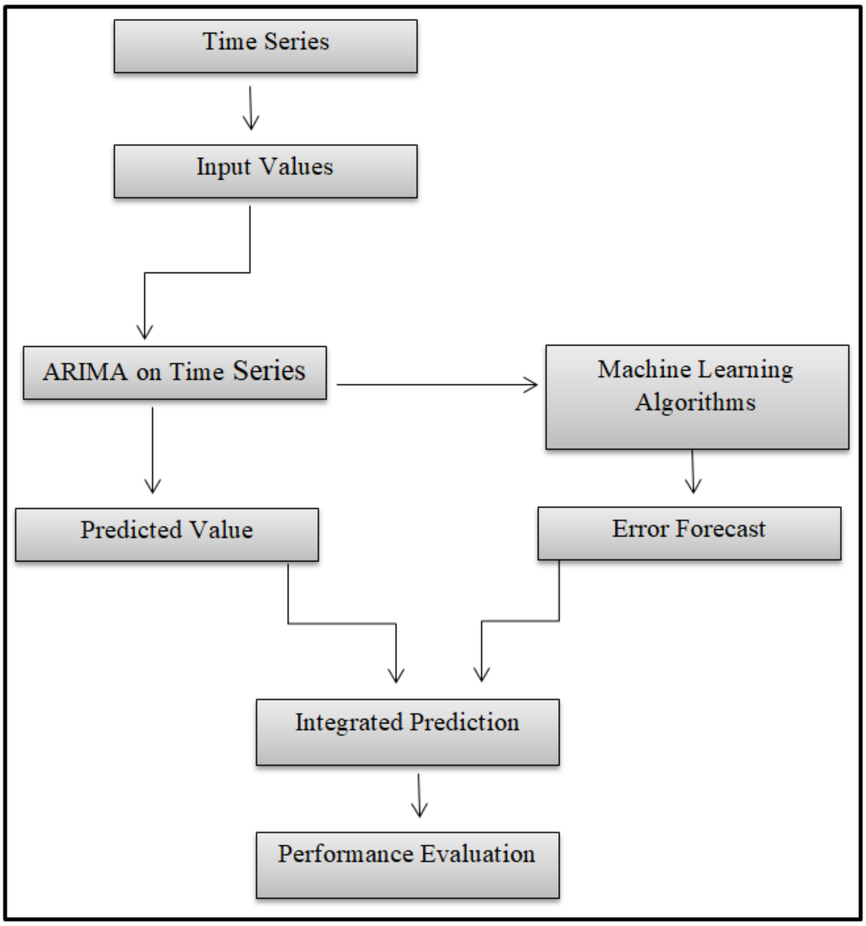

Figure 4: Workflow of the proposed work

4 Design Specification

The primary aim of the thesis is to create a model that would be capable to detect and forecast the price of dairy production in Ireland. To achieve this, a concept of time series was introduced along with cross validation (CV) that could be implemented on the dataset so obtained. The dataset was acquired from European Union repository and contained large amount of data. Hence, this data needed to be filtered and pre-processed before feeding this data to the training and testing phase. Data processing proved to be a major contribution in the entire working as, filtering out unnecessary data results in achieving higher accuracy of the model. Hence, the data was fed to the pro-processing model from the dataset so collected. The dataset was in the form of hourly price fluctuations that took place in Ireland. This data was acquired and fed with respect to a span of time and on the basis of historical events that took place in the past. Once this data was fed to the processing phase, next the data processor, removed and filtered all the irrelevant data present and had it passed to the building model. It is in this module that the actual execution of the work takes place. An accurate procedure training and testing was carried out with splitting the data using cross validation techniques. Once the data was split into a ratio of 80:20 was decided wherein; 80 percent of the data was given to the training phase and 20 percent to the testing phase. A proper execution of data took place in this model building phase. Later, the trained data was fed to the forecasting module. It was in this phase that the machine learning algorithms were used for prediction. All the ML forecasting algorithms were implemented using their normal executional phase. The concept of time series and CV are accompanied by these machine learning algorithms.

The diagram below depicts the architecture of the proposed work.

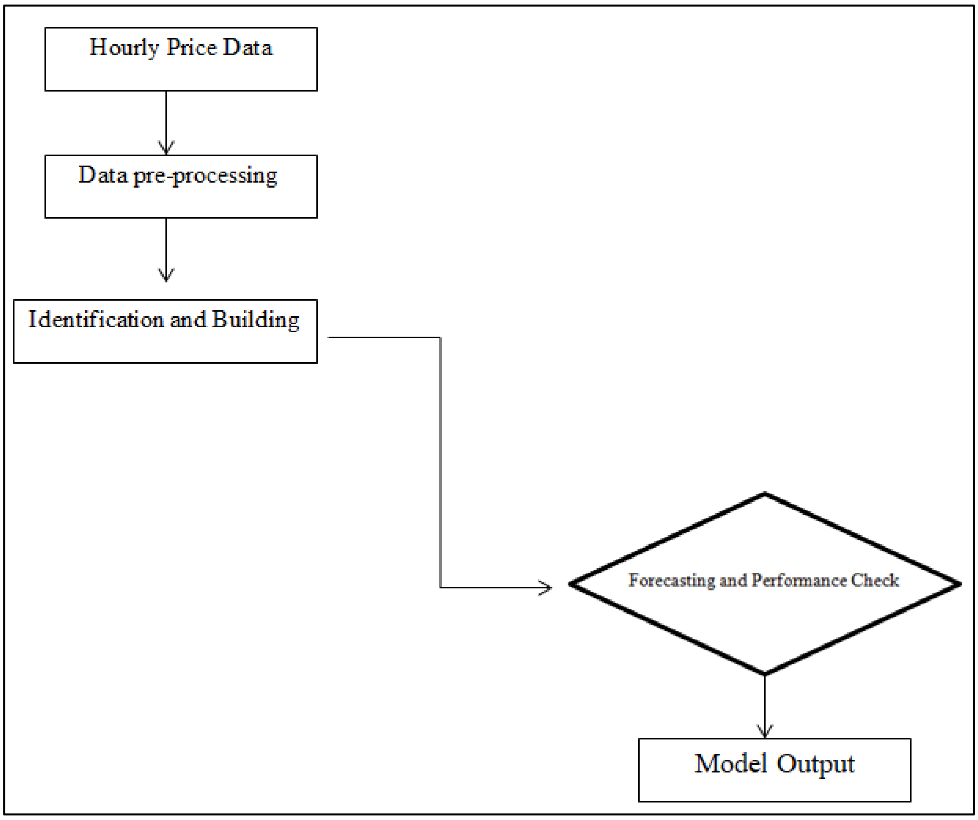

Figure 5: Working Architecture of the Proposed System

5 Implementation Details

The fundamental aim of the thesis is to implement a time series model so that it could forecast the price of dairy products using machine learning techniques such as ARIMA, ARIMA GARCH, SEM and SARIMA. To carry out the implementation the dataset was obtained from European Union repository. This repository consisted of a wide variety of dairy products. The dataset contained monthly price of dairy products from 1990 to 2021. Since the case study of this thesis is implemented on the dataset of Ireland, further forecasting of price occurs on Irish dairy products. The model was implemented using the time series concept wherein the time series of events were fed as input to the model to execute univariate forecasting. The dataset so obtained was classified into training and testing sets; where the whole dataset was utilized for learning purpose and just the last columns from dataset have been used for the testing purpose. The diagram below gives an overall view of the implementation process.

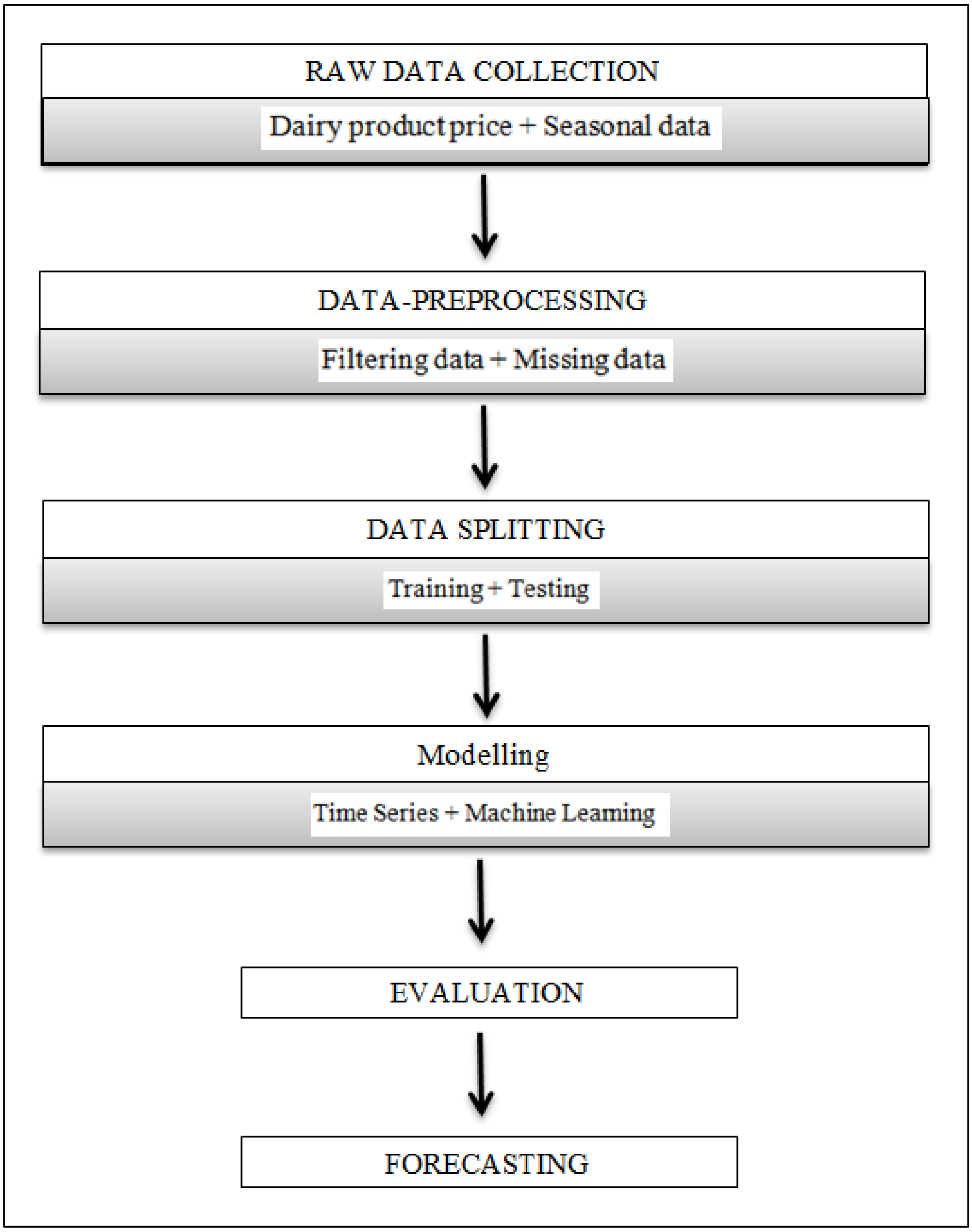

Figure 6: Implementation Details of the proposed work

The entire implementation took place for price fluctuations in “butter” that would take place due to seasonal changes in Ireland. Further the model was implemented using four machine learning algorithms that were finally evaluated against averaging parameters.

4.1 Training and Testing

A model undergoes the training and testing phase, in which the machine is initially trained with the dataset so obtained from the repository. The testing part is generally done in the later stages wherein the model is checked for accuracy. However, this splitting of data in to training and testing is done on temporal data. Every time a split takes place, the model uses different records present from the dataset. Since the thesis makes use of time stamp model, the input which is fed is historical data containing past values. These past values are fed to the model on a time axis. Hence, the process takes place through a time stamp model. However, this splitting of data occurs in a sequence and is splitted in an 80:20 ratio wherein Eighty percent of the data is utilized for training the model, while twenty percent is used to test it. Once the model has been validated, further procedure of evaluation takes place. For our study, for training purpose, the entire dataset was used and fed as an input to the time series model. Later we filtered the dataset and used the last columns for testing purpose. Butter prices of previous 12 months were used in the testing phase.

6 Results and Discussion

This section briefs the evaluation metrics used to forecast the model. Below are the mentioned concepts that are used to evaluate the final results.

- Mean squared error
  The functionality of this metric is to calculate the average square of all the forecasted errors in the testing phase. The advantage of this metric is that it manages all the positive and negative variances present in the dataset. However, its evaluation involves several units during the training phase; thereby making it a limitation.
- Root Mean Squared Error
  This metric is responsible to calculate the squared root obtained from mean error and further calculates the average error between actual data and forecasted data. The limitation of mean error is however outperformed by root mean square error as the units of error and target are the same while implementing and calculating the mathematical operations required for root mean square error.

Below graph depicts the visualization of data for "butter" obtained from the dataset:

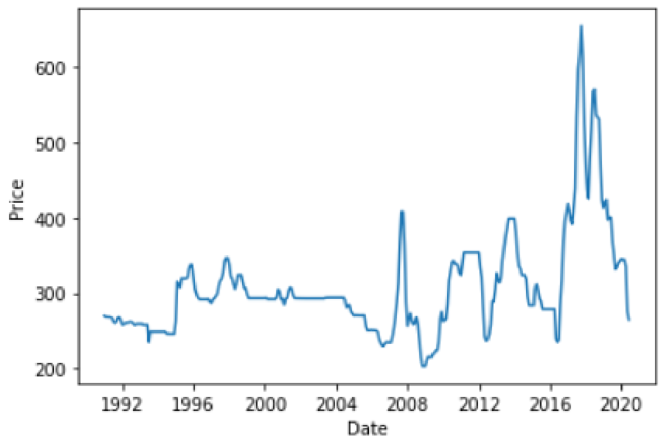

Figure 7: A time series illustration of butter prices from 1992-2020

6.1 Experiment 1- ARIMA Model

The aim of this experiment is to predict the seasonal prices of Butter product using ARIMA model and forecasting future Butter values for the next three years. The model was executed in two phases:

training and testing. Butter prices of previous 12 months were used in the testing phase and finally the forecasting was done for the next 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

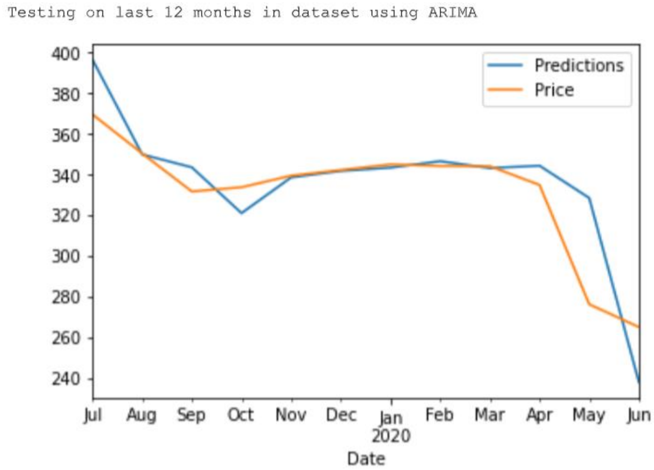

Figure 8: A prediction graph depicting butter prices using ARIMA model with month details on the x-axis and price of butter on the y-axis

The graph so generated above, depicts the actual price values in orange line and the predicted values by the model in blue line. Evaluation metrics of Mean Squared Error and Root Mean Squared Error were further calculated, and an error factor of 19.60 was observed using the ARIMA model.
Below graph illustrates forecast of butter prices for the next three years.

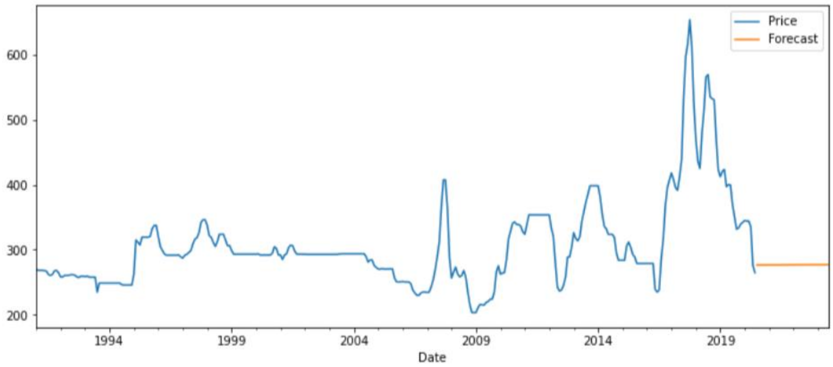

Figure 9: A forecasting graph depicting butter prices using ARIMA model with month details on the x-axis and price of butter on the y-axis

The above graph illustrates butter prices from 1994 to 2019 in blue line and the orange line forecasts the price for the next three years.

6.2 Experiment 2- ARIMA GARCH Model

The aim of this experiment model is to predict the seasonal prices of Butter product using ARIMA-GARCH model and forecasting future Butter values for the next three years. The model was executed in two phases: training and testing. Butter prices of previous 12 months were used in the testing phase and finally the forecasting was done for the next three years.

Table 2: Evaluation Metrics of ARIMA GARCH Model

EVALUATION METRICS -2 : ARIMA GARCH MODEL	
Root Mean Square Error	300.56
Mean Squared Error	90338.87

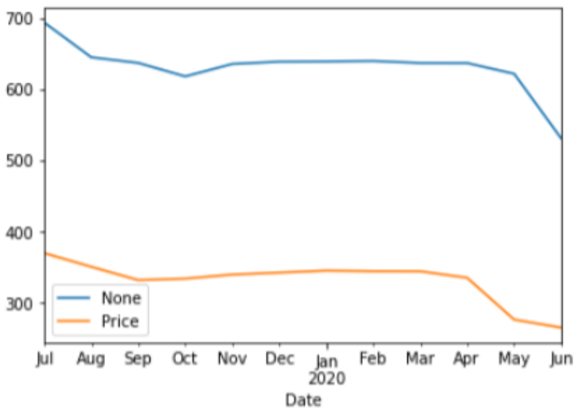

Figure 10: A prediction graph depicting butter prices using ARIMA- GARCH model with month details on the x-axis and price of butter on the y-axis

The graph so generated above, depicts the actual price values in orange line and the predicted values by the model in blue line. Evaluation metrics of Mean Squared Error and Root Mean Squared Error were further calculated, and an error factor of 300.56 was observed using the ARIMA-GARCH model. As seen in the graph, ARIMA-GARCH model is not the appropriate model for predicting seasonal prices using Irish Dairy Products dataset.
Below graph illustrates forecast of butter prices for the next three years.

Below graph illustrates forecast of butter prices for the next three years.

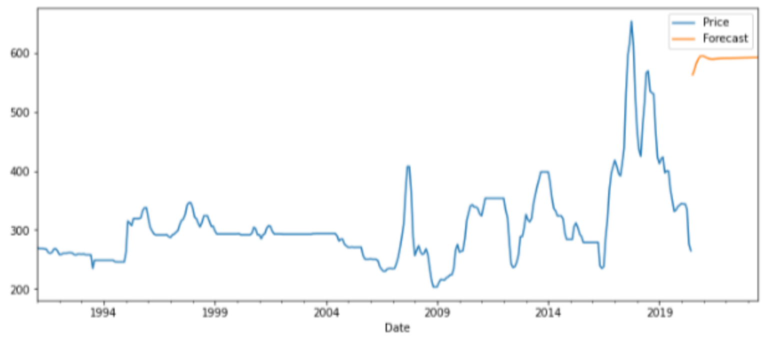

Figure 11: A forecasting graph depicting butter prices using ARIMA – GARCH model with month details on the x-axis and price of butter on the y-axis

The above graph illustrates butter prices from 1994 to 2019 in blue line and the orange line forecasts the price for the next three years. As clearly seen from the above graph, ARIMA-GARCH model does not performs well in forecasting Butter Prices.

6.3 Experiment 3- Simple Exponential Model

The aim of this experiment model is to predict the seasonal prices of Butter product using Simple Exponential model and forecasting future Butter values for the next three years. Butter prices of previous 12 months were used in the testing phase and finally the forecasting was done for the next three years.

Table 3: Evaluation Metrics of SEM Model

EVALUATION METRICS -3 : SEM MODEL	
Root Mean Square Error	26.51
Mean Squared Error	703.19

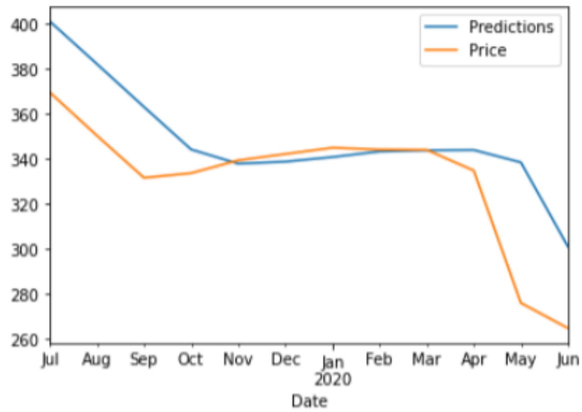

Figure 12: A prediction graph depicting butter prices using SEM model with month details on the x-axis and price of butter on the y-axis

The graph so generated above, depicts the actual price values in orange line and the predicted values by the model in blue line. Evaluation metrics of Mean Squared Error and Root Mean Squared Error were further calculated, and an error factor of 26.51 was observed using the SEM model.

Below graph illustrates forecast of butter prices for the next three years.

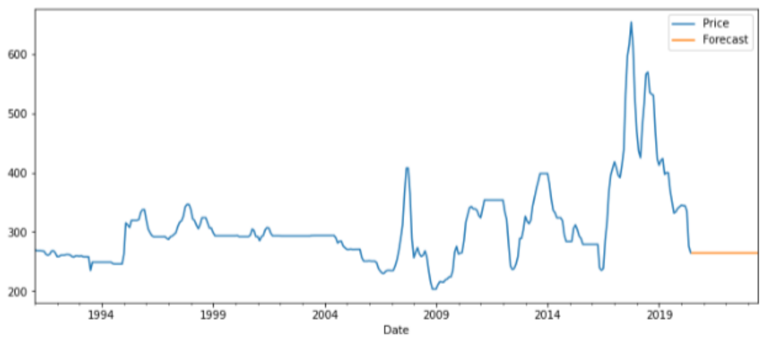

Figure 13: A forecasting graph depicting butter prices using SEM model with month details on the x-axis and price of butter on the y-axis

The above graph illustrates butter prices from 1994 to 2019 in blue line and the orange line forecasts the price for the next three year.

6.4 Experiment 4- SARIMA Model

The aim of this experiment model is to predict the seasonal prices of Butter product using SARIMA model and forecasting future Butter values for the next three years. Butter prices of previous 12 months were used in the testing phase and finally the forecasting was done for the next three years.

Table 4: Evaluation Metrics of SARIMA Model

EVALUATION METRICS -4 : SARIMA MODEL	
Root Mean Square Error	18.52
Mean Squared Error	343.208

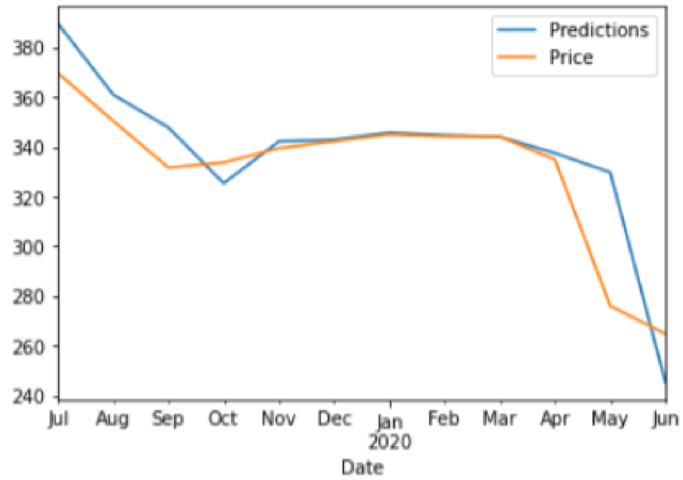

Figure 14: A prediction graph depicting butter prices using SARIMA model with month details on the x-axis and price of butter on the y-axis

The graph so generated above, depicts the actual price values in orange line and the predicted values by the model in blue line. Evaluation metrics of Mean Squared Error and Root Mean Squared Error were further calculated, and an error factor of 18.52 was observed using the SARIMA model. As seen in the above graph, SARIMA model gives the best prediction for Butter values.
Below graph illustrates forecast of butter prices for the next three years.

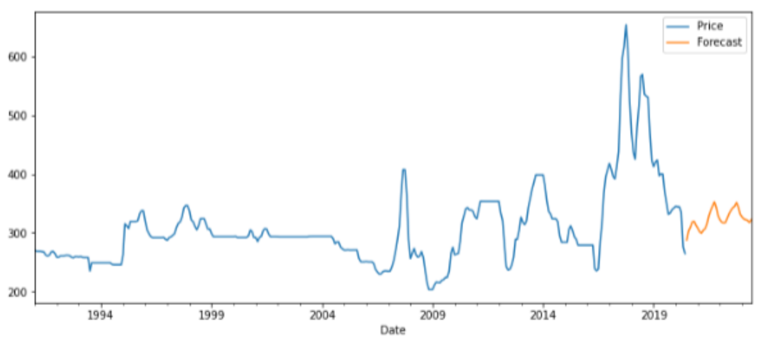

Figure 15: A forecasting graph depicting butter prices using SARIMA model with month details on the x-axis and price of butter on the y-axis

The above graph illustrates butter prices from 1994 to 2019 in blue line and the orange line forecasts the price for the next three years.

7. Conclusion and Future Work

The primary objective of the thesis was to implement a time series model to forecast prices of Irish dairy products. The thesis was implemented using ARIMA model and machine learning models. The entire workflow of the model took place by obtaining dataset from the European repository and executing training and testing phases on it. Full dataset was used on the training phase and the data of previous 12 months were used for the testing phase. A time series concept was then applied wherein the price data of butter for the last 12 months was given as input to the model. Later four machine learning models were implemented that were responsible to forecast the price fluctuations for the next 3 years. On result evaluation that was calculated using root mean square error, it was observed that the SARIMA model outperformed the rest of the models by providing the least squared mean error of 18 percent.

In the future, this work can further be expanded to forecast dairy prices of all the products involved, with a much larger data from multiple countries. Also, the same concepts can further be used on a much larger span of time series historical events that can assist to forecast similar applications.

Acknowledgement

I would like to thank the National College of Ireland and its School of Computing's support and technical personnel for their continuous encouragement. I would want to convey my gratefulness to my supervisors Paul Stynes and Pramod Pathak, whose skills and experience were crucial in the development of the study objective and methodologies. Your insightful comments inspired me to refine my ideas and increase the quality of the research. In addition, I'd want to express my gratitude to my parents for their unwavering support and wise counsel.

References

Adediran SA, Ratkowsky DA, Donaghy DJ, Malau-Aduli AEO (2012) Comparative evaluation of a new lactation curve model for pasture-based Holstein- Friesian dairy cows. J Dairy Sci 95:5344–535

FAO (2020) Food outlook, biannual report on global food markets. Italy, Rome

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

Lohano DH, Soomro MF (2006) Unit root test and forecast of milk production in Pakistan. Int J Dairy Sci 1(1):63–696

Akter, Sh., and Rahman,S.(2010) Milk supply forecasting for a dairy cooperative in the UK using univariate time series modelling techniques, In book: Agricultural Economics: New Research (pp.227–242), Chapter:11, Publisher: Nova Science Publishers, New York, Editors: Thomas H. Lee

M. W. Kadous. Learning comprehensible descriptions of multivariate time series. In ICML, pages 454–463, 1999

J. J. Rodriguez, J. J. R. Guez, and C. J. Alonso. Boosting interval-based literals: Variable length and early classification. 2002

S. S. Skiena. The algorithm design manual: Text, volume 1. Springer Science & Business Media, 1998

J. Schmidhuber. Deep learning in neural networks: An overview. Neural Networks, 61:85–117, 2015

L. Ye and E. Keogh. Time series shapelets: a new primitive for data mining. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 947–956. ACM, 2009

Z. Xing, J. Pei, and S. Y. Philip. Early classification on time series. Knowledge and information systems, 31(1):105–127, 2012

Z. Xing, J. Pei, G. Dong, and S. Y. Philip. Mining sequence classifiers for early prediction. In SDM, pages 644–655. SIAM, 2008

O'Connell, A., MCPARLAND, S., Reugg, P. L., O'brien, B. & Gleeson, D. 2015. Seasonal trends in milk quality in Ireland between 2007 and 2011. Journal of Dairy Science

Ahmed, F., Shah H., Raza, I. & Saboor, A. 2011. Forecasting milk production in Pakistan. Pakistan Journal of Agricultural Research, 24

Bolotova, Y. V. 2017. An Analysis of Retail Fluid Milk Pricing in the Eastern United States

Rosa, F., Weaver, R. & Vasciaveo, M. 2016. Dairy Price commodity in Italy: volatility and forecast after milk quotas. Proceedings in Food System Dynamics, 112-115

Hansen, B. G. 2020. Forecasting prices of dairy commodities – a comparison of linear and nonlinear models. Irish Journal of Agricultural and Food Research, 59

Colas, G., and G. Brice. "Seasonal variations of the fertilizing capacity of the deep-frozen ram semen." In Proceedings of the eighth international congress on animal reproduction, vol. 4, pp. 977-980. 1976

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

Harris, Kim S., and Raymond M. Leuthold. "A comparison of alternative forecasting techniques for livestock prices: A case study." North Central Journal of Agricultural Economics (1985): 40-50