UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

MARCELA SILVA GUIMARÃES VASCONCELLOS

# A DATA-BASED APPROACH TO NEWSVENDOR PROBLEMS SUBJECT TO PURCHASE PRICE UNCERTAINTY

Recife
2021

MARCELA SILVA GUIMARÃES VASCONCELLOS

# A DATA-BASED APPROACH TO NEWSVENDOR PROBLEMS SUBJECT TO PURCHASE PRICE UNCERTAINTY

Dissertation presented to the Graduate Program in Industrial Engineering at Universidade Federal de Pernambuco, as partial requisite for obtaining the title of Master of Industrial Engineering.

**Concentration field**: Operations Research.

**Advisor**: Profa. Dra. Isis Didier Lins.

Recife

2021

Catalogação na fonte:
Sandra Maria Neri Santiago, CRB-4 / 1267

V331d Vasconcellos, Marcela Silva Guimarães.
A data-based approach to newsvendor problems subject to purchase price uncertainty / Marcela Silva Guimarães Vasconcellos. – 2021.
67 f.: il., figs., tabs.

Orientadora: Profa. Dra. Isis Didier Lins.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-graduação em Engenharia de Produção. Recife, 2021.
Inclui referências.

1. Engenharia de produção. 2. Problema do jornaleiro. 3. Política de estoque. 4. Machine learning. 5. Previsão de séries temporais. 6. Otimização. I. Lins, Isis Didier (Orientadora). II. Título.

UFPE

658.5 CDD (22. ed.) BCTG/2022-204

MARCELA SILVA GUIMARÃES VASCONCELLOS

# A DATA-BASED APPROACH TO NEWSVENDOR PROBLEMS SUBJECT TO PURCHASE PRICE UNCERTAINTY

Dissertation presented to the Graduate Program in Industrial Engineering at Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, as partial requisite for obtaining the title of Master of Industrial Engineering.

**Concentration field**: Operations Research.

Approved in: 24/02/2021.

**EXAMINATION COMMITTEE**

Profa. Dra. Isis Didier Lins (Advisor)
Universidade Federal de Pernambuco

Prof. Dr. Márcio das Chagas Moura (Internal Examinator)
Universidade Federal de Pernambuco

Prof. Dr. Andrew Trapp (External Examinator)
Worcester Polytechnic Institute

## ABSTRACT

Poor procurement decisions, especially involving perishable or short life-cycled products, which will have to be disposed of, can cost companies large portions of their profits. The newsvendor problem addresses inventory decisions to assist retailers in deciding just the right order quantity while still subject to uncertainty. Efficient time series forecasting techniques, including the use of machine learning models, have helped improve financial results by offering insight on future outcome-based decisions. In this dissertation, a comprehensive study was developed around the situation in which a retailer is faced with the problem of stochastic purchase prices and must decide when is the best day to place an order, as well as how much to buy to restock his perishable supply. To support the decision-making process, the problem was modeled as a variant of the newsvendor problem, subject to two decision variables: when to place the order and how much to buy. SARIMA, Prophet, MLP, RNN, and LSTM models were used for time series forecasting and were assessed in their ability to support the decision-making by forecasting future purchase prices. All forecasting-based decisions outperformed the zero-information scenario in terms of total costs. Two models (MLP and RNN) outperformed the others in terms of supporting the decision of when to buy.



Keywords: newsvendor problem; inventory policy; machine learning; time series forecasting; optimization.

# RESUMO

Más decisões de compras, especialmente as que envolvem produtos perecíveis ou de ciclo de vida curto que terão de ser descartados, podem custar às empresas uma grande parte de seus lucros. O problema do jornaleiro trata de decisões acerca de estoque para ajudar os varejistas a decidir a quantidade ideal para cada pedido realizado, embora ainda sujeito à incerteza. Técnicas eficientes de previsão de séries temporais, incluindo o uso de modelos de aprendizado de máquina, ajudaram a melhorar os resultados financeiros, oferecendo uma visão sobre futuras decisões baseadas em resultados. Nesta dissertação, foi desenvolvido um estudo abrangente em torno de um cenário em que um varejista está sujeito ao problema de preços de compra estocásticos e deve decidir quando é o melhor dia para fazer um pedido, assim como quanto comprar para reabastecer seu estoque de perecíveis. Para apoiar o processo de tomada de decisão, o problema foi modelado como uma variante do problema do jornaleiro, sujeito a duas variáveis de decisão: quando fazer o pedido e quanto comprar. Os modelos SARIMA, Prophet, MLP, RNN e LSTM foram utilizados para previsão de séries temporais e avaliados em sua capacidade de apoiar a tomada de decisão ao prever futuros preços de compra. Todas as decisões baseadas em previsões superaram o cenário sem informação em termos de custos totais. Dois modelos (MLP e RNN) superaram os outros em termos de apoio à decisão de quando comprar.

Palavras-chave: problema do jornaleiro; política de estoque; machine learning; previsão de séries temporais; otimização.

**LIST OF FIGURES**

## LIST OF TABLES

**LIST OF ABBREVIATIONS AND ACRONYMS**

| | |
|---|---|
| Adam | Adaptive moment estimation |
| AR | Autoregressive |
| ARIMA | Autoregressive Integrated Moving Average |
| ARMA | Autoregressive Moving Average |
| BFGS | Broyden–Fletcher–Goldfarb–Shanno |
| DNN | Deep Neural Networks |
| IPCA | National Price Index to the Broad Consumer |
| LBFGS | Limited-memory Broyden–Fletcher–Goldfarb–Shanno |
| LSTM | Long Short-Term Memory |
| MA | Moving Average |
| MFNP | Multi-feature Newsvendor Problem |
| ML | Machine Learning |
| MLP | Multi-Layer Perceptron |
| ReLU | Rectified Linear Unit |
| RNN | Recurrent Neural Networks |
| SARIMA | Seasonal Autoregressive Integrated Moving Average |
| SGD | Stochastic Gradient Descent |

**SUMMARY**

# 1 INTRODUCTION

Inventory management is an important field of study in management sciences (NEALE *et al.*, 2004; MUNYAO *et al.*, 2015). Although keeping high stock levels might lead to holding costs and fixed capital, it is also helpful in reducing shortage risks, opportunity costs of unsold items, and keeping the overall supply chain flowing. In some industries, however, this topic is more critical than in others. Consider a chocolate manufacturer that needs to stock up on Easter eggs. It must produce, distribute, and sell all its Easter eggs in time for customers to buy them. If it runs out of products, by the time it programs, produces and re-distributes the extra items, the holiday has passed, along with all the demand for the item. Therefore, it must guarantee enough items the first time around. On the other hand, if it orders too many, it will have to deal with enormous amounts of wasted inventory. The problem faced by the chocolate company is an example of a newsvendor problem, in which a producer must decide on an order size that is just the right amount because ordering either more or less will be harmful to the business.

The first skeleton of the newsvendor problem was originally proposed by EDGEWORTH (1888), who investigated the case of banks that need to keep a certain amount of money in cash to serve those clients who wish to withdraw money from their accounts, subject to the unlikely chance of all clients deciding to empty their accounts at the same time. It was first called the newsboy or newsvendor problem by MORSE & KIMBALL (1951), who named it after a situation in which a newspaper seller must place a single order with the number of papers he intends to sell during the following day. If he orders too many, he will lose money because no one will be willing to purchase the leftover product the following day. On the other hand, if he orders too few, he will earn less profit that day than he could have, not to mention displeasing customers that might choose another newsstand the following day. In this context, he is not able to place complementary orders throughout the day.

This problem is present in many industries such as airlines, supermarkets, entertainment, fashion, technology, or perishable food manufacturers (CHOI, 2012). Examples of applications of this problem are airline overbooking decisions, replenishing stocks of perishable products, and quantifying the production lot of seasonal or short-cycled products (PORTEUS, 2008). Specific situations can also be modeled by different variants of the classic problem, allowing it to be applied to a diversity of scenarios such as situations involving multiple products (SHAO & JI, 2006; ABDEL-MALEK & MONTANARI, 2005), multiple periods (KIM *et al.*, 2015; BEHRET & KAHRAMAN, 2010), interdependent demands (LOTFI et al., 2020), quantity discount on selling prices (KHOUJA & MEHREZ, 1996; ZHANG et al., 2020), and many others (KHOUJA, 1999).

The main criteria to characterize a newsvendor-type problem are i) the existence of a random demand for a given product, ii) a single order being placed for each time period, iii) known costs of ordering excessively, called overage, and of ordering short of the demand, also referred to as underage (PORTEUS, 2008). The problem is centered around deciding on the most profitable order size. It is an optimization problem, tied to a search for the best decision-making for a given scenario, and a prediction problem, because the decision will always depend on forecasts or assumptions about the customer's demand behavior. The classic solution requires making assumptions about the unknown demand, whether it be a single value, such as a measure of central tendency, or a probability distribution (SILVER *et al.*, 2017).

The widespread development of data-based predictions contributed to the emergence of new methods to predict this expected demand, which is then used as input to an optimization model with total overage and underage costs as the objective function (BERTSIMAS & THIELE, 2005). Machine Learning (ML) methods have become increasingly popular in time series forecasting (GILLILAND, 2020) for their ability to model nonlinear behavior without specific parametrization, unlike usual classic models. RUDIN & VAHN (2013) suggest that applying data-based techniques in the newsvendor problem-solving process can significantly improve results, reducing total cost. In HUBER *et al.* (2019), the exploration of data-based methods is analyzed as an important source of information that can aid inventory management in newsvendor-like problems, including by predicting the order quantity directly after training a machine on how to minimize costs, ignoring the mathematical formulation of the optimization problem.

The demand uncertainty, however, is not always the most difficult uncertainty to master. In retail, Normal and Gamma distributions have proven to fit well the demand for fast-moving items, as long as the mean is significantly larger than the standard error and historical data is known (RAMAEKERS & JANSSENS, 2008). Most of the newsvendor literature has been directed to modeling the uncertainty in demand but has not yet explored the context of supplier uncertainty (ULLAH *et al.*, 2019). Alterations in the supply of perishable products can be significant, especially due to situations such as seasonality of crops, droughts, amongst other climatic and macroeconomic factors that lead to fluctuating market prices (GARGANO & TIMMERMANN, 2014). A retailer that does not wish to see empty shelves and displeased customers will be subject to paying higher prices for its products unless it can anticipate these alterations and use them to its advantage. In almost all newsvendor literature, overage and underage prices have been considered as constant parameters, although they are directly affected by purchase price alterations. There is a need to better understand the effect of stochastic purchase prices, especially of perishable items, in the newsvendor optimization problem and decision-making.

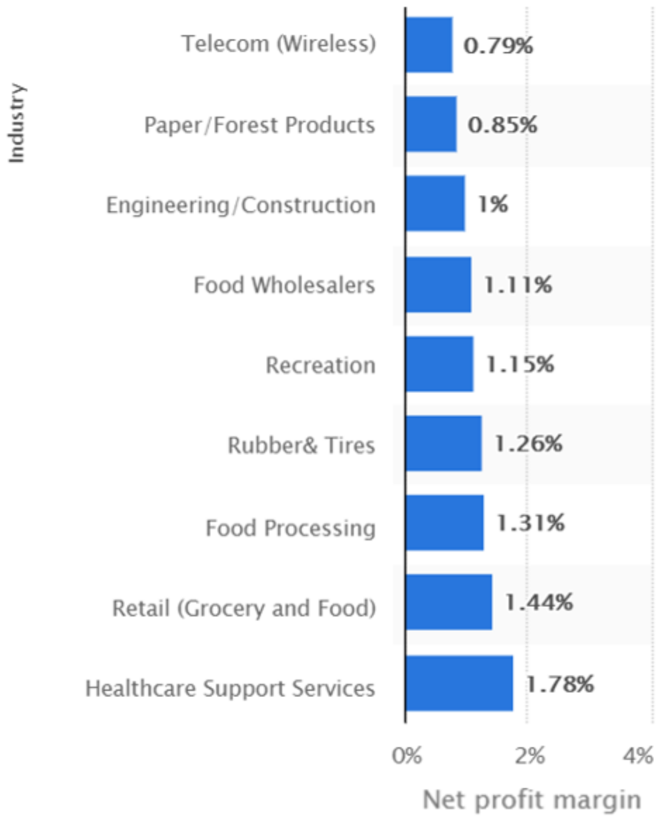
In this work, a newsvendor problem subject to purchase price uncertainty is proposed and analyzed. To illustrate the case, a retailer that must purchase perishable products on a weekly basis is considered. In order to respect newsvendor properties, the purchase decision to supply for a certain week must be made the week before. So, although it must make a one-time purchase decision for the following week and cannot replenish its stock as the week goes by, it can decide which is the best day to put in the order. The decision then becomes not only how much to order but also which day of the previous week. This allows the decision-maker to account for possible situations that affect its purchase price.

This work aims at evaluating the cost performance of newsvendor decisions that take purchase price stochasticity into consideration and comparing it to that of the strategy obtained by considering only demand uncertainty. In order to achieve this, the traditional newsvendor problem is adapted to account for purchase price-dependent overage and underage costs. Future prices are estimated by use of time series forecasting techniques such as the seasonal autoregressive integrated moving average (SARIMA) (BOX *et al.*, 2015) and Prophet models (TAYLOR & LETHAM, 2018), as well as ML models, such as the Multi-Layer Perceptron (MLP) (ROSENBLATT, 1958), and Deep Neural Networks (DNN), such as Recurrent Neural Networks (RNN) (RUMELHART *et al.*, 1986) and Long Short-Term Memory (LSTM) networks (HOCHREITER & SCHMIDHUBER, 1997), which are highly sophisticated prediction models that have been successful in time series forecasting (ZHANG, 2001). Predictions are then used as input to the optimization model that is proposed to jointly decide the ideal quantity and timing of the order.

## 1.1 JUSTIFICATION

Grocery and food retail is one of the industries with the lowest profit margins (Figure 1), usually falling between 1 and 2% (STATISTA, 2020). In this type of commerce, financial sustainability is ensured by a high volume of sales. In a highly competitive context such as retail, marking up prices will increase profit margins but may have severe effects on volume, jeopardizing the business model (MOHAMMED, 2009). Also, this type of strategy burdens the customers with paying for the increased profits, which can be achieved through the implementation of intelligent business solutions. The retail market serves all social classes, given its role to provide basic supplies such as food and hygiene products. Therefore, it is an essential service, and charging high prices for products is a form of exploration of even exclusion of low-income individuals from access to basic needs. With respect to that social role, retailers should seek alternative strategies to achieve better business performance.

Figure 1 – Industries with lowest (positive) profit margins in the U.S. in January 2020.




Source: Statista, S. (2020).

A viable alternative to increase profit margins is to reduce costs and expenses, which can be achieved by designing solid procurement and inventory policies. BIERY (2017) discusses the significance of managing inventory subject to such low profit margins. The retailer cannot afford to lose a customer who will rapidly resort to the competition if he or she doesn't find the desired products on the shelf. On the other hand, overstocking shelves might lead to spoiling material, which will incur high losses that may be difficult to cover by the low profit margins.

An additional impact of ordering excessively is producing large amounts of trash, which not only wastes natural resources, but often returns overly processed, or in plastic wraps, and pollutes the environment. Production and transportation of each discarded product also count negatively towards the retailer's carbon footprint. Transport alone accounts for 24% of $CO_2$ emissions from energy (RITCHIE(a), 2020). Beef, for example, a highly perishable product, is one of the strongest contributors to carbon emissions, contributing as much as 60kg of CO2-equivalents to the atmosphere, compared to other products such as corn, for example, which contributes only 1kg (RITCHIE(b), 2020). Therefore, it is important to act responsibly when estimating the truly needed number of products in order to maintain an environmentally sustainable activity. Having a procurement and stock policy that provides just the right balance can be critical to business sustainability.

The newsvendor problem describes this exact scenario (EDGEWORTH,1888). It focuses especially on the tradeoff between losses incurred from unmet demand and those due to spoilage of products. It addresses the decision regarding what order size to place in each fixed time-step for products that expire, become obsolete, or lose part of their value in the following time-step. As a way of reducing not only spoilage and opportunity costs but also product purchase costs, this work proposes a variant of the newsvendor in which purchase price variability is considered. The optimization problem is expanded to combine the decisions of how much to order and when to place the order.

Although current state-of-the-art has suggested a high potential of total cost reduction from the application of data-based techniques in the newsvendor problem-solving process (RUDIN & VAHN, 2013), literature shows that, in the greater part, only more traditional methods such as Random Forests (BREIMAN, 2001) and shallow MLPs have been applied to this problem (BERTSIMAS & THIELE, 2005; RUDIN & VAHN, 2013; BERTSIMAS & KALLUS, 2014). This shows there is still space to explore more advanced methods.

According to MCLAUGHLIN *et al.* (2015), procurement decisions can be especially complex in the context of product perishability and supplier-distributor fragmentation, which leads to a greater need for produce buyers to make decisions using current, accurate, and fast-changing information. Sophisticated methods, such as DNNs, have been more recently developed involving neural networks that are capable of modeling complex, nonlinear patterns (SAMARASINGHE, 2016). Predictions on sequential data can be improved by applying such models that were designed specifically for this type of data, such as Long Short-Term Memory (LSTM) (HOCHREITER & SCHMIDHUBER, 1997) and Recurrent Neural Networks (RNN) (RUMELHART *et al.*, 1986). Advanced Artificial Intelligence tools have been known to contribute to relevant problems in supply chain management contexts (GUIMARÃES *et al.*, 2019) and will therefore be explored in this work.

## 1.2 OBJECTIVES

This section lays out the main focus of this work with respect to what it intends to achieve.

### 1.2.1 General Objectives

This study's general goal is to propose a methodology for defining optimal order quantity and timing in a newsvendor problem subject to purchase price uncertainty based on time series forecasting techniques. The performance of SARIMA, Prophet, ML, and DNN are investigated and their impact on cost results are compared to a strategy that does not consider the existing purchase price uncertainty.

### 1.2.2 Specific Objectives

- To investigate the current state-of-the-art works that model a newsvendor-type situation in which the newsvendor is subject to uncertainty regarding its purchase prices, as well as lay the fundamental grounds in theory behind the classical formulation of the problem and time series forecasting techniques;
- To state the modified version of the optimization problem and how it differs in theory from the classical newsvendor once the purchase price fluctuations are considered;
- To gather quality historical data including purchase price and demand time series, as well as underage and overage parameters or, if necessary, simulate values that are based on real-life behavior and can therefore be used to adequately evaluate the performance of the compared methods in real scenarios;
- To choose adequate packages, and implement the algorithms for each of the selected forecasting techniques, and objective function performance evaluation;
- To evaluate and compare the performance of models used to estimate future purchase price based on historical data using each technique;
- To solve the optimization problem using the estimated purchase prices as input;
- To evaluate the methodology and compare it to currently applied strategies disregarding purchase price variations.

## 1.3 METHODOLOGY

This study proposes a novel approach to the decision-making process in a variant of the newsvendor problem that includes stochastic purchase prices practiced by the newsvendor’s supplier. This quantitative approach combines time series prediction models with a theoretical mathematical background to obtain a methodology that optimizes the decision in the proposed variant. Multiple strategies are tested, and their effects on the final cost are described. Although the nature of each selected technique may suggest an intuition behind which models should work best when applied to this particular type of time series, it is not a goal of this study to provide evidence of causal relations. Therefore, it can be classified as descriptive research, regarding its goals (FONTELLES et al., 2009).

In respect to its nature, the research is theoretical, due to the methodology’s generalized application to any newsvendor-type problem subject to purchase price uncertainty. Regarding its procedures, the research involves modeling to describe a variant of the newsvendor problem mathematically and proposing a methodology to solve it.

This work will be developed in Python, using Jupyter Notebook, and all used packages are described in the text.

The study requires the following steps to achieve the general purpose:

1. Literature review;
2. Optimization problem modeling;
3. Data collection;
4. Time series forecasting algorithm implementations;
5. Purchase price estimation;
6. Model evaluations and comparison;
7. Optimization model solution;
8. Methodology evaluation and conclusion.

## 1.4 DISSERTATION STRUCTURE

This dissertation is divided into 5 sections, starting with this introduction, and containing all the step-by-step involved in the logical structuring of this research. The content of the following sections is briefly described below:

**Section 2:** the theoretical background and literature review of essential concepts related to the Newsvendor Problem and time series forecasting techniques;

**Section 3:** gives a detailed description of the proposed methodology in this work;

**Section 4:** presents the results of the implementation of the proposed method on a dataset that represents meat prices;

**Section 5** provides some concluding remarks and comments about future works.

# 2 THEORETICAL BACKGROUND AND LITERATURE REVIEW

The first part of this section introduces the newsvendor problem in mathematical terms and explores the main developments on the research on this problem, especially those that are relevant to this work. The second part describes the different time series forecasting methods that will be applied in Section 3.

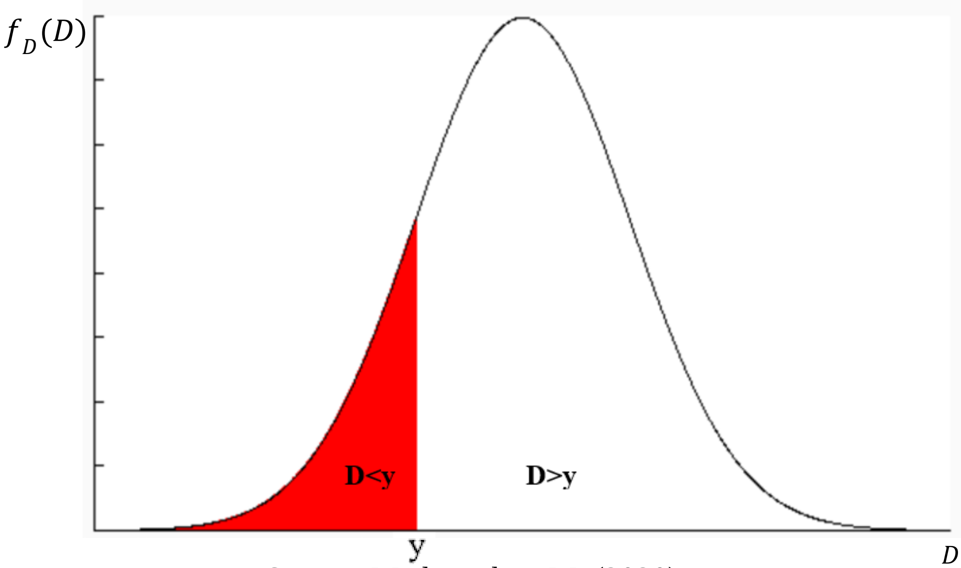
## 2.1 THE NEWSVENDOR PROBLEM

The problem seeks to minimize the expected total cost (C) to which the newsvendor is exposed, considering the unit costs ($c_o$) of overage and the unit opportunity cost of underage ($c_u$) are known. Considering an unknown demand (D), one must seek which value of order size (y) allows for this minimum. The objective function is defined in Equation 2.1.

$$\min_{y} C(y) = E[\, c_u * \max\{0, D - y\} + c_o * \max\{0, y - D\} \,] \quad (2.1)$$

To illustrate this expected value, it is possible to use the case of a Normal probability distribution for the random variable that represents the demand, with density function $f_D(D)$ (Figure 2).

Figure 2 – Demand probability density function.




Source: Mathcracker, M. (2020).

The expected value of the amount between brackets is the sum of all possible values of the expression, multiplied by the probability of each of these values occurring (Equation 2.2). It is simple to observe that, if D ≠ y, only one of the two added factors can be greater than zero as either $\max\{0, D-y\}$ or $\max\{0, y-D\}$ is zero. They are both 0 only for D = y. Therefore, we can look separately at the cases when D > y and when D < y.

$$C(y) = \int_{y}^{\infty} c_u(D-y)dF(D) + \int_{-\infty}^{y} c_o(y-D)dF(D) \tag{2.2}$$

Rearranging Eq. 2.2, it is possible to obtain Eq. 2.3 and, subsequently, Eq. 2.4.

$$C(y) = c_u \int_{y}^{\infty} DdF(D) - c_u * y \int_{y}^{\infty} dF(D) + c_o * y \int_{-\infty}^{y} dF(D) - c_o \int_{-\infty}^{y} DdF(D) \tag{2.3}$$

$$C(y) = c_u \int_{y}^{\infty} DdF(D) - c_u * y * \big(1 - F(y)\big) + c_o * y * F(y) - c_o \int_{-\infty}^{y} DdF(D) \tag{2.4}$$

The derivative of this expression is equated to zero (Equation 2.5) to obtain its point of minimum:

$$-c_u * y * f(y) - c_u * \big(1 - F(y)\big) + f(y) * c_u * y + c_o * F(y) + f(y) * c_o * y - c_o * y * f(y) = 0 \tag{2.5}$$

Rearranging Equation 2.5, it is possible to find the critical fractile (Equation 2.6), explicitly proposed by WHITIN (1953).

$$F(y) = P(D \leq y) = \frac{c_u}{c_o + c_u} \tag{2.6}$$

This demonstration shows that, for a continuous space of solutions, the optimal solution to this problem depends solely on a relation between the underage and overage unit costs called the critical fractile. The desired order quantity then corresponds to the value for which the probability of the actual demand being entirely met by the ordered quantity, with or without residual stock, is equal to this critical fractile.

Many variants of this problem have been explored throughout the years, including special cases of demand behavior or the associated costs. In cases where the demand is price-dependent, for example, it is possible to include an additional decision variable representing the price to be charged by the retailer. This is called the price-setting newsvendor problem (WHITIN, 1955). YAO *et al.* (2006) explore how to model price-dependent stochastic demand and then achieve the combined pricing and inventory solutions. PETRUZZI & DADA (1999) review this extension of the problem and explore the value of information, comparing expected costs from situations with perfect demand information and price-setting policies subject to imperfect information. A more recent approach investigates the case of dynamic pricing: although the newsvendor may not place additional orders throughout the week, it may adjust its price according to the demand during the selling season (ULLAH *et* *al.*, 2019), which has proved to increase total profit. Major current developments on the price-setting newsvendor problem are investigated by DEYONG (2019).

Another scenario involves nonlinear purchase costs given by an oracle, that is, when unit purchase costs are not defined, and the cost of a certain lot size is given by the supplier, with no explanation regarding the function used to calculate it. The classic single-period newsvendor optimization model cannot be solved in this scenario, but a polynomial-time approximation can be obtained (HALMAN *et al.*, 2012). The present work considers that unit purchase costs are not affected by order size.

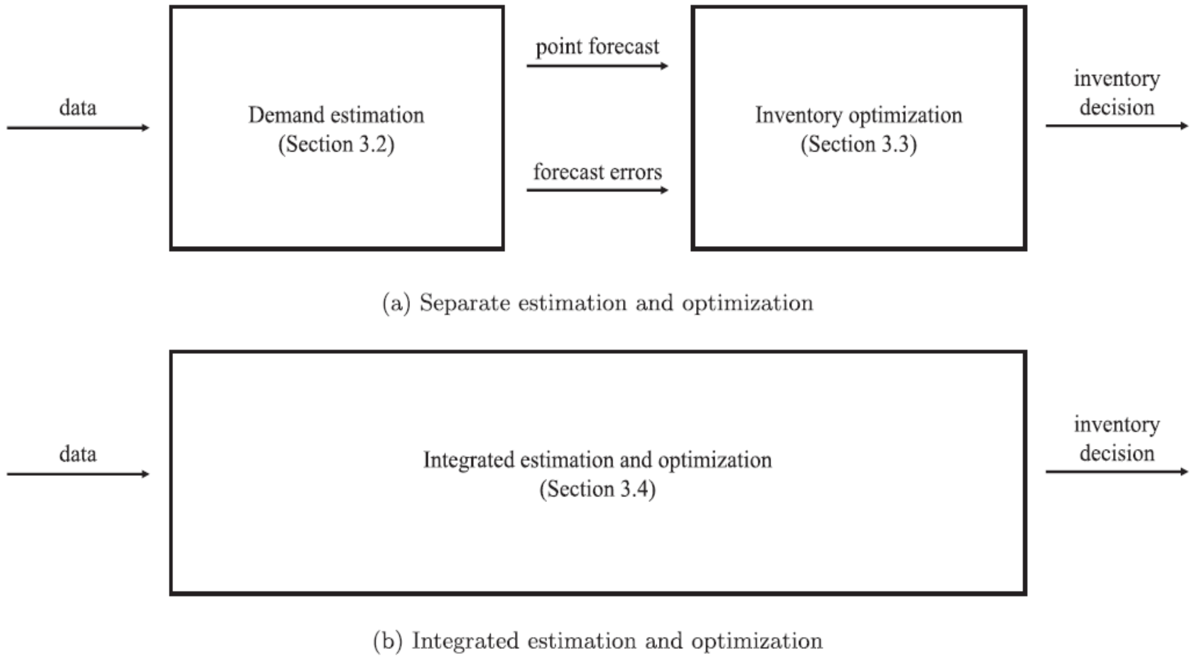
Demand in a certain time period might also depend on previous behavior. LOTFI *et al.* (2020) assess a two-period newsvendor problem, considering the possibility of interdependent demand and the effects of previously unsatisfied demand or residual products. In other formulations of the problem, the retailer does not have to discard unused products the following time period, but they somehow make use of them. HUA *et al.* (2020) discuss the case in which a retailer can exchange the remaining products for other items that might be serviceable. This exchange must, however, be less interesting than ordering only the needed amount of the initial item. Otherwise, there would be no overage cost, mischaracterizing a newsvendor situation. In the era of Big Data, available information such as weather conditions, monthly seasonality, and store location can be used to improve inventory policies, including in the newsvendor situation. The case in which exogenous variables are taken into consideration in the prediction and optimization tasks is called the multi-feature newsvendor problem (MFNP) (OROOJLOOYJADID *et al.*, 2016). Further variants of the traditional problem can be found in QIN *et al.* (2011).

Solving the traditional newsvendor problem and many of its variants requires knowing or making assumptions about the probability distribution that describes the demand behavior (SILVER, 2017). This information in practice is unknown, and wrong assumptions can lead to poor predictions and, consequently, bad inventory decisions (BAN & RUDIN, 2019). As an alternative, data-based predictions can be used both in estimating the most likely demand and its associated uncertainty, which can then be used as input to the optimization model. Besides offering an alternative approach to prediction, data-based models can incorporate external information into the newsvendor problem. This can be achieved by using ML models, especially neural networks related to sequential data, which can also be partnered with traditional time series forecasting techniques in order to present improved results.

Oroojlooyjadid, Snyder, and Takáč (2016) use DNNs in an integrated methodology to address both forecasting and optimization steps of the MFNP, as illustrated in Figure 3(b).

Several neural networks with two and three hidden layers were generated with randomly selected numbers of neurons and were then iteratively trained. After each iteration, the worst-performing networks were discarded until only the three highest-performing ones remained. Using the proposed approach showed better results than traditional Machine Learning approaches, such as Random Forests, and previously proposed approaches to the newsvendor problem (BERTSIMAS; THIELE, 2005; RUDIN; VAHN, 2013; BERTSIMAS; KALLUS, 2014). In this case, however, the model works as a black box, hiding the mathematical theory behind the inventory management problem and learning how to minimize holding and opportunity costs based on historical data, with little space for intuitive comprehension of what is motivating the choice of order quantity. ZHANG & GAO (2017) propose an improvement to the loss function used by most MFNP integrated solution approaches. Furthermore, BAN & RUDIN (2018) tested both integrated and two-step approaches and found that the latter outperforms the former.

Figure 3 – Integrated estimation and optimization solution for the MFNP.




Source: Huber, J. *et al.* (2019)

HUBER *et al.* (2019) further analyze the exploration of data as an important source of information that can aid inventory management in newsvendor-like problems. Data has been explored in different forms and steps of the available solution methodologies, and the study investigates which approach shows the best results for an MFNP. Data can be used for demand estimation through ML or more traditional times series forecasting techniques, in which case the optimal solution is reached as a derivation of the traditional critical fractile solution (Figure 3 (a)). In this case, the probability that is being compared to the critical fractile refers to the prediction error, not the demand value. This can be done by assuming a certain type of distribution and estimating parameters such as the standard deviation or observing empirical errors. The latter constitutes a second level for which historical data can be incorporated into the newsvendor solution.

A third level in which the use of data is investigated in this study is directly estimating the optimal order quantity from the available data (Figure 3(b)). This may be useful to find hidden patterns or behaviors that might contribute to better results. In contrast, it says little about the intuition behind the values provided by the machine and ignores the mathematical insights on inventory optimization in the newsvendor problem. The results show that using data-based methods in the first level can significantly contribute to cost reduction, whereas adding data information in the second level did not provoke relevant changes. Moreover, using a single model to integrate both phases of the problem did not obtain good results compared to making estimations and then optimizing as separate tasks. The ML techniques applied to this problem, however, were a simple MLP with a single hidden layer and Gradient Boosted Decision Trees, which are not state-of-the-art models used for time series prediction. This study applies more recent models, which are commonly applied to time series prediction, such as RNN and LSTMs, which can also be coupled with more traditional time series forecasting techniques that are very effective in modeling patterns such as seasonality and trends.

In the newsvendor context, an additional decision that can be added to the problem is the time of the purchase. Although this has been a discussed issue in stock policy literature (HU *et al.*, 2012), it is analyzed for long-term contracts and not for frequent short-cycled purchases, as in the case of the newsvendor. Due to the primary assumption that the newsvendor must place a single order for a given time period, few analyses of this problem have considered the possibility of anticipating or postponing said purchase. HU & SU (2018) show that in a newsvendor situation subject to price-dependent demand and stochastic purchase price, decisions can be improved by combining both pricing and procurement decisions. In order to respect newsvendor conditions, purchases are required to be made before the selling season starts, in what the authors call the pre-season. The decision regarding when to purchase involves product price and transportation costs, as well as holding costs.

In this context, the present work also seeks to determine the best time to purchase. However, it does not consider holding costs for anticipated orders because of the perishable nature of the products. With seasonal products, it is possible to anticipate orders with no effect on the selling season by fixing the delivery date irrespective of the moment the order is placed. Short shelf life food products, on the other hand, cannot be received and stored in advance, or they will expire before the week ends. In the case considered in this work, products are always received at the beginning of the week, regardless of the day the order was placed. When a product is bought will affect its costs, and consequently, the overage and underage costs associated with the newsvendor problem and the order quantity decision.

In respect to inventory policies subject to stochastic purchase prices, it has been shown that optimal replenishment policies are price-dependent and perform much better than strategies that do not consider future price variation (BERLING & MARTÍNEZ-DE-ALBÉNIZ, 2011). However, these proposed methods do not consider the single-order restriction of the newsvendor and product perishability. Also, they use geometric Brownian motion and the Ornstein-Uhlenbeck process to model purchase price behavior (SCHWARTZ, 1997). In HU & SU (2018), the behavior of the purchase price curve was modeled by the Black-Scholes equation (BLACK & SCHOLES, 1973). This study investigates the use of data-based models in purchase price modeling.

Therefore, this research is novel given its application of data-based forecasting to address the stochastic purchase price issue in newsvendor problems. Few works have included purchase prices in the newsvendor decision, and these have not applied data-based techniques to model this uncertainty.

## 2.2 TIME SERIES FORECASTING

One of the most consecrated methods for forecasting is the autoregressive integrated moving average (ARIMA) model (BOX *et al.*, 2015). It is an aggregation of both the autoregressive (AR) and the moving average (MA) models for stationary time series, with an additional factor that considers a possible stationary behavior of the series after a certain number of differentiations. The seasonal autoregressive integrated moving average (SARIMA) model is a modification of the ARIMA, which accounts for the seasonal variation of the series in order to make a better prediction. A SARIMA (p, d, q)(P, D, Q)m model is described by 7 values that describe the series' behavior and a time-step value's dependency on previous values of the series: i) p: Trend autoregression order, ii) d: Trend difference order, iii) q: Trend moving average order, iv) P: Seasonal autoregressive order, v) D: Seasonal difference order, vi) Q: Seasonal moving average order, v) m: The number of time steps for a single seasonal period. Special cases of this model, such as the autoregressive moving average

(ARMA) and the original simple AR and MA models, can be achieved by setting certain parameters to one or zero, as described in Table (1). Other widely used traditional methods are variants of the exponential smoothing approach, such as Simple Exponential Smoothing and Holt-Winters (BROWN & MEYER, 1961; GARDNER, 2006). In this work, the SARIMA will be used as a traditional model to be compared to more recent methods.

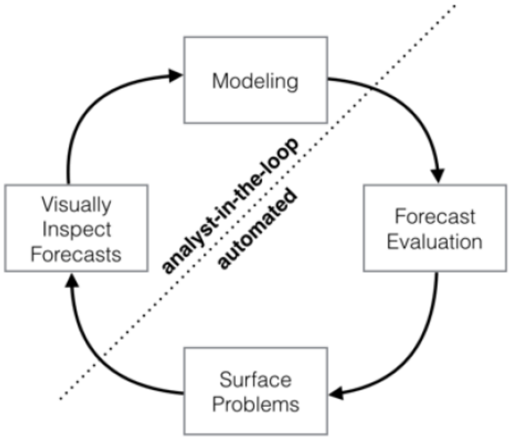
Table 1 - Special Cases of SARIMA

| Model | Parameters $(p,d,q)(P,D,Q)_m$ |
|---|---|
| Autoregressive (AR) | $(p,0,0)(0,0,0)_1$ |
| Moving Average (MA) | $(0,0,q)(0,0,0)_1$ |
| Autoregressive Moving Average (ARMA) | $(p,0,q)(0,0,0)_1$ |

Source: The author (2021)

The Prophet model (TAYLOR & LETHAM, 2018) is a recent linear model proposed by Facebook that offers an innovative approach to time series modeling by combining statistical forecast and judgmental forecast. The former is represented by a modular regression model with seasonality, trend, and holiday components. After adjusting the parameters to the data, specific forecasts are flagged for revision by a specialist. The structure of the proposed model admits user-defined parameters and options to allow the analyst to recursively feed the model for improvements in what the authors call Analyst-in-the-Loop modeling (Figure 3). The analysts can adjust how strongly predictions are influenced by seasonal trends using smoothing parameters. It is also possible to define upper and lower bounds for predictions to limit the effect of trends, which can be useful if a change of pattern or growth saturation point is expected. Finally, the user can aid the model in identifying irregular behavior by inputting specific dates to be treated as holidays. It is especially applicable in scenarios subject to strong seasonality and holiday-related variations. The authors state that the method has shown lower mean absolute percentage error than widely used models such as ARIMA and exponential smoothing.

Figure 4 - Diagram of Analyst-in-the-loop user-machine collaboration.

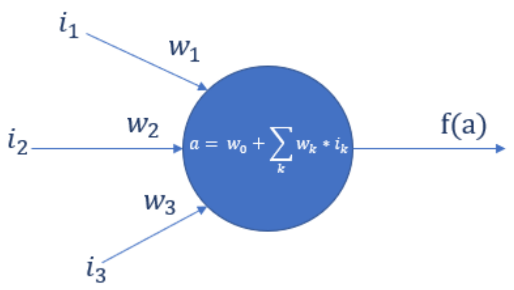



Source: Taylor, S. J. & Letham, B. (2018)

A special class of demand forecasting models has received much attention recently: Machine Learning models (ZHANG, 2001; GILLILAND, 2020). Proposing an alternative to many traditional approaches that require assumptions to be made about the probability distribution that best describes the demand, these models discharge the analyst of the responsibility of building a solution around hypotheses that are unknown and often imprecise. Among these models, neural networks stand out for their ability to find hidden patterns in the data and provide predictions based on insight learned by the model itself.

Multi-Layer Perceptrons (MLP) (ROSENBLATT, 1958) are the most fundamental neural networks. They are models consisting of connected elements called neurons that combine data through a function, called the activation function, and feed other elements of the network. A simple neuron, also called a Perceptron, consists of an input, the input's coefficient, an activation function, and an output (Figure 5). The neuron accepts multiple inputs by multiplying each value by its own coefficient, also called "weight", and adding them to a constant value. The resulting value is then transformed by the activation function, which will produce the output.

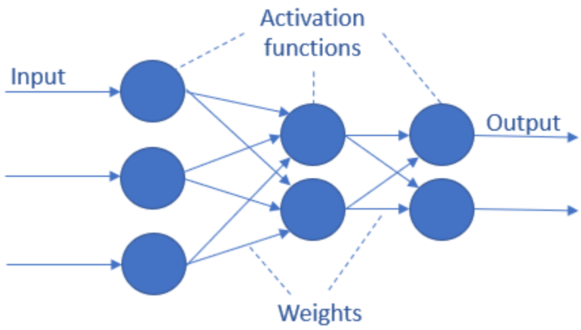
Figure 5 - A Perceptron.




Source: The author (2021)

A network is formed when a collection of neurons are connected, and one neuron's output becomes the other one's input. Networks are organized in layers of neurons. The first one is the input layer, which deals directly with inputted data. The final layer is the one that produces the model's outputs. The number of neurons in this layer should correspond to the desired size of the output. Any layer in between is called a hidden layer, and they each receive as input the outputs from the previous layer and pass on their outputs to the following layer. In fully connected layers, also called dense, the input of a neuron is the weighted sum of all outputs from the previous layer. Figure 6 illustrates a fully connected MLP.

Figure 6 - Example of Multi-Layer Perceptron.




Source: The author (2021)

A network's architecture is a description of its fundamental traits such as the number of layers, the number of neurons in each layer, and which connections between layers are made. Defining the architecture is one of the first steps in building an MLP model. The sizes of the input and of the output are key to defining the number of neurons in the input and output layers. The activation functions should also be determined and will largely depend on the desired output and learning task. MLPs can be trained for a regression or classification task. In either case, the output, architecture, and activation function should be chosen accordingly. Examples of activation functions are described in Table 2.

The process of training a network means adjusting its weights in order to obtain outputs as close as possible to the target values. All training is based on a database containing examples of sets of inputs and their corresponding target values. After a random initialization of the weights of the network, a Loss Function is calculated, which describes the difference to be minimized between the model's outputs and their corresponding target values from the database. Examples of loss functions can be seen in Table 3.

Table 2 – Activation functions

| Name | Formula | Observation |
|---|---|---|
| Sigmoid | $\frac{1}{1+e^{-x}}$ | Especially useful in predicting a probability because output ranges between zero and one. |
| Softmax | $\frac{e^{z_j}}{\sum_K e^{z_k}}$ | Used in multiclass classification |
| Hyperbolic Tangent (tanh) | tanh(x) | Values range between -1 and 1. More frequently used in classification tasks |
| Rectified Linear Unit (ReLU) | max (0,x) | Widely used in regression tasks |

Source: The author (2021)

Table 3 – Loss Functions

| Name | Formula | Task |
|---|---|---|
| Mean Squared Error (MSE) | $\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$ N= number of samples | Regression |
| Mean Squared Logarithmic Error (MSLE) | $\frac{1}{N}\sum_{i=1}^{N}(\log(y_i + 1) - \log(\hat{y}_i + 1))^2$ | Regression |
| Cross-Entropy | $-\sum_{i=1}^{k} y_i \log(\hat{y}_i)$ k= number of classes | Classification |

Source: The author (2021)

Due to the complexity of the problem, a computational method is used in order to minimize the loss function. The use of variants of Gradient Descent is very common. These methods calculate the current solution's partial derivatives, and adjust the variables one "step" towards the opposite direction of the derivative, usually decreasing the value of the Loss function. This process is repeated until reaching a stopping criterion, such as a local minimum or a maximum number of iterations. The partial derivatives of the loss function with respect to the weights are calculated through a process called Backpropagation.

A widely used optimization method is Stochastic Gradient Descent (SGD) (ROBBINS & MONRO, 1951), in which the parameter update is done not after gradients are calculated for the entire dataset, but for each point in a randomly chosen order. Other methods propose changes to how the size and direction of the “step” are defined. Adaptive Moment Estimation (Adam) (KINGMA & BA, 2015), for example, is a method in which the “step” size is a function not only of a pre-defined parameter but also of the exponentially decaying averages of past gradients and squared gradients. This avoids getting stuck on local minima and especially saddle points.

Quasi-Newton methods are also good optimizing alternatives, as they simplify Newton’s method by approximating the Hessian matrix instead of calculating it. A widely applied method is the Broyden–Fletcher–Goldfarb–Shanno (BFGS) (BROYDEN, 1970; FLETCHER, 1970; GOLDFARB, 1970; SHANNO, 1970). In solving large problems, a simplified version of the method, called Limited-memory BFGS (LBFGS) (LIU & NOCEDAL, 1989), is often used.

Once the weights are defined by the optimization method, the model is trained and ready to be applied in predicting labels to new inputs. In order to test the ability of the model to make new predictions, a portion of the data is set aside from the training set and the model’s accuracy is assessed by comparing the test set’s target outputs to the obtained predictions. When multiple parameters to a model need to be compared, a validation set is also set aside from the training process and is used to compare accuracies from each combination of parameters. The final model’s accuracy is then measured on the test set.

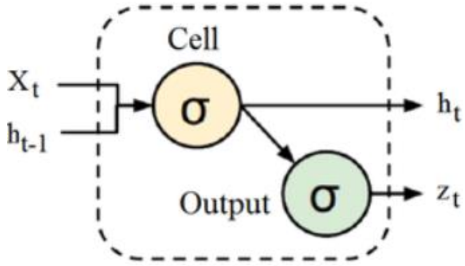
With vast applicability due to their flexibility in design and architecture, MLPs are powerful investigators of possible combinations and frequently helpful in recognizing complex patterns. In time series forecasting, MLPs are useful not only in making predictions but also in combining predictions from different models into hybrid models, making use of strengths from different techniques (OLIVEIRA, 2020).

More recently, some of these models have developed into more powerful ones, capable of learning very complex patterns. These are known as Deep Neural Networks (DNN) (LE CUN *et al.*, 2015). Two of them stand out, due to their architecture designed especially for sequential data, as is the case with time-indexed data: RNN and LSTM.

Recurrent Neural Networks (RNN) (RUMELHART *et al.*, 1986) are a variant of neural networks formed by a sequence of neurons that take as input not only the given variables but also a hidden unit passed on by the previous one. It can receive multiple inputs in a given sequence and produce multiple outputs or a single one, according to the desired output format.

These configurations are called “many-to-many” and “many-to-one”, respectively. It is also possible to obtain results from a single input, in which case one can have “one-to-many” or “one-to-one”. Figure 7 depicts a cell that makes up an RNN, in which X is the input, h is the hidden unit, z is the output and t is the index of the cell in the sequence. Due to their configuration, RNNs can identify patterns related to the sequence of the variables.

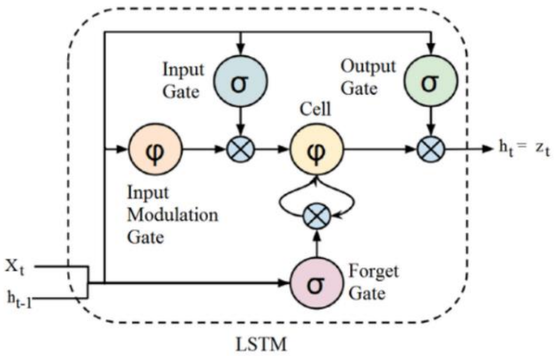
Figure 7 - Illustration of an RNN cell




Source: Fayyaz, M. *et al.* (2016)

RNNs only receive hidden unit information from the previous cell, causing it to dismiss information from previous time-steps. In order to fill that gap, Long Short-Term Memory (LSTM) networks (HOCHREITER & SCHMIDHUBER, 1997) follow the same sequential pattern of an RNN but were designed to hold on longer to previous information by adding functions, called gates, that interact to decide at each time-step how much of previous information is kept in the hidden unit, and how much is updated by the current time-step information. Figure 8 describes in further detail the information flow in a single LSTM cell.

Figure 8 - Illustration of an LSTM cell



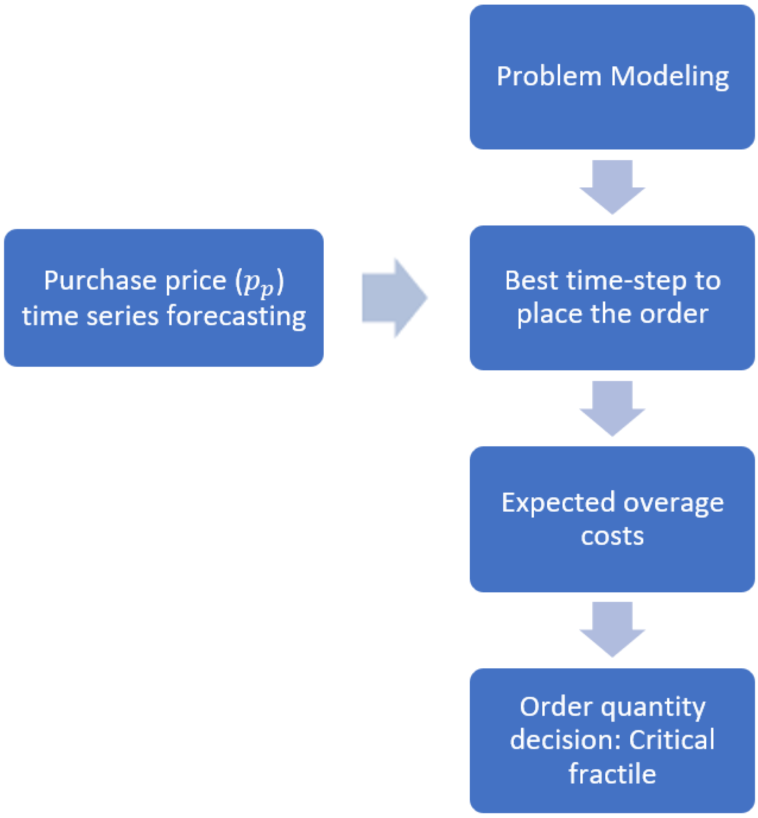

Source: Fayyaz, M. *et al.* (2016)

Each method applies different mathematical reasoning to model the series and try to predict its future behavior. By adopting all of them as potential candidates for the proposed methodology, it is possible to compare performances and observe if any particular method was better capable of capturing the specific pattern of the time series. This will be part of the methodology described in Section 3.

# 3 PROPOSED METHODOLOGY

In order to develop a methodology for determining ideal order placement timing and quantity in newsvendor situations subject to purchase price fluctuations, the classical newsvendor problem is first adapted. The optimization problem uses purchase price estimations in order to determine the optimal timing in which to place the purchase, as well as the order quantity. Firstly, the price estimations are obtained by autoregressive models. This time series forecasting based on historical data is performed using SARIMA, Prophet, MLP, RNN, and LSTM. These values are then used as the overage cost in the optimization model to obtain the optimal order quantity and timing. Figure 9 illustrates the proposed methodology.

Figure 9 - Methodology flowchart




Source: The author (2021)

## 3.1 NEWSVENDOR SITUATION

The context of this problem is drawn around retailers that must refresh their perishable products every cycle of *s* fixed time steps. At the beginning of each cycle, they must dispose of the previous inventory and begin selling the new one. This respects all three criteria for newsvendor-type problems: i) the retailers are subject to an unknown random demand.; ii) a single order is placed for each time period, with the additional factor that they must decide at which time step of the initiating cycle they will place the order for the following one; iii) overage costs are a function of the purchase price and underage costs are a function of the selling price.

The problem is inspired by a Brazilian supermarket chain that makes weekly purchases of perishable items. The prices of most of these items are subject to fluctuations due to environmental and economic factors such as droughts, pests, international trade deals, among others. The retailer seeks to make the best decision regarding when and how much to purchase, given this scenario. This work proposes the use of data-based purchase price prediction methods as input to the decision-making process. In order to test the quality of the replenishment decisions obtained by the proposed methods, real historical prices are used. In this case, meat was chosen as the item for analysis due to its adherence to model characteristics: stable demand, fluctuating purchase prices based on exogenous factors, and high perishability.

It is reasonable to assume a Normal distribution as the probabilistic demand model. According to RAMAEKERS & JANSSENS (2008), Normal and Gamma distributions have proven to fit well the demand for fast-moving items in retail. Also, assuming a Normal distribution has been shown to perform at a similar level of neural network demand prediction techniques applied to the newsvendor problem, with reduced solution complexity (HUBER *et al.*, 2019).

In order to maximize its profit, the retailer must decide on two decision variables: the order size y and purchase time t. Five weekly discrete time steps are considered in each decision horizon (s=5), representing the weekdays available for the procurement department to place the order for the following week’s material.

## 3.2 OPTIMIZATION MODEL

The mathematical model was adapted to the proposed newsvendor variant by first analyzing the objective function. When subject to fixed purchase prices, the unit purchase time (t) is not considered in the objective function (Equation 2.1) because it does not affect the parameters involved. In the case of a stochastic purchase price ($p_p$), however, it directly impacts the overage costs ($c_o$), and overall cost now becomes a function of time. Underage costs are considered to be a function of the selling price ($p_s$), and will be further discussed. The time-dependency of the overage costs is described in the revised objective function (Equation 3.1).

$$\min_{y,t} C(y,t) = E\big[\, c_u(p_s) * \max\{0, D-y\} + c_o(p_p(\mathrm{t})) * \max\{0, y-D\}\big] \tag{3.1}$$

This overage cost is traditionally translated as the difference between the unit purchase price and the unit salvage value (HILL, 2017). In the context of the supermarket, no salvage revenue was considered, due to the health hazard of expired food products. Also, disposal of the goods was regarded as a fixed cost, since there is already available infrastructure and discarding scrap material takes a negligible portion of the employee's time. Therefore, the unit overage cost is solely described by the unit purchase price.

The underage cost is often described as the unit profit (SWAMIDASS, 2000). Considering a food retail profit margin of 2%, the underage cost would be 2% of the selling price ($p_s$). However, PERONA *et al.* (2001) argue that not only lost profits but also opportunity costs should be considered in a lost sale. They show that, although estimated lost sale costs for downstream supply chain members are generally regarded as the lost contribution margin alone and lie below 2%, actual costs are about 11% of potential revenue due to demand absorption by the competition. Plugging this value into (Equation 3.1) for a fixed decision horizon, the expected total cost is obtained as follows, in which $p_p$ is the corresponding purchase price on the fixed time-step (Equation 3.2):

$$\min_{y,t} C(y,t) = E\big[\, (0.11 * p_s * \max\{0, D-y\}) + (p_p(\mathrm{t}) * \max\{0, y-D\})\big] \tag{3.2}$$

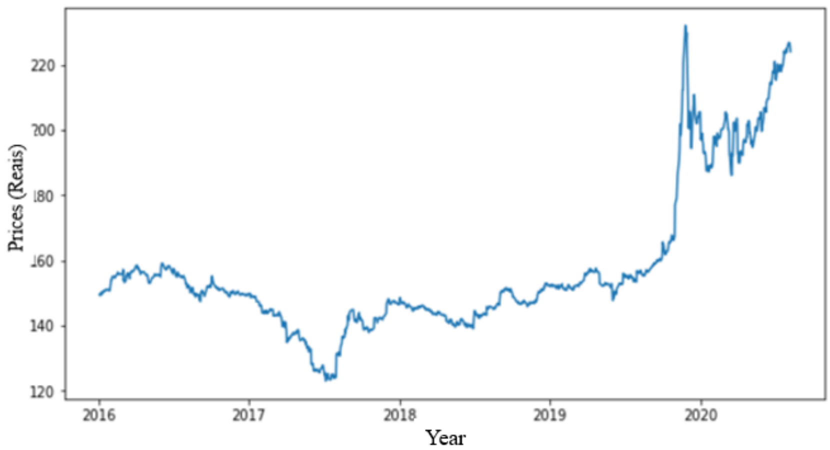
The objective function is monotonically non-decreasing with respect to $p_p$, since $\max\{0, y-D\}$ is strictly non-negative, and the underage term is fixed for a single decision horizon. Therefore, the t that minimizes C is the one that minimizes $p_p$. The behavior of the stochastic variable will be analyzed and estimated based on the available historical values, and several methods will be compared in order to achieve the minimal $p_p$.

## 3.3 DATASET DESCRIPTION

For analyzing and deciding upon the purchase prices of meat, one must hold historical data of the variable's daily behavior. Due to sparse data held by the supermarket that inspired this research with respect to daily purchase price historical behavior – the only data points available are those of dates in which there was, in fact, a purchase – it is necessary to use an additional time series with high correlation and daily availability. Meat price fluctuations in Brazil are measured by the National Price Index to the Broad Consumer (IPCA, from Portuguese: Índice Nacional de Preços ao Consumidor Amplo) (IBGE, 2020). This would be an ideal reference for the actual price that is being analyzed but it is only available on a monthly basis. A daily-indexed database showing a high correlation with this series would then be the desirable basis for this work.

Multiple daily-indexed series were tested for correlation with the IPCA index, such as the Live Cattle Futures prices in dollars, and the Feeder Cattle Futures prices in dollars (FML, 2020). Other series could not be used for daily forecasting purposes, such as Food Price indexes from the United Nation's Food and Agriculture Organization, for they were only available on a monthly basis. A Pearson correlation of 0.896 was found with the monthly average of the Live Cattle Bovespa Futures prices (FML, 2020), from the Brazilian stock market. Based on this high correlation, the Live Cattle Futures prices variations were used to estimate daily meat prices. Publicly available daily values for this index from Jan 4, 2016, to Aug 7, 2020, were extracted from an open database and used to model the series and estimate prices (FML, 2020). The most recent available data was chosen, dating back to over four complete year cycles to allow for enough information on the behavior of the series. The resulting time series is shown below in Figure 10.

Figure 10 - Timeline of Live Cattle BOVESPA stock prices




Source: The author (2021)

## 3.4 PREPROCESSING

The original dataset consists of a table containing multiple information on the Live Cattle Bovespa Futures price series and is read as a data frame in Python's Pandas library (MCKINNEY, 2010) (Table 4). The daily market closing prices were used as the main variable of interest. The date column is used as the index for the research's database. Daily highs, lows, and opening prices, as well as traded volume and percentage price variation, are also available but are not considered in the context of this work.

Table 4 – Live Cattle Bovespa Futures price series

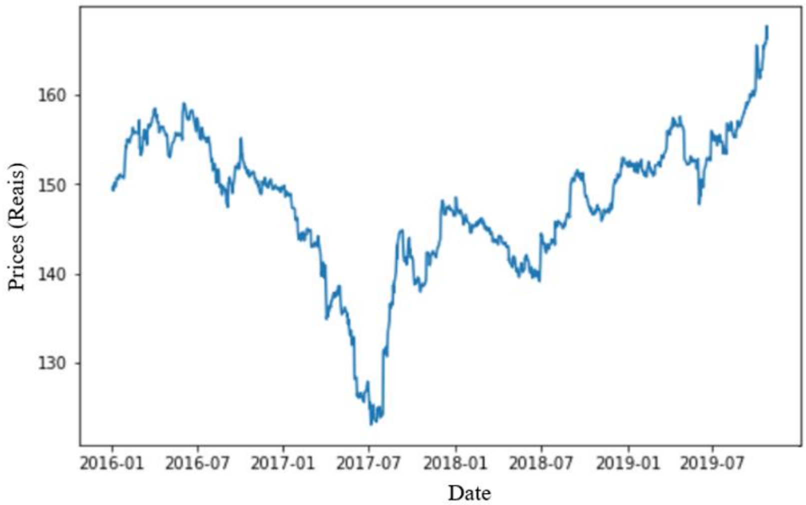
| | Date | Price | Open | High | Low | Vol. | Change % |
|---|---|---|---|---|---|---|---|
| 0 | Aug 07, 2020 | 224.05 | 226.00 | 226.00 | 224.05 | 0.35K | -0.73% |
| 1 | Aug 06, 2020 | 225.70 | 224.50 | 226.80 | 224.50 | 0.08K | -0.42% |
| 2 | Aug 05, 2020 | 226.65 | 226.70 | 226.95 | 225.90 | 0.52K | -0.02% |
| 3 | Aug 04, 2020 | 226.70 | 227.10 | 227.10 | 226.70 | 0.00K | -0.04% |
| 4 | Aug 03, 2020 | 226.80 | 226.95 | 227.40 | 226.30 | 0.01K | 0.53% |

Source: The author (2021)

Brazilian holidays are not seen in the database because the stock market is closed on those dates. Therefore, the data frame had missing entries, and accordingly, completion was necessary in order to contemplate every weekday in this time frame. Missing days were considered to have the same value as the previous day since the market was not open to trade.

By observing the graph in Figure 10, a significant change in series behavior can be seen in November of 2019. According to specialists (GESSULLI AGRIBUSINESS, 2020), this increase was due to a high rise in meat demand from China because of the swine fever outbreak at that time. After the record high, the series continued to show an unstable pattern due to an added effect of COVID-19 in 2020. When compared to the historical series from 2016 to 2019, all values since November 2019 could be considered outliers according to Tukey's fences (TUKEY, 1977). This means that values are over 1.5 times the inter-quartile range above the third quartile of historical data until then. Although the series may eventually stabilize at a new level or permanently assume this exceptional pattern, it is still not possible to infer future behavior from 2020 values due to such atypical factors such as COVID-19 and the swine fever. Also, given the length of the data and general forecasting methodology, assessing both patterns together would mean that the first and longer pattern would be used to train the data and the altered pattern would approximately correspond to a test set. This would compromise the models' performance assessment. Therefore, for the sake of extracting insight for this study, all data since the November 2019 peak were excluded from the observations. The data was kept until the last week that did not contain alterations related to this November 2019 peak but remained within the Tukey's fences criterion. The resulting series contains 995 points, or 199 weeks, and ranges from Jan 4, 2016, to October 25, 2019. The series values are shown in Figure 11.

Figure 11 - Live Cattle BOVESPA stock prices excluding exceptional pattern



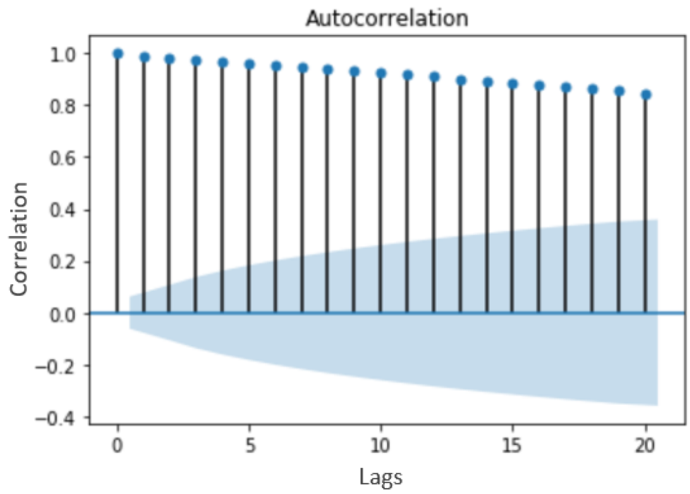

Source: The author (2021)

Before applying the predictive models, the values from the series were scaled using Python's MinMaxScaler from Scikit-Learn library's preprocessing package ("sklearn.preprocessing") (PEDREGOSA *et al.*, 2011). This scaler takes each value, subtracts the minimum value from the database in which it was fit, and divides the result by the range of the fitting database, which is the difference between its maximum and minimum values. The scaler was fit on the training data and used to transform both training and test data.

## 3.5 AUTOCORRELATION AND PARTIAL AUTOCORRELATION FUNCTIONS

An important aspect of the time series which can be used to provide insight into its pattern is its autocorrelation (COWPERTWAIT, 2009). The Autocorrelation Function (ACF) describes how each point in the series is related to its predecessors by comparing the series to its delayed copy. The size of the delay is called the lag to which the series is compared. A commonly used way of representing the correlation is the ACF graph, which provides the autocorrelation for each lag. The ACF graph for the Live Cattle time series, up to 20 lags, is provided in Figure 12. Since each week is represented by 5 weekdays, 20 lags would be 4 weeks, close to a month.

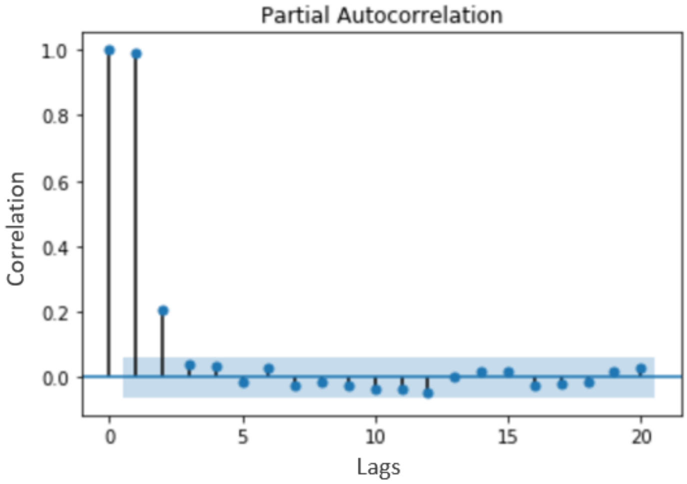
Figure 12 – Time series autocorrelation




Source: The author (2021)

The shaded region indicates the significance region, beyond which the autocorrelation is considered different from zero. The series shows high autocorrelation values for all lags up to 20, with a slight decrease at each lag. However, a function that better describes the individual influence of each lag is the Partial Autocorrelation Function (PACF). The PACF indicates the correlation of the series with each lag, once removed the linear dependence of the series with lower lags. That is to say that redundant information is removed, and the PACF indicates which lags still provide information on the pattern of the series once lower lags are already known. The PACF graph is shown below (Figure 13) and indicates that only lags 1 and 2 are relevant beyond the significance region and using greater lags is unlikely to provide more insight into future values of the series.

Figure 13 – Time series partial autocorrelation




Source: The author (2021)

The ACF and PACF were obtained through the "plot_acf" and "plot_pacf" functions from the statsmodels.graphics.tsaplots package in Python (SEABOLD *et al.*, 2010).

## 3.6 FORECASTING MODELS

This section details how the selected forecasting models were applied including the libraries and parameters that were used.

### 3.6.1 SARIMA

The first forecasting model applied to the data was the SARIMA, which was implemented using the pmdarima library (SMITH *et al.*, 2018) in Python, created to replicate the widely used auto.arima library available in R. Considering that the procurement team must decide on what day will be held the purchase before the week begins, the model considered a five-step-ahead prediction made at the end of every week. The first seventy-five percent of the dataset was used to train the model, and the other twenty-five percent of samples were used as a test set. The function takes as input, at first, the complete training set, and uses it to predict the first 5 time-steps (one week) of the test set. Then, the first week of the test data is added to the training set and is used to retrain the data to predict the second week. This is iteratively repeated as each week is added to the training data to predict the following week, until there are no more weeks left to be predicted. That is to say, the parameters of the ARIMA were recalibrated at the end of every week before each five-step-ahead prediction was made.

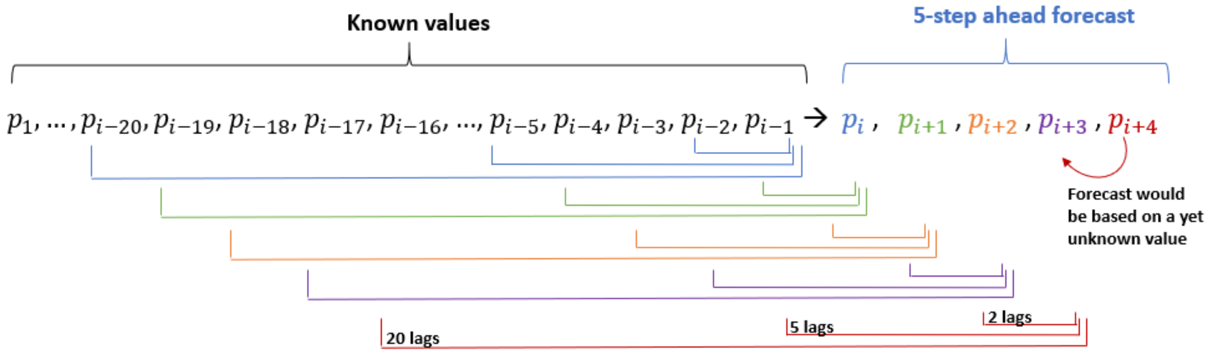
### 3.6.2 Prophet

The second approach used was the Prophet model (TAYLOR & LETHAM, 2018). The open-source library made available in Python by its authors was used ("fbprophet" package). As input, it uses the date-indexed time series, as it uses pre-coded date information such as day of the week or holiday in its predictions. In order to obtain the prediction for the following week, it must make a 7-step-ahead prediction, since the data always ends on a Friday, and must predict up to the following Friday. Therefore, the frequency is set to 'D', which stands for daily, and the period input is set to 7. However, only Monday through Friday predictions are saved. And the model is iteratively fit and called on to predict each week. Similarly to the previous model, a 75-25 train-test split was used for this model, with the earliest 145 consecutive weeks in the training set and the latest 50 weeks in the test set.

### 3.6.3 Multi-Layer Perceptron

The third model that was applied was an MLP that took as input the previous values for the time series. In order to choose which previous values should influence the time-steps that need to be predicted, an autocorrelation analysis was conducted. As shown in the PACF graph, only lags 1 and 2 offer relevant information on the values of the series and should therefore be used in the prediction. However, in order to compare results and test possibilities, three different MLP inputs were tested: with lags up to 2, up to 5, and up to 20. That is, the lags suggested by the PACF, the entire previous week, and the 4 previous weeks (nearly a month of information), respectively. An interesting observation is that, since the context of the problem involves a 5-step-ahead prediction, not all lags are yet known for all the points that are to be predicted (Figure 14). For example, if the prediction is made at the end of the week, lags 1 through 4 are unknown for the following Friday. However, the ACF and PACF graphs show there is redundant information between the first two lags and all the others, up to lag 20. Since all lags showed high ACF values, but not PACF values, higher lags should be able to offer the same autocorrelation information that would be offered by lower lags, if they were known.

Figure 14 – Some lags for future time-steps are yet unknown



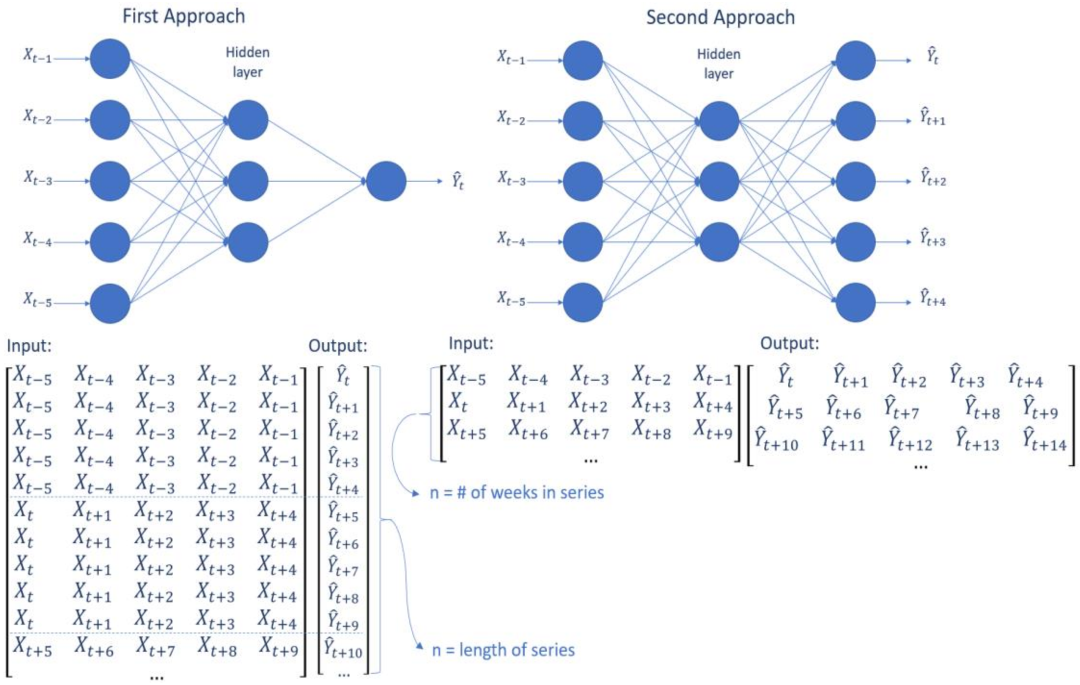

Source: The author (2021)

Two approaches were considered in order to obtain the 5-step-ahead model (Figure 15). The first one was considering the output to be a one-dimensional vector with the size of the test set, that is, representing each day. And the input associated with each output would be the corresponding lags from the time series. The shape of the input, in this case, is the size of the test set as the number of rows and the selected number of lags as the number of columns.

Since the decision is always made at the end of each week for the following one, all lags are known for Monday ($Y_t$), which is the first day of the week, and therefore there is available information on all days that precede it. However, for Tuesday's prediction ($Y_{t+1}$), lag 1 would be Monday, whose value is still unknown. Therefore, the closest known lags are used, that is, the last days of the previous week. This will correspond to the same values used for Monday's prediction. Because there is no updated information throughout the week – the forecast is made for an entire week ahead – all days of the week will be predicted based on the same closest available lags. When predicting based on lags 1 and 2, all days of the week will use the past Friday and Thursday as input. The same logic is repeated for all desired numbers of lags. The most recent 2, 5, or 20 points available to be used as input are the same for each Monday through Friday. Using the exact same input for all 5 days of the week resulted in the same prediction for every day of a single week. That is, in the context of aiding the optimization problem in deciding which day of the week to make the purchase, this model was not useful.

A second approach was then considered by obtaining a single output as an array of length 5 for each week, representing its 5 days. Although a single set of lags was used to predict the entire array, the trained model would know that prices vary throughout the week and offered different values for each day. Therefore, the output is now a matrix with 5 columns and the number of weeks in the test set as the number of lines. The input is also a matrix that still has the number of lags as the number of columns, but now the number of rows represents the number of weeks in the test set.

Figure 15 – Some lags for future time-steps are yet unknown




Source: The author (2021)

An MLP with a single hidden layer was created using the MLPRegressor from Scikit-Learn’s “neural_network” package. Multiple combinations of parameters were tested on the validation set, with variations in the number of neurons, activation function of the hidden layer, and optimizer. Table 5 contains the different values that were tested for each of these parameters. All possible combinations were tested.

Table 5 – MLP Parameters

| Parameter | Tested values |
|---|---|
| Number of neurons in the hidden layer | {1, 10, 20, 50, 100} |
| Activation function | {tanh, relu} |
| Optimization method | {Adam, SGD, LBFGS} |

Source: The author (2021)

The model was validated to choose the best parameters. The validation model was taken from what would be the training set. Instead of the 75/25 split applied to previous models, a 50/25/25 split was made. The training set consisted of 99 weeks, the validation set in 50 weeks, and the test set in the remaining 50 weeks.

Each model was trained 100 times on the training set for 1000 epochs and tested on the validation set. For each input shape (number of lags), the model with the best average Mean Squared Error was selected. Training and validation sets were then concatenated in order to retrain the model with the best performing parameters. The model was then tested on the test set, from which the performance metrics were obtained.

### 3.6.4 Recurrent Neural Networks

The fourth and fifth models are two kinds of recurrent neural networks: RNNs and LSTMs. Using the same logic as the second approach to the MLP, the inputs are up to 2, 5, or 20 lags for each week, and the output is the prediction for each day of the week. The networks are implemented using the Keras library in Python (CHOLLET, 2015). A sequential model is created using Keras and an RNN or LSTM layer is added, followed by a dense layer with 5 cells, in order to produce the 5-steps-ahead prediction. Varying sizes for the RNN or LSTM layers were tested. The values are listed in Table 6. Mean squared error (MSE) is used as the loss function to compile the Sequential Model, and three optimization methods are tested, as seen in Table 6. Models with each possible combination of parameters were trained 30 times on the training set for 1000 epochs and tested on the validation set. For each input shape (number of lags), the model with the best average Mean Squared Error was selected. Once the best choices of parameters were determined, the champion model was retrained on training and validation sets combined. The performance of the models was measured on the test set.

Table 6 – RNN and LSTM Parameters

| Parameter | Tested values |
|---|---|
| Number of neurons in the RNN or LSTM layers | {1, 10, 20, 50, 100} |
| Optimization method | {Adam, SGD, LBFGS} |

Source: The author (2021)

### 3.6.5 Performance metrics

In order to compare the forecasting models, two metrics were used. The first one was the Mean Squared Error of the test set predictions versus target values. In the case of stochastic models, such as the neural networks that are subject to random weight initialization, the average Mean Squared Error of 100 training cycles was used.

The second performance metric is related to the retailer context. In reality, the best prediction model, in this case, is the one that supports the best decision policy, generating the least overall costs. Therefore, the suggested metric describes the average weekly cost dispensed by the retailer if every week they purchase on the day in which is predicted to have the lowest value of the week, according to the model used. One of the baselines for this metric is when the retailer knows the best day to purchase before the week starts, which we call the Oracle strategy. It is an ideal and non-realistic scenario, but it describes the lower bound for the cost and best-case scenario. The second baseline used is the zero-information scenario in which the retailer selects a random day every week to purchase. Since each day of the week has an equal chance of being selected, the estimated cost of that policy a given week is the average of its days. The average weekly cost of that policy is the average cost of all days in the test set.

These two metrics are used to compare the forecasting models in their ability to provide insight into the best day to make the purchase, that is, the best time t. However, the joint (y,t) decision is also evaluated after the forecasts are inputted into the optimization problem.

## 3.7 DECISION-MAKING

Once the lowest purchase price has been identified for a certain week, t (the variable representing the decision of when to place the order) is set to the day it is predicted to occur, which will implicate in the lowest expected value for $c_o(\mathrm{t})$, offering the greatest chance to make the choice that will contribute less to the objective function that should be minimized.

The classic newsvendor solution can then determine the order amount (y), that is, choosing the value of y so that the probability of the demand being fully served by the purchased amount be equal to the critical fractile (Equation 2.6). In this context, the critical fractile can be rewritten as Equation 3.3, with $F^{-1}$ as the inverse of the cumulative distribution function that governs the stochastic demand.

.

$$y^* = F^{-1}\left(\frac{0.11 * p_s}{p_p + 0.11 * p_s}\right) \tag{3.3}$$

In order to evaluate the quality of the joint decision, random demands will be generated 10,000 times for the test set time horizon of 50 weeks. The demands will follow a Normal probability distribution with a mean of 1,000, and three levels of standard deviation will be tested (50, 150, and 300). These standard deviation values were defined as a grid to evaluate performance when the problem is subject to multiple levels of uncertainty. The value of 1,000 was chosen as a reference value.

Real historical data from the Live Cattle times series will be used for the purchase prices. The selling price ($p_s$) parameter is determined by market strategy, taking into account the average prices practiced by the competition. Its value will be adjusted every 4 weeks based on the IPCA.

Total overage and underage costs for the 50 weeks (as an average of the 10,000 replications) is the performance metric used to compare the proposed joint-decision strategies obtained from each prediction model and Equation 3.3. The baseline against which to decide if the strategy is useful is the zero-information situation. In this scenario, the retailer randomly chooses a day in which to place the order and consistently uses the mean demand (1,000 units) as the order quantity.

The 10,000 total cost values from each strategy are compared using the Kruskal-Wallis non-parametric statistic test (KRUSKAL & WALLIS, 1952). The test is performed using the “kruskal” function from the Scipy library’s “stats” package (VIRTANEN, 2020). The test indicates whether the different groups of values originate from the same distribution, without making any assumption about the probability distribution from which they originate. In this context, it tells if results from the different strategies are significantly different or if they can be considered to follow roughly the same pattern. For a significance value of α=0.05, the null hypothesis that the groups share the same population median will be rejected if the test’s resulting p-value is greater than 0.05.

# 4 RESULTS

This section presents the results obtained from the application of the forecasting models as well as from the simulation of their usage in the optimization process.

## 4.1 PURCHASE PRICE ESTIMATION

Table 7 summarizes MSE and cost performances for each forecasting model and compares them to the proposed oracle and zero-information (Random day) baselines, as described in Section 3.5.5. The "Additional Cost" column refers to the percent increase in average weekly unit costs of meat if compared to the perfect information scenario, that is, the Actual best day (Oracle).

Table 7 - Quality of Purchase Price Predictions

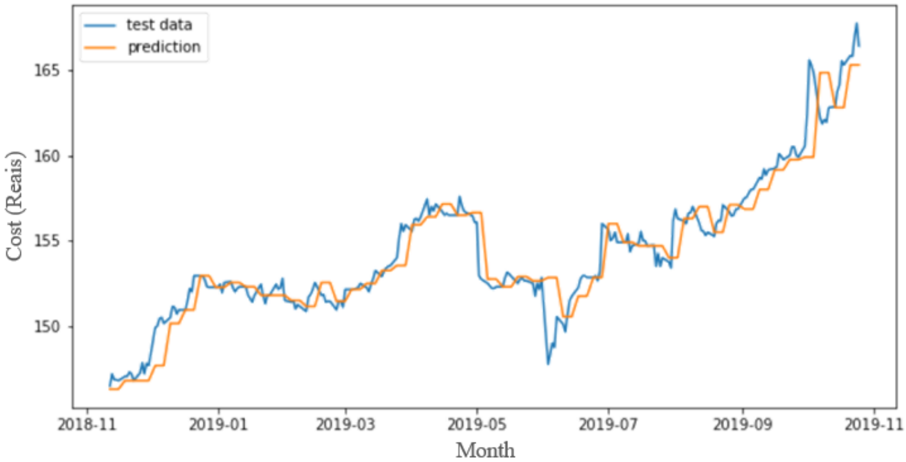
| **Purchase day policy** | ***Average Unit Cost*** | ***Additional Cost*** | ***Cost Performance*** | ***MSE*** | ***MSE Performance*** |
|---|---|---|---|---|---|
| **Actual best day (Oracle)** | **153.730400** | **0.0000%** | **1** | **N/A** | **N/A** |
| RNN (up to lag 20) | 154.230400 | **0.3252%** | **2** | 0.00160994 | **3** |
| MLP (up to lag 20) | 154.322000 | **0.3848%** | **3** | 0.00177662 | **8** |
| **Random day (Historical average)** | **154.343640** | **0.3989%** | **4** | **N/A** | **N/A** |
| SARIMA | 154.343640 | **0.3989%** | **4** | 0.00134465 | **1** |
| RNN (up to lag 5) | 154.348600 | **0.4021%** | **6** | 0.00161587 | **4** |
| LSTM (up to lag 5) | 154.357000 | **0.4076%** | **7** | 0.00200494 | **10** |
| Prophet | 154.357200 | **0.4077%** | **8** | 0.00515356 | **11** |
| LSTM (up to lag 2) | 154.394800 | **0.4322%** | **9** | 0.00158377 | **2** |
| MLP (up to lag 5) | 154.411800 | **0.4432%** | **10** | 0.00172827 | **7** |
| MLP (up to lag 2) | 154.417400 | **0.4469%** | **11** | 0.00164193 | **6** |
| LSTM (up to lag 20) | 154.441200 | **0.4624%** | **12** | 0.00196408 | **9** |
| RNN (up to lag 2) | 154.552800 | **0.5350%** | **13** | 0.00161636 | **5** |

Source: The author (2021)

The SARIMA model (Figure 16) turned out quite simple and could not identify clear patterns in the apparently complex time series. It did not capture any trend or seasonality factors in the series, and the best fit order was, in fact, an ARIMA (0,1,0). This suggests the series is close to a random walk, which means that at each time-step, the random variable moves slightly up or down, with no clear pattern, and the best guess of what the next time-step value will be is the last previously known value. This simple model showed the best MSE outcomes out of all that were tested: 0.00134465. Entering the original data into the model, without previously scaling the data, was also tested but resulted in a similar MSE.

However, since the model resulted in a random walk, a 5-step-ahead prediction outputted 5 equal values for all days of the week. This was found to be true for all 50 weeks of the test set. Therefore, in terms of selecting the best day to place an order, the model was no better than the original scenario of not having any information. Therefore, in terms of average weekly cost performance, it was considered to have the same value as the random choice baseline.

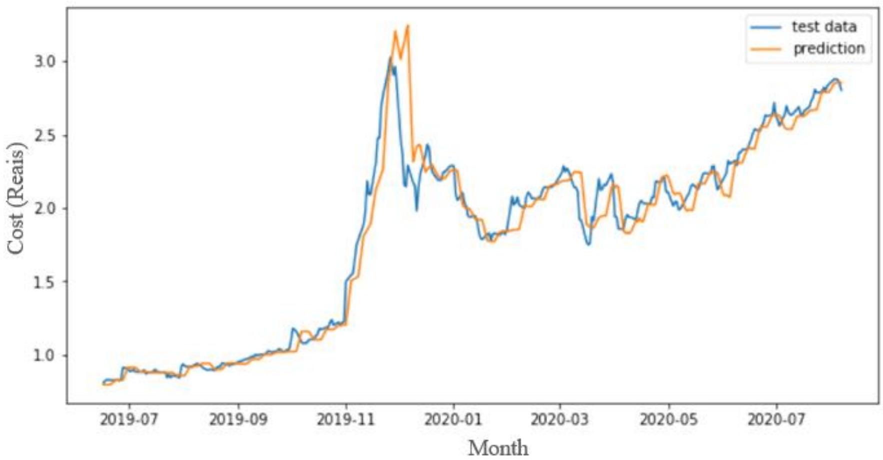
Figure 16 - SARIMA purchase price prediction.




Source: The author (2021)

SARIMA is usually very good at identifying patterns. When the full series, before removal of the pandemic pattern, was modeled, it obtained an ARIMA (1, 1, 4), and the 5-step-ahead forecasting of the test set offered different values for the different days of the week (Figure 17).

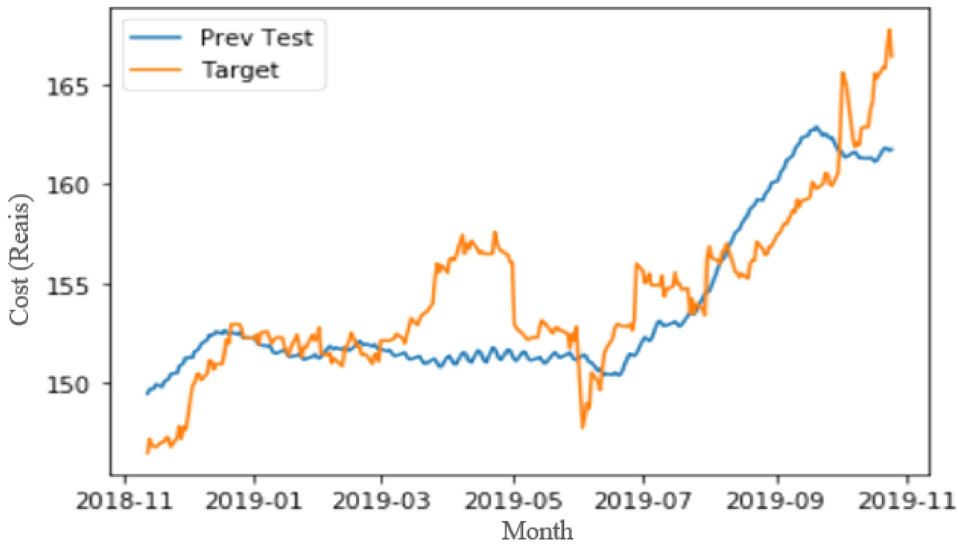
Figure 17 - SARIMA purchase price prediction (including new pattern).




Source: The author (2021)

On the other hand, the Prophet model offered different values for each day of the week and therefore could offer insight into the decision of when to buy. However, its forecasts were the least accurate of all. Although the model includes trend and seasonality parameters, such as the SARIMA, it was not able to capture a few significant trends around the first semester of 2019. The model had an MSE of 0.00515356. Figure 18 compares Prophet's predictions for the test set ("Prev Test") to the target series.

Figure 18 –Prophet purchase price prediction

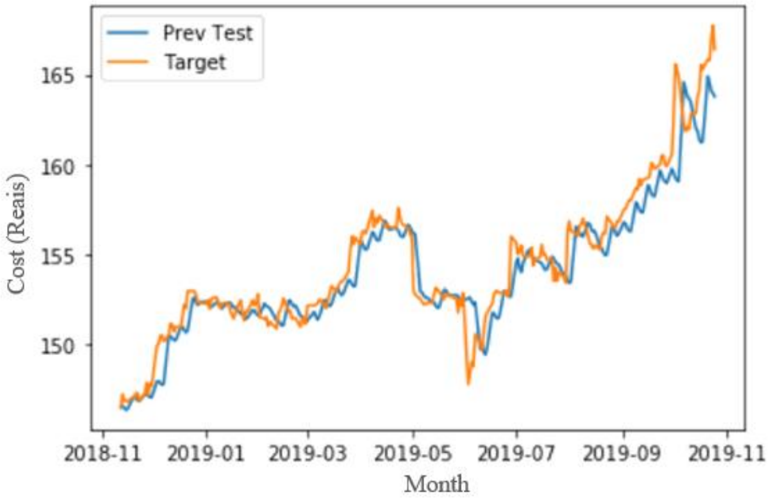



Source: The author (2021)

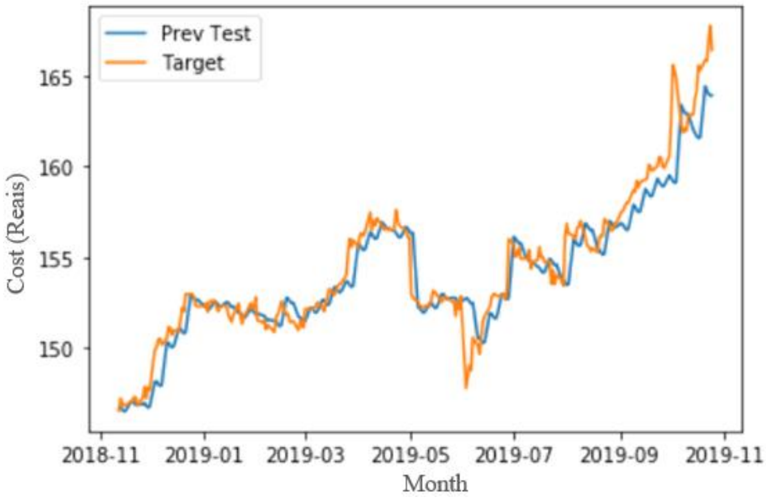
The first approach considered for the MLP was not an interesting model to support the decision-making since a daily prediction with all points belonging to the same week sharing the same features resulted in the same predicted value for the entire week, as described in Section 3.5.3. Therefore, the second approach was chosen to be investigated in this work. Different performances were found for each tested maximum number of lags (2, 5, and 20). The three models had similar MSE results, turning out in adjacent places on the MSE ranking (6$^{th}$, 7$^{th}$, and 8$^{th}$ places, respectively). However, in terms of being used as a basis for the order day decision, the MLP using up to 20 lags performed very differently from the others and was one of the only two models that outperformed the random day strategy. The MLPs using up to 2 and 5 lags were amongst the worst performing models in terms of cost. Figure 19 compares the test set real data to its predictions ("Prev Test") by the MLP models that take up to 2 (Fig. 19a), 5 (Fig. 19b), and 20 lags (Fig. 19c) as input.

Figure 19 –MLP purchase price prediction

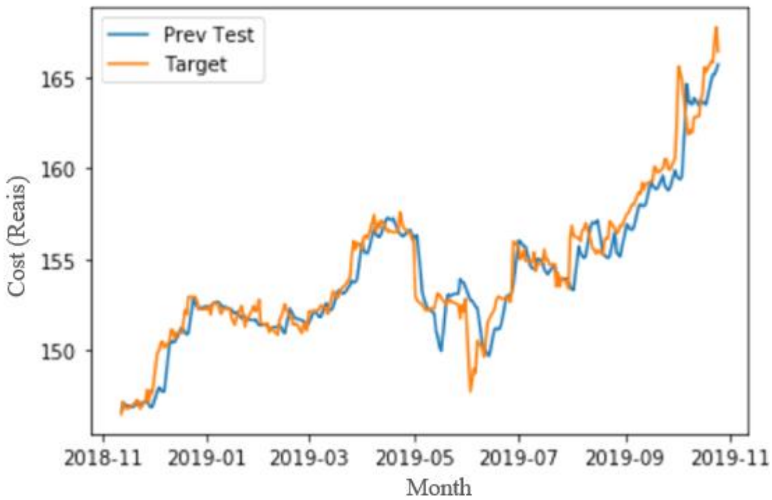
a. Input: Lags 1 and 2




b. Input: Lags 1 to 5




c. Input: Lags 1 to 20

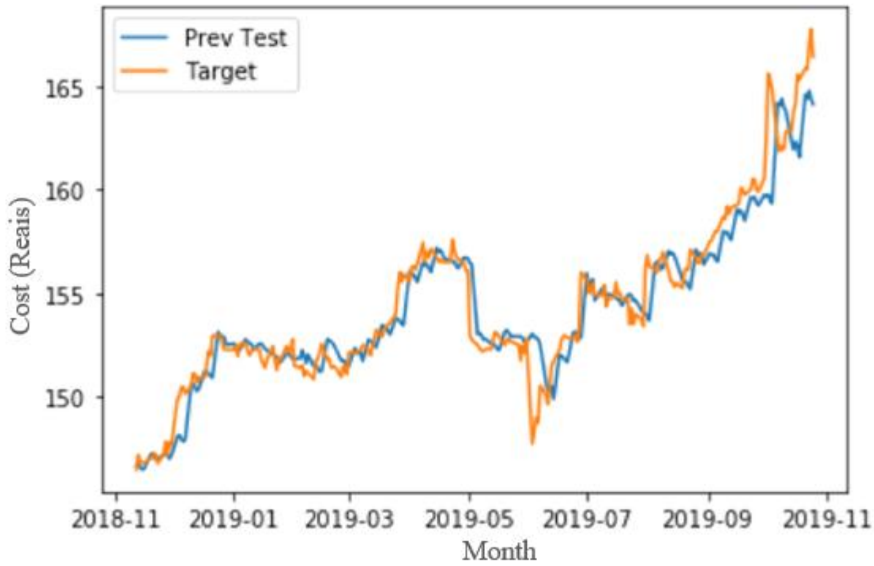



Source: The author (2021)

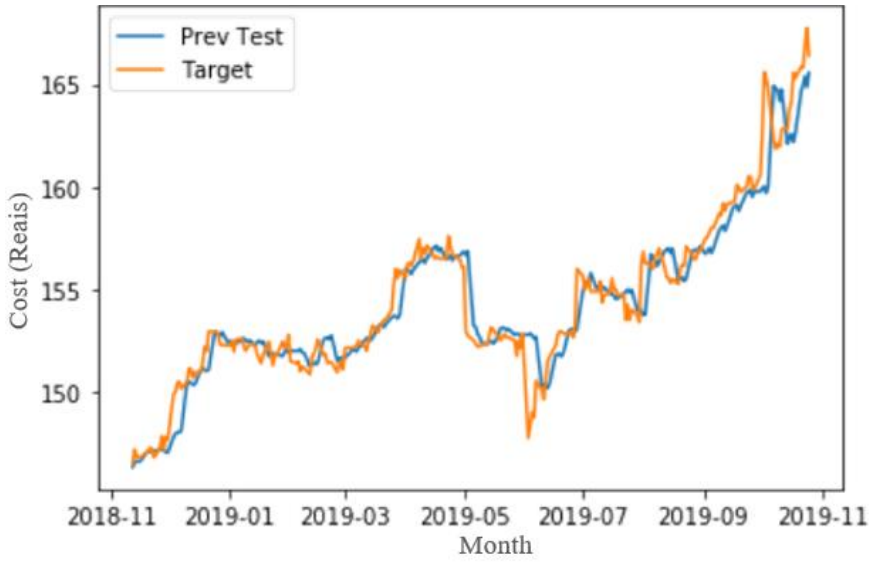
The RNNs had a good overall MSE performance, ranking 3rd, 4th, and 5th amongst all models. The 5th-ranked was the RNN that took lags 1 and 2 as input, which obtained an MSE of 0.00161636. Although it did not perform badly compared to the others in terms of MSE, it was the worst model in terms of average weekly unit cost. The RNN that used up to 5 lags of information performed slightly better in terms of MSE, obtaining an average of 0.00161587, and much better in terms of average unit cost, losing to the random strategy baseline by a small margin. Still, it did not manage to outperform it and therefore is not considered useful in improving decisions. The better ranked of all three was the RNN that used lags 1 through 20 in its forecasts, which obtained an MSE of 0.00160994. Its biggest accomplishment, however, was being the top-ranked model in terms of improving unit costs by pointing out the best day to put in the order. Figure 20 shows all three models' predictions ("Prev Test") versus the target values. From the graphs, the models that use fewer lags seem to fit the series better at the beginning, whereas using 20 lags seemed to fit the series better towards the end, especially capturing the growth pattern. The 20 lags memory made it underestimate values at the beginning, taking longer to adjust to the new stabilized value after a steep growth.

Figure 20 –RNN purchase price prediction

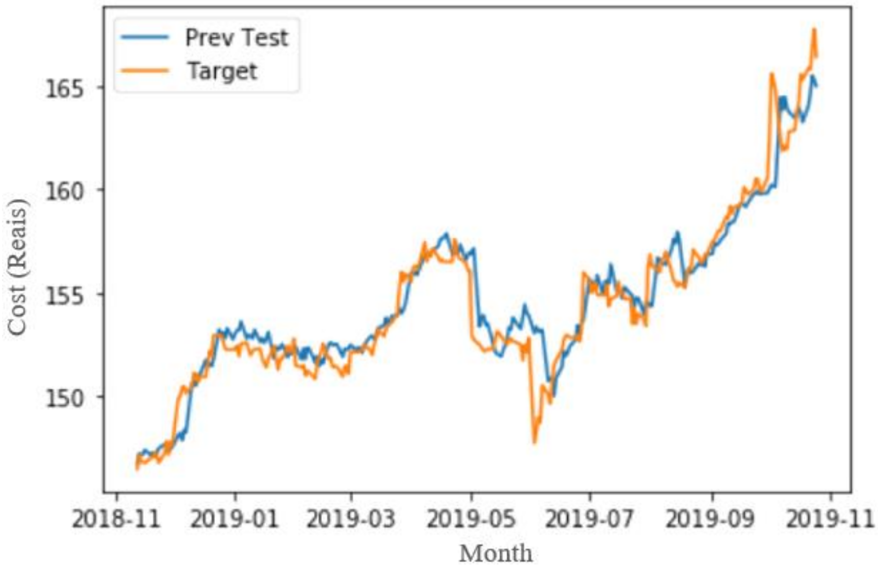
a. Input: Lags 1 and 2



b. Input: Lags 1 to 5




c. Input: Lags 1 to 20

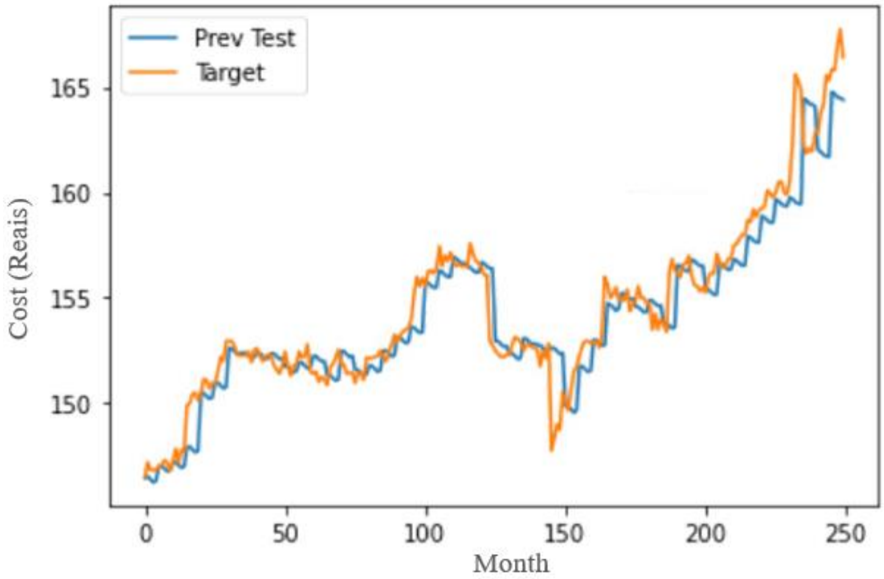



Source: The author (2021)

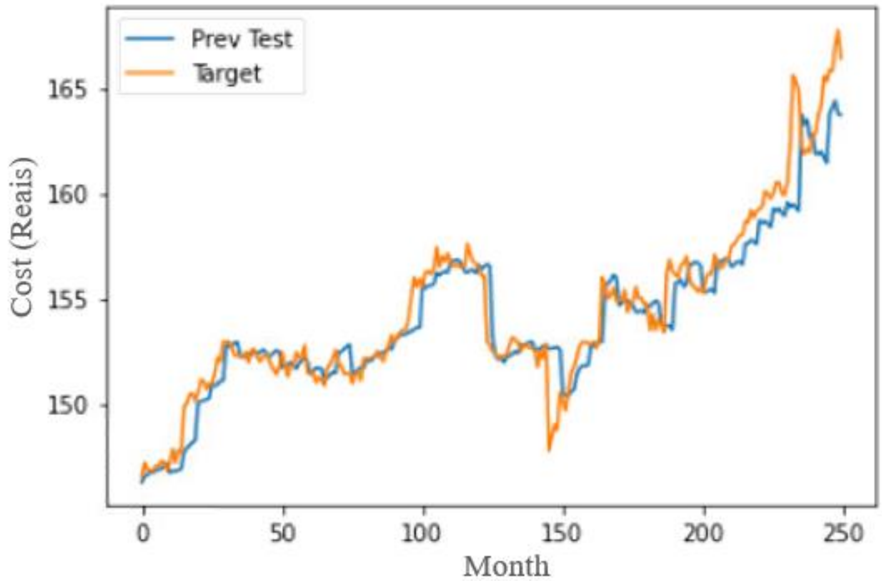
Although both MLP and RNN using 20 lags were the best-performing and only useful models, the long memory of the LSTM (both by definition and by the number of lags inputted) did not seem to help it. The 20-lag LSTM was the second-worst performer in terms of cost, and 9th out of 11 in the MSE ranking. The 5-lag LSTM was 10th out of 11 in the MSE ranking, losing only to the Prophet model, although it was the best-placed LSTM in terms of cost performance. However, it still lost to the average strategy baseline, SARIMA, and its RNN 5-lag counterpart, showing that it was not a very interesting model in this case. The LSTM using only lags 1 and 2 showed a very good MSE result, obtaining 2nd place in the ranking. However, in terms of the problem's need for guidance regarding what day to buy, it ranked 9th and lost to many other strategies, including the zero-information average strategy baseline. Figure 20 compares the test set real data to its predictions ("Prev Test") by the LSTM models that take up to 2 (Fig. 21a), 5 (Fig. 21b), and 20 lags (Fig. 21c) as input.

Figure 21 – LSTM purchase price prediction

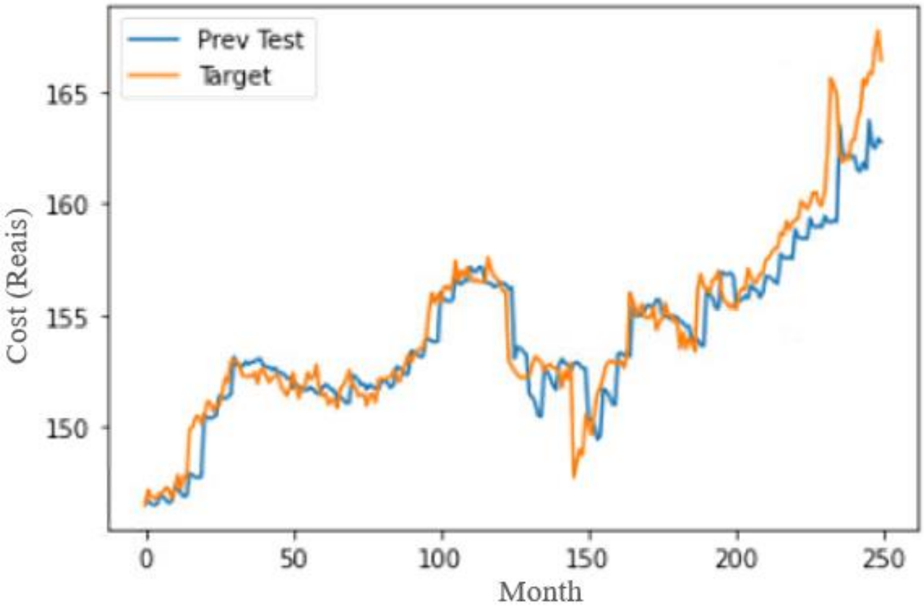
a. Input: Lags 1 and 2




b. Input: Lags 1 to 5



c. Input: Lags 1 to 20




Source: The author (2021)

Preliminary results from this investigation were submitted, accepted for publication, and presented in the 2020 IEEE Symposium Series on Computational Intelligence (GUIMARÃES *et al.*, 2020). However, only the purchase price predictions were analyzed then. The next step is assessing the joint decision of when and how much to buy by refocusing on the optimization model.

## 4.2 OPTIMIZATION MODEL COST PERFORMANCE

Once all purchase prices and consequently predicted overage costs are calculated, optimal order quantities are defined using the obtained critical fractile, and the random demands are simulated in order to evaluate and compare total cost performances. Three separate groups of results were obtained, one for each different value of the demand distribution's standard deviation.

The 10,000 total cost values, adding each iteration's overage and underage costs, are stored in a Pandas data frame. Using the "describe" function, overall descriptive statistics were obtained for each model's results, as shown in Tables 8, 9, and 10.

Table 8 – Total cost descriptive statistics, D ~ N(1000, $50^2$)

| | Random | SARIMA | Prophet | MLP 2 lags | MLP 5 lags | MLP 20 lags | RNN 2 lags | RNN 5 lags | RNN 20 lags | LSTM 2 lags | LSTM 5 lags | LSTM 20 lags |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **count** | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 |
| **mean** | 172355.97 | 79298.16 | 79301.33 | 79310.57 | 79307.31 | 79307.12 | 79303.43 | 79298.60 | 79308.81 | 79299.98 | 79310.56 | 79320.93 |
| **std** | 30602.24 | 10374.45 | 10397.03 | 10391.68 | 10389.97 | 10390.09 | 10397.84 | 10384.89 | 10390.63 | 10383.90 | 10392.76 | 10397.66 |
| **min** | 77164.12 | 48483.93 | 48303.21 | 48490.88 | 48481.01 | 48510.50 | 48487.20 | 48483.26 | 48573.35 | 48485.89 | 48475.80 | 48506.94 |
| **25%** | 150995.58 | 71890.56 | 71872.82 | 71881.25 | 71876.47 | 71896.72 | 71871.27 | 71874.42 | 71889.41 | 71878.82 | 71878.79 | 71889.91 |
| **50%** | 171010.78 | 78414.62 | 78403.74 | 78433.34 | 78423.65 | 78433.16 | 78417.63 | 78414.91 | 78434.05 | 78416.38 | 78428.62 | 78441.14 |
| **75%** | 192668.40 | 85537.07 | 85561.60 | 85560.61 | 85545.04 | 85549.81 | 85551.60 | 85537.88 | 85553.71 | 85546.06 | 85550.41 | 85571.90 |
| **max** | 303968.17 | 130357.10 | 130318.22 | 130459.68 | 130489.45 | 130505.64 | 130495.97 | 130481.78 | 130543.35 | 130419.20 | 130501.81 | 130471.49 |

Source: The author (2021)

Table 9 – Total cost descriptive statistics, D ~ N(1000, $150^2$)

| | Random | SARIMA | Prophet | MLP 2 lags | MLP 5 lags | MLP 20 lags | RNN 2 lags | RNN 5 lags | RNN 20 lags | LSTM 2 lags | LSTM 5 lags | LSTM 20 lags |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **count** | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 |
| **mean** | 516902.78 | 238551.54 | 238564.86 | 238588.83 | 238576.54 | 238576.08 | 238565.34 | 238553.51 | 238582.04 | 238555.70 | 238591.87 | 238614.62 |
| **std** | 89657.39 | 30981.40 | 31049.86 | 31035.27 | 31027.65 | 31025.15 | 31046.46 | 31010.91 | 31025.12 | 31006.95 | 31034.79 | 31041.56 |
| **min** | 217872.26 | 151103.19 | 151254.45 | 150965.24 | 150979.23 | 150999.98 | 150879.52 | 150994.32 | 151006.23 | 151009.07 | 150977.85 | 150972.18 |
| **25%** | 453250.30 | 216335.53 | 216375.21 | 216427.68 | 216384.81 | 216376.88 | 216344.04 | 216365.20 | 216381.28 | 216352.60 | 216361.46 | 216399.27 |
| **50%** | 513605.38 | 236101.05 | 236052.16 | 236139.86 | 236144.39 | 236130.81 | 236127.34 | 236126.15 | 236111.19 | 236131.73 | 236157.28 | 236154.63 |
| **75%** | 576008.38 | 257389.79 | 257492.60 | 257517.40 | 257492.53 | 257445.50 | 257499.60 | 257401.93 | 257465.34 | 257418.81 | 257465.62 | 257483.88 |
| **max** | 890204.94 | 392815.32 | 393796.04 | 393163.15 | 393136.46 | 393133.98 | 393326.08 | 393053.41 | 393019.14 | 393110.13 | 393160.51 | 393663.34 |

Source: The author (2021)

Table 10 – Total cost descriptive statistics, D ~ N(1000, $300^2$)

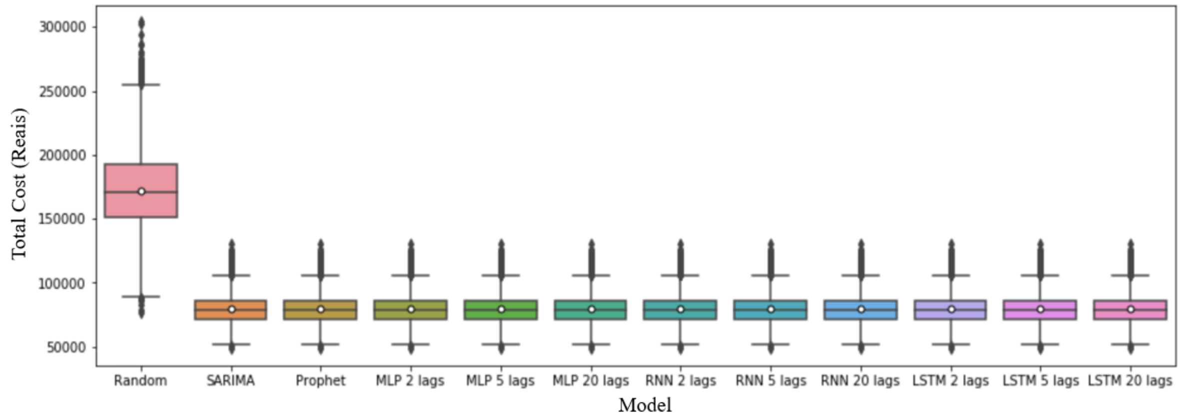
| | Random | SARIMA | Prophet | MLP 2 lags | MLP 5 lags | MLP 20 lags | RNN 2 lags | RNN 5 lags | RNN 20 lags | LSTM 2 lags | LSTM 5 lags | LSTM 20 lags |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **count** | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 |
| **mean** | 103453... | 475846.29 | 475865.14 | 475919.69 | 475897.43 | 475896.79 | 475875.57 | 475849.95 | 475910.91 | 475856.36 | 475922.93 | 475977.54 |
| **std** | 184126.90 | 62264.42 | 62400.86 | 62374.43 | 62363.58 | 62356.79 | 62401.61 | 62329.74 | 62360.76 | 62320.29 | 62375.06 | 62397.51 |
| **min** | 483795.71 | 286229.23 | 285364.60 | 286065.60 | 286073.61 | 285961.92 | 285911.74 | 286033.15 | 285994.76 | 286090.43 | 286039.17 | 286002.37 |
| **25%** | 907097.12 | 430981.29 | 431005.94 | 430918.93 | 430963.36 | 430931.60 | 430903.83 | 430988.74 | 430981.15 | 430940.84 | 430988.78 | 431021.24 |
| **50%** | 102572... | 469699.02 | 469714.52 | 469849.48 | 469793.91 | 469795.33 | 469737.23 | 469738.29 | 469902.15 | 469736.56 | 469848.53 | 469843.82 |
| **75%** | 115527... | 514279.98 | 514563.10 | 514437.03 | 514406.98 | 514441.56 | 514377.59 | 514294.50 | 514427.15 | 514294.52 | 514402.59 | 514525.43 |
| **max** | 190984... | 798097.80 | 800502.85 | 799439.71 | 799246.20 | 799627.47 | 799879.30 | 798589.22 | 799227.20 | 799074.24 | 799263.11 | 799890.46 |

Source: The author (2021)

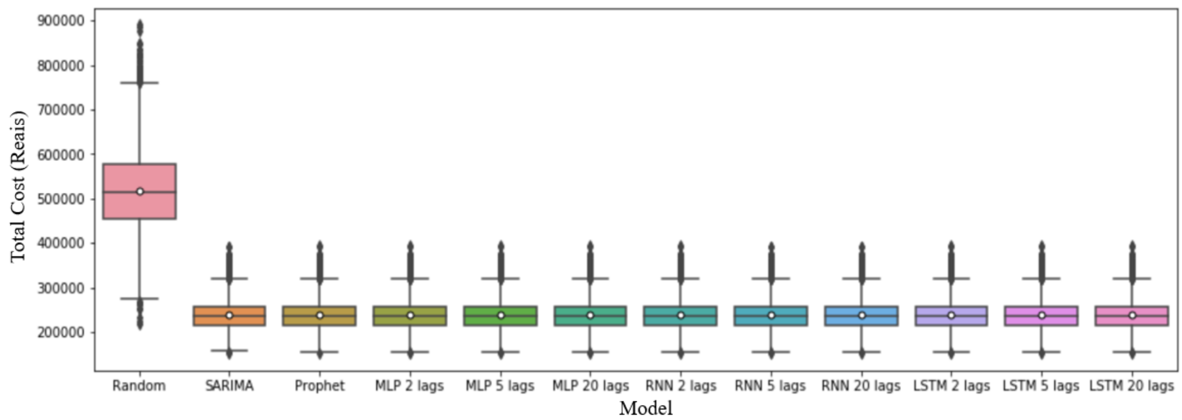
Although it is possible to extract some insight from the tables, they are more clearly visualized through the boxplots (Figure 22). By looking at the three y-axes, it is possible to tell that overall underage and overage costs largely increase as the uncertainty in demand increases for all assessed strategies. The greater uncertainty appears to affect all of the evaluated strategies similarly, as all three graphs look alike, except for their scales. Therefore, no strategy seems to stand out in terms of dealing with larger uncertainties.

Figure 22 – Total cost boxplot

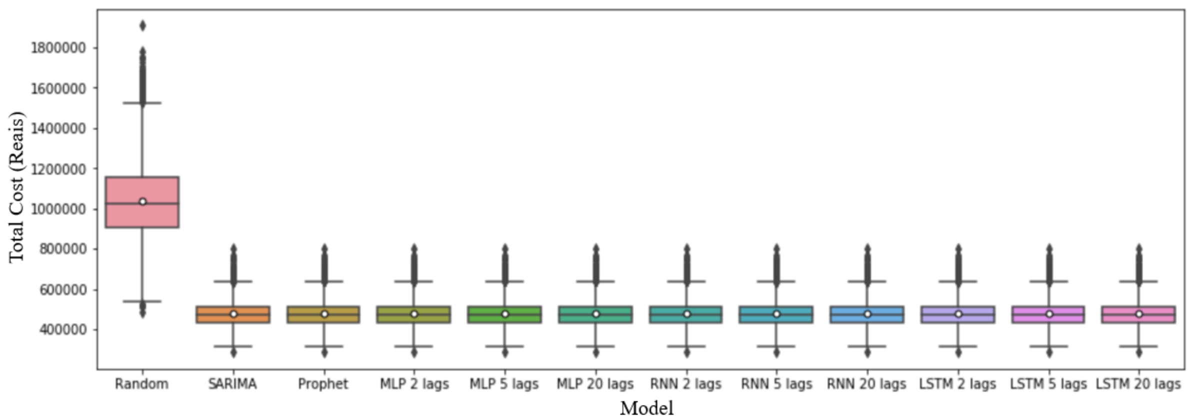
a. D ~ N(1000, $50^2$)




b. D ~ N(1000, $150^2$)




c. D ~ N(1000, $300^2$)




Source: The author (2021)

One of the most striking observations is how the random strategy's mean and median values are, in all three scenarios, roughly twice as expensive as each of the others' measures of central tendencies. This suggests that there is a relevant advantage to using any one of them over not having a defined prediction-based strategy. The Kruskal-Wallis non-parametric test was applied to costs from all of them to test the hypothesis that the choice of strategy has a significant impact on the total cost outcome, that is, they do not all present statistically similar results. Different tests were performed for each level of demand uncertainty. All three tests obtained a p-value smaller than $10^{-16}$, which indicates a strong rejection of the null hypothesis that all strategies, including the baseline, obtain values with similar medians and distributions. The test neither indicates which nor how many of the compared samples show different patterns. That is, further investigation is needed in order to identify if this result is due to the notably different behavior of the random strategy alone, or if other strategies have also shown significantly different results from the others.

Since all strategies seem very similar when compared to the baseline, the latter was temporarily removed from the observations. A heatmap applied to the descriptive statistics helps visualize and compare the methods. For a demand standard deviation of 50 (Table 11), the 20 lag LSTM showed the highest mean and median amongst all methods. It also showed the highest third quartile. The other neural networks that made predictions based on the previous 20 lags were also costly methods: the RNN showed the second-highest median and highest minimum and maximum, whereas the MLP showed the highest first quartile, but also high minimum, median and maximum. In terms of mean, the most expensive methods, besides the 20-lag LSTM, are the 2-lag MLP and the 5-lag LSTM. The lowest mean was obtained by SARIMA, which also obtained the lowest third quartile and second-lowest maximum. Other low means were obtained by the 5-lag RNN and the 2-lag LSTM. Although obtaining only the fourth-best mean, the Prophet model obtained the lowest minimum, median, and maximum, showing a very good result compared with the other models.

Table 11 – Total cost heatmap, D ~ N(1000, $50^2$)

| | SARIMA | Prophet | MLP 2 lags | MLP 5 lags | MLP 20 lags | RNN 2 lags | RNN 5 lags | RNN 20 lags | LSTM 2 lags | LSTM 5 lags | LSTM 20 lags |
|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 |
| mean | 79298.16 | 79301.33 | 79310.57 | 79307.31 | 79307.12 | 79303.43 | 79298.60 | 79308.81 | 79299.98 | 79310.56 | 79320.93 |
| std | 10374.45 | 10397.03 | 10391.68 | 10389.97 | 10390.09 | 10397.84 | 10384.89 | 10390.63 | 10383.90 | 10392.76 | 10397.66 |
| min | 48483.93 | 48303.21 | 48490.88 | 48481.01 | 48510.50 | 48487.20 | 48483.26 | 48573.35 | 48485.89 | 48475.80 | 48506.94 |
| 25% | 71890.56 | 71872.82 | 71881.25 | 71876.47 | 71896.72 | 71871.27 | 71874.42 | 71889.41 | 71878.82 | 71878.79 | 71889.91 |
| 50% | 78414.62 | 78403.74 | 78433.34 | 78423.65 | 78433.16 | 78417.63 | 78414.91 | 78434.05 | 78416.38 | 78428.62 | 78441.14 |
| 75% | 85537.07 | 85561.60 | 85560.61 | 85545.04 | 85549.81 | 85551.60 | 85537.88 | 85553.71 | 85546.06 | 85550.41 | 85571.90 |
| max | 130357.10 | 130318.22 | 130459.68 | 130489.45 | 130505.64 | 130495.97 | 130481.78 | 130543.35 | 130419.20 | 130501.81 | 130471.49 |

Row max

Row min

Source: The author (2021)

For a demand standard deviation of 150 (Table 12), on the other hand, Prophet obtained expensive minimum and maximum costs. The 20-lag LSTM continued to hold the highest mean and showed the second-highest median. The bottom-ranked median this turn belonged to the 5-lag LSTM. The most expensive first and third quartiles belonged to the 2-lag MLP, which also showed a high mean and median compared to the others. As a positive highlight, SARIMA presented the lowest mean, first quartile, third quartile, and maximum.

Table 12 – Total cost heatmap, D ~ N(1000, $150^2$)

| | SARIMA | Prophet | MLP 2 lags | MLP 5 lags | MLP 20 lags | RNN 2 lags | RNN 5 lags | RNN 20 lags | LSTM 2 lags | LSTM 5 lags | LSTM 20 lags |
|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 |
| mean | 238551.54 | 238584.86 | 238588.83 | 238576.54 | 238576.08 | 238565.34 | 238553.51 | 238582.04 | 238555.70 | 238591.87 | 238614.62 |
| std | 30981.40 | 31049.86 | 31035.27 | 31027.65 | 31025.15 | 31046.46 | 31010.91 | 31025.12 | 31006.95 | 31034.79 | 31041.56 |
| min | 151103.19 | 151254.45 | 150965.24 | 150979.23 | 150999.98 | 150879.52 | 150994.32 | 151006.23 | 151009.07 | 150977.85 | 150972.18 |
| 25% | 216335.53 | 216375.21 | 216427.68 | 216384.81 | 216376.88 | 216344.04 | 216365.20 | 216381.28 | 216352.60 | 216361.46 | 216399.27 |
| 50% | 236101.05 | 236052.16 | 236139.86 | 236144.39 | 236130.81 | 236127.34 | 236126.15 | 236111.19 | 236131.73 | 236157.28 | 236154.63 |
| 75% | 257389.79 | 257492.60 | 257517.40 | 257492.53 | 257445.50 | 257499.60 | 257401.93 | 257465.34 | 257418.81 | 257465.62 | 257483.88 |
| max | 392815.32 | 393796.04 | 393163.15 | 393136.46 | 393133.98 | 393326.08 | 393053.41 | 393019.14 | 393110.13 | 393160.51 | 393663.3 |

Row max

Source: The author (2021)

Row min

Following its pattern, the 20-lag LSTM did not perform well for a demand standard deviation of 150 (Table 13). It obtained worse mean and first quartile costs, and second worse maximum. Prophet presented the highest maximum and 3rd quartile, although also obtaining the lowest minimum and second-lowest median. The second and third lowest mean belonged to the 5-lag RNN and the 2-lag LSTM, which also ranked well in terms of the median and third quartile. The best mean, however, was obtained by SARIMA, which also showed the lowest median, third quartile, and maximum, although it had the highest minimum. Once again, the model was a positive highlight.

Table 13 – Total cost heatmap, D ~ $N(1000, 300^2)$

| | SARIMA | Prophet | MLP 2 lags | MLP 5 lags | MLP 20 lags | RNN 2 lags | RNN 5 lags | RNN 20 lags | LSTM 2 lags | LSTM 5 lags | LSTM 20 lags |
|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 |
| mean | 475846.29 | 475865.14 | 475919.69 | 475897.43 | 475896.79 | 475875.57 | 475849.95 | 475910.91 | 475856.36 | 475922.93 | 475977.54 |
| std | 62284.42 | 62400.86 | 62374.43 | 62363.58 | 62356.79 | 62401.61 | 62329.74 | 62360.76 | 62320.29 | 62375.06 | 62397.51 |
| min | 286229.23 | 285364.60 | 286065.60 | 286073.61 | 285961.92 | 285911.74 | 286033.15 | 285994.76 | 286090.43 | 286039.17 | 286002.37 |
| 25% | 430981.29 | 431005.94 | 430918.93 | 430963.36 | 430931.60 | 430903.83 | 430988.74 | 430981.15 | 430940.84 | 430988.78 | 431021.24 |
| 50% | 469699.02 | 469714.52 | 469849.48 | 469793.91 | 469795.33 | 469737.23 | 469738.29 | 469902.15 | 469736.56 | 469848.53 | 469843.82 |
| 75% | 514279.98 | 514563.10 | 514437.03 | 514406.98 | 514441.56 | 514377.59 | 514294.50 | 514427.15 | 514294.52 | 514402.59 | 514525.43 |
| max | 798097.80 | 800502.85 | 799439.71 | 799246.20 | 799627.47 | 799879.30 | 798589.22 | 799227.20 | 799074.24 | 799263.11 | 799890.46 |

Row max
Row min

Source: The author (2021)

In order to determine how statistically significant were the differences between the costs obtained from the application of each method, the Kruskal-Wallis test was performed once again for each level of demand variance, this time excluding only the random strategy. In this case, the p-value were all approximately 1 (0.9999999999822967, 0.9999999999887775, and 0.9999999999868868, for standard deviations 50, 150, and 300, respectively). All three values are much larger than the significance value of 0.05 and indicate that the null hypothesis that all remaining strategies follow similar distributions should not be rejected for any level of demand uncertainty. These extremely high p-values are a strong indicator that, although performing slightly differently from each other, the differences between strategies are not statistically significant in any of the evaluated cases, and the strategies can be considered to show similar performances. It can also be inferred that the first Kruskal-Wallis result indicates that, in fact, the baseline performed significantly worse than all the other methods.

## 4.3 DISCUSSION

Observing only the decision of when to buy, the 20-lag RNN and the 20-lag MLP showed better results than the zero-information scenario and could reduce costs related to the unit price at the time of the purchase. These savings on regular, high-volume product procurement can save companies millions of dollars. Other proposed models did not outperform the random strategy when isolating the decision of when to place the order. It was expected that models that used more historical information on the series would show a better performance. Curiously, the 20-lag LSTM did not offer good forecasting results. This is unexpected because of the similarity between LSTM and RNN models.

When both order quantity and timing decisions were considered simultaneously, all models managed to obtain around 50% lower overall underage and overage costs than the baseline, with similar performances among themselves. Because there was no model that offered a statistically relevant better underage and overage cost performance, the computational time should be taken into account when choosing a model to be applied. This is especially true since the problem involves a weekly purchase, that is, running the model every week to make a decision. In our experiments, SARIMA showed to be the most efficient model, taking shortly over a minute to train. LSTM, on the other hand, was the longest and took 13 hours to train. Table 14 shows training and forecasting times for each method. All models, except the LSTM, ran in an IntelCore i5-6200U CPU, with 6GB RAM. The LSTM did not run on this CPU and had to be run on an Intel i9 9900K, with 32 GB of memory. Once again the LSTM showed an unexpectedly different running time from the RNN, although both codes are practically identical. Further investigation is needed to understand this behavior.

Table 14 – Model training and forecasting time

| Model | Time (h:m:s) |
|---|---|
| SARIMA | 00:01:20.176236 |
| Prophet | 00:29:10.871661 |
| MLP Lag 2 | 00:01:31.217546 |
| MLP Lag 5 | 00:02:04.624071 |
| MLP Lag 20 | 00:04:44.148654 |
| RNN Lag 2 | 00:39:08.157913 |
| RNN Lag 5 | 00:39:32.411578 |
| RNN Lag 20 | 00:40:03.740074 |
| LSTM Lag 2 | 13:05:27.428923 |
| LSTM Lag 5 | 13:08:37.278873 |
| LSTM Lag 20 | 13:01:41.578850 |

Source: The author (2021)

The fact that a slightly better contribution of the 20-lag RNN and the 20-lag MLP in deciding on the best day to make the purchase did not show up in the joint-decision results suggests that, for this series, the contribution of all the models in aiding in the order quantity decision was more significant than these two's ability to point out the best order timing. This most likely happens because the series resembles a random walk and did not show a clear pattern upon which the methods could provide better insight. Historical price information from other series might show greater differences in prices throughout the week and have larger gains from selecting the best day.

Since the order quantity decision is calculated based on an expected purchase price, having a forecast instead of relying on the average demand proved to be useful, even when using models with lower forecasting accuracies. Prophet, for instance, showed the highest MSE in the time series forecasting but was arguably the best basis for decision-making in the low-variance scenario for the demand. SARIMA, which modeled the series as a simple random walk, had excellent performance in both standard deviations of 150 and 300 for the demand. By analyzing their performance, it is clear that their insight on the future behavior of the series was useful decision-guiding information. Only an adequate time series modeling will show if price fluctuations might be estimated by a simple random walk, or if more complex patterns will be identified. This insight has shown to be effective in estimating overage prices and supporting the purchase quantity decision.

# 5 CONCLUSIONS AND FURTHER WORK

In this dissertation, a comprehensive study was developed around the situation in which a retailer is faced with the problem of stochastic purchase prices and must decide when is the best day to place an order, as well as how much to buy to restock his perishable supply. In order to support the decision-making process, the problem was modeled as a variant of the newsvendor problem, subject to two decision variables: when to place the order and how much to buy. After adapting the mathematical model, the decision of when to buy was centered around the expected purchase price. Therefore, multiple time series forecasting techniques were assessed in their ability to support the decision-making.

All the assessed decision-making strategies outperformed the original zero-information scenario. Results suggest that the greatest contribution of using a forecasting model is having a reasonable overage estimation and using it to support the order quantity decision. All models, although forecasting slightly different prices, offered much better order sizes than the baseline strategy of buying the mean demand.

More insight can be obtained by expanding the current investigation to address the new behavior the series has shown since November 2019, which was a limitation to the scope of this work. Also, a scenario with a fixed lead time after placing the order can be explored in how it would impact the decision versus the fixed delivery day that was considered in this work.

Future work involving other successful time series forecasting methods should show promising results in improving overage estimation, which impacts both discussed decisions. Recent methodologies use a combination of linear and nonlinear methods to extract different patterns from the series. Linear models can include ARMA, Holt-Winters, Prophet, among others, whereas the nonlinear relations can be modeled using machine learning methods, such as neural networks. These methods are then combined by use of linear or nonlinear models and are often referred to as “hybrid” models (OLIVEIRA, 2020; BABU; REDDY, 2014; MATTOS NETO; CAVALCANTI; MADEIRO, 2017; ZHANG, 2003). Hybrid models have also shown competitive results in state-of-the-art global forecasting competitions (MAKRIDAKIS *et al.*, 2020).

The use of more advanced preprocessing techniques can also aid in extracting features from the time series, which can help models better identify patterns and improve their forecasts. Empirical Mode Decomposition (HUANG *et al.*, 1998), for example, is a preprocessing method that decomposes the series into multiple signals, identifying possibly many patterns within a single series. This may help identify different sources of uncertainty, and address them separately, investigating the best way to deal with each one.

Also, procurement of perishable products is highly influenced by meteorological conditions, pests, trade deals, and other environmental or macroeconomic phenomena, which lead to a greater need for produce buyers to make decisions using current, accurate, and fast-changing information. Natural Language Processing (NLP) (MANNING & SCHUTZE, 1999) can extract recently published relevant information from the internet and use it as a feature to accurately anticipate the best timing to place an order. Predictions can then be used as input to the optimization model that is proposed to decide the ideal quantity and timing of the order jointly. NLP has shown multiple successful applications in supply chain risk management (GUIMARÃES *et al.*, 2019).

The decision of when to buy can also be aided by a classification model, that can decide which of the 5 days is best. This may show a better performance than trying to forecast the actual value on a certain day. However, this policy would have to be paired with some kind of forecasting method to decide on the purchase price parameter to be used in the critical fractile, since this is the most relevant difference between the methods assessed in this work and the random strategy.

In terms of improving the order quantity decision, which has shown to have a great impact on overall costs, rethinking the overage and underage parameters might lead to a better decision. Since underage usually depends on the profit, this is directly impacted by the practiced purchase price. Studying the mathematical implications of having the underage be a function of the purchase price, as well as the 11% opportunity cost parameter, might improve the critical fractile solution. Different parameters for the demand behavior can also be tested, as well as other distributions. Gamma distribution is an adequate alternative (RAMAEKERS & JANSSENS, 2008).

Applying these potential improvements can significantly impact not only purchase price predictions, but also optimal order quantity predictions. This can easily be translated into the numerous practical applications of newsvendor-type problems, incurring significant cost reductions. This joint-decision strategy can be transformed, for example, into a technological product to be used by retailers to support their weekly purchase decision. As mentioned in the discussion, at this point all strategies showed similar performances, so using a faster method such as SARIMA would be more appropriate for a tool that will be handled weekly, and most likely extended to work not only for beef but all other perishable products.

# REFERENCES


ABDEL-MALEK, L. L.; MONTANARI, R. On the multi-product newsboy problem with two constraints. *Computers & Operations Research*, v. 32, n. 8, p. 2095-2116, 2005.

ARROW, K.; HARRIS, T.; MARSCHAK, J. Optimal inventory policy. *Econometrica*, v. 19, p. 250–272, 1951.

BABU, C. N.; REDDY, B. E. A moving-average filter-based hybrid ARIMA–ANN model for forecasting time series data. *Applied Soft Computing*, n. 23, p. 27-38, 2014.

BAN, G.-Y.; RUDIN, C. The big data newsvendor: practical insights from machine learning. *Operations Research*, v. 67, n. 1, p. 90–108, 2019.

BEHRET, H.; KAHRAMAN, C. A multi-period newsvendor problem with pre-season extension under fuzzy demand. *Journal of Business Economics and Management*, v. 11, n. 4, p. 613-629, 2010.

BERLING, P.; MARTÍNEZ-DE-ALBÉNIZ, V. Optimal inventory policies when purchase price and demand are stochastic. *Operations Research*, v. 59, n. 1, p. 109-124, 2011.

BERTSIMAS, D.; KALLUS, N. From predictive to prescriptive analytics. Management Science, v. 66, n. 3, p. 1005-1507, 2020.

BERTSIMAS, D.; THIELE, A. *A data-driven approach to newsvendor problems*. Cambridge, MA: Massachusetts Institute of Technology Press, 2005.

BIERY, M. E. These industries generate the lowest profit margins, 2017. Forbes. https://www.forbes.com/sites/sageworks/2017/09/24/these-industries-generate-the-lowest-profit-margins/#d4fd000f49d2 [Online; accessed Oct. 2, 2020]

BIRD, S.; LOPER, E.; KLEIN, E. *Natural Language Processing with Python. O'*Reilly Media Inc, 2009.

BLACK, F.; SCHOLES, M. The pricing of options and corporate liabilities. *Journal of Political Economy*, n. 81, p. 637–659, 1973.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. *Time series analysis*: forecasting and control. 4 ed. John Wiley & Sons, 2015.

BREIMAN, L. Random forests. *Machine Learning*, v. 45, p. 5-32, 2001.

BROYDEN, C. G. The convergence of a class of double-rank minimization algorithms: 2. The new algorithm. *IMA Journal of Applied Mathematics*, v.6, n.3, p.222–231, 1970.

BROWN, R. G.; MEYER, R. F. The fundamental theory of exponential smoothing. *Operations Research*, v. 9, p. 673-685, 1961.

CHOI, T. *Handbook of Newsvendor Problems*: Models, Extensions and Applications. New York: Springer, 2012.

CHOLLET, F. *et al.* Keras, 2015. https://keras.io [Online; accessed Oct. 2, 2020]

COWPERTWAIT, P. S. P.; METCALFE, A. V. *Introductory time series with R*. New York: Springer, 2009.

DEYONG, G. D. The price-setting newsvendor: review and extensions. International Journal of Production Research, DOI: 10.1080/00207543.2019.1671624, 2019.

EDGEWORTH, F. Y. The mathematical theory of banking. *Journal of the Royal Statistical Society*, v. 51, n. 1, p. 113-127, 1888.

FAYYAZ, M.; SAFFAR, M. H.; SABOKROU, M.; FATHY, M.; HUANG, F.; KLETTE, R. STFCN: Spatio-temporal fully convolutional neural network for semantic segmentation of street scenes. In: *Asian Conference on Computer Vision*. Springer, Cham, p. 493-509, 2016.

FLETCHER, R. A new approach to variable metric algorithms. *The Computer Journal*, v. 13, n. 3, p. 317-322, 1970.

FONTELLES, M. J.; SIMÕES, M. G.; FARIAS, S. H.; FONTELLES, R. G. S. Metodologia da pesquisa científica: diretrizes para a elaboração de um protocolo de pesquisa. *Revista Paraense de Medicina*, v. 23, n. 3, p. 1-8, 2009.

FUSION MEDIA LIMITED (FML), Historical Data – Live Cattle Contract, https://br.investing.com/commodities/live-cattle [Online; accessed on Aug. 7, 2020].

GARDNER, E. S. Exponential smoothing: the state of the art – Part II. *International Journal of Forecasting*, v. 22, n. 4, p. 637-666, 2006.

GARGANO, A.; TIMMERMANN, A. Forecasting commodity price indexes using macroeconomic and financial predictors. *International Journal of Forecasting*, v. 30, n. 3, p. 825-843, 2014.

GESSULLI AGRIBUSINESS. Como a Peste Suína Africana influenciou os hábitos de consumo na China?, 2020. https://www.suinoculturaindustrial.com.br/imprensa/como-a-peste-suina-africana-influenciou-os-habitos-de-consumo-na-china/20200626-105653-o859 [Online; accessed Dec. 1, 2020]

GILLILAND, M. The value added by machine learning approaches in forecasting. *International Journal of Forecasting*, v. 36, n. 1, p. 161-166, 2020.

GOLDFARB, D. A family of variable-metric methods derived by variational means. *Mathematics of Computation*, v. 24, n. 109, p. 23-26, 1970.

GUIMARÃES, M. S.; LINS, I. D.; MOURA, M. C. Artificial Intelligence applications in supply chain risk assessment. *Anais do Congresso da Associação Brasileira de Análise de Riscos, Segurança de Processo e Confiabilidade – ABRISCO*, Rio de Janeiro, 2019.

GUIMARÃES, M. S.; LINS, I. D.; MOURA, M. C.; DUARTE, H. O. Application of Data-Based Prediction Methods in Newsvendor Problems Subject to Purchase Price Uncertainty. *Proceedings of 2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, Canberra, 2020.

HALMAN, N.; ORLIN, J. B.; SIMCHI-LEVI, D. Approximating the nonlinear newsvendor and single-item stochastic lot-sizing problems when data is given by an oracle. *Operations Research*, v. 60, n. 2, p. 429-446, 2012.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural Computation*, v. 9, n. 8, p. 1735-1780, 1997.

HU, X; MUNSON, C. L.; FOTOPOULOS, S. B. Purchasing decisions under stochastic prices: Approximate solutions for order time, order quantity and supplier selection. *Annals of Operations Research*, v. 201, p. 287–305, 2012.

HU, X; SU, P. The newsvendor's joint procurement and pricing problem under price-sensitive stochastic demand and purchase price uncertainty. *Omega*, v. 79, p. 81–90, 2018.

HUA, G. et al. The newsvendor problem with barter exchange. *Omega-International Journal of Management Science*, v. 92, UNSP 102149, 2020.

HUANG, N. E.; SHEN, Z.; LONG, S. R.; WU, M. C.; SHIH, H. H.; ZHENG, Q.; YEN, N.; TUNG, C. C.; LIU, H. H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, v. 454, n. 1971, p. 903-995, 1998.

HUBER, J.; MÜLLER, S.; FLEISCHMANN, M.; STUCKENSCHMIDT, H. A data-driven newsvendor problem: From data to decision. *European Journal of Operations Research*, v. 278, p. 904-915, 2019.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE) – "Tabela 1419 - Índice Nacional de Preços ao Consumidor Amplo (IPCA)," Sistema IBGE de Recuperação Automática - SIDRA, 2020.

KHOUJA, M. The single-period (news-vendor) problem: literature review and suggestions for future research. *Omega*, v. 27, n. 5, p. 537-553, 1999.

KHOUJA, M.; MEHREZ, A. A multi-product constrained newsboy problem with progressive multiple discounts. *Computers & Industrial Engineering*, v. 30, n. 1, p. 95-101, 1996.

KIM, G.; WU, K.; HUANG, E. Optimal inventory control in a multi-period newsvendor problem with non-stationary demand. *Advanced Engineering Informatics*, v. 29, n. 1, p. 139-145, 2015.

KINGMA, D. P.; BA, J. L. (2015). Adam: a Method for Stochastic Optimization. *Proceedings of International Conference on Learning Representations*, p. 1–13, 2015.

KRUSKAL, W. H.; WALLIS, W. W. Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, v. 47, n. 260, p. 583-621, 1952.

LE CUN, Y.; BENGIO, Y.; HINTON, G. Deep Learning. Nature, v. 521, p. 436-444, 2015.

LIU, D. C.; NOCEDAL, J. On the limited memory BFGS method for large-scale optimization. *Mathematical Programming*, v. 45, p. 503–528, 1989.

LOTFI, R. Interdependent demand in the two-period newsvendor problem. *Journal of Industrial and Management Optimization*, v. 16, n. 1, p. 117-140, 2020.

MAKRIDAKIS, S.; SPILIOTIS, E.; ASSIMAKOPOULOS, V. The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, v. 36, p. 54–74, 2020.

MATHCRACKER. Normal Probability Grapher. https://mathcracker.com/normal-probability-grapher [Online; accessed Dec. 1, 2020]

MATTOS NETO, P. S.; CAVALCANTI, G. D.; MADEIRO, F. Nonlinear combination method of forecasters applied to pm time series. *Pattern Recognition Letters*, n. 95, p. 65-72, 2017.

MCKINNEY, W. Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference (SciPy)*. Austin: Stéfan van der Walt and Jarrod Millman, p. 56-61, 2010.

MCLAUGHLIN, E. W.; PARK, K. S.; HAWKES, G. F. *Produce industry procurement:* Changing preferences and practices. E.B. 2015-10. Ithaca: Cornell University, 2015.

MUNYAO, R. M. Role of inventory management practices on performance of production department: a case of manufacturing firms. *International Journal of Economics, Commerce and Management*, v. 3, n. 5, p. 1625-1656, 2015.

MOHAMMED, R. Why high profit margins don’t prove smart pricing, 2009. https://hbr.org/2009/11/so-you-think-you-know-how-to-g/ [Online; accessed Oct. 2, 2020]

MORSE, P. M.; KIMBALL, G. E.; GASS, S. I. *Methods of operations research.* Courier Corporation, 2003.NEALE, J. J.; TOMLIN, B. T.; WILLEMS, S. P. The role of inventory in. superior supply chain performance. In: *The practice of Supply Chain Management: where theory and application converge*. Boston: Springer, 2004. cap 3, p. 31-59.

OLIVEIRA, J. F. L. *et al.* A hybrid optimized error correction system for time series forecasting. *Applied Soft Computing*, v. 87, 2020.

OROOJLOOYJADID, A.; SNYDER, L. V.; TAKÁČ, M. Applying Deep Learning to the Newsvendor Problem. *IISE Transactions*, v. 52, n. 4, p. 444-463, 2020.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825-2830, 2011.

PERONA, M.; CIGOLINI, R.; ADANI, M.; BIONDI, R.; GUZZETTI, S.; JENNA, R.; CHESSA, M.; AGELLARA, S. The integrated management of logistic chains in the white goods industry: A field research in Italy. *International Journal of Production Economics*, v. 69, p. 227-238, 2001.

PETRUZZI, N. C.; DADA, M. Pricing and the newsvendor problem: a review with extensions. *Operations Research*, v. 47, n. 2, p.183-194, 1999.

PORTEUS, E. L. The newsvendor problem. In: CHHAJED, D.; LOWE, T. J., ed. *Building Intuition*. International Series in Operations Research & Management Science, v. 115. Boston: Springer, 2008. chap 7, p.115-134.

RAMAEKERS, K.; JANSSENS, G. K. On the choice of a demand distribution for inventory management models. *European Journal of Industrial Engineering*, v. 2, n. 4, p. 479-491, 2008.

REHUREK, R.; SOJKA, P. Software framework for topic modeling with large corpora. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. Malta: European Language Resources Association, p. 45-50, 2010.

RITCHIE, H. (a) Cars, planes, trains: where do CO2 emissions from transport come from?, 2020. https://ourworldindata.org/co2-emissions-from-transport#:~:text=Transport%20accounts%20for%20around%20one,CO2%20emissions%20from%20energy%5D.&text=Road%20travel%20accounts%20for%20three,buses%20%E2%80%93%20which%20contribute%2045.1%25. [Online; accessed Dec. 29, 2020]

RITCHIE, H. (b) You want to reduce the carbon footprint of your food? Focus on what you eat, not whether your food is local, 2020. https://ourworldindata.org/food-choice-vs-eating-local [Online; accessed Dec. 29, 2020]

ROBBINS, H.; MONRO, S. A stochastic approximation method. *The annals of mathematical statistics*, p. 400-407, 1951.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, v. 65, n. 6, p. 386, 1958.

RUDIN, C.; VAHN, G. *The big data newsvendor*: Practical insights from machine learning analysis. Cambridge, MA: MIT Sloan School of Management, 2014.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *Nature*, v. 323, p. 533–536, 1986.

SAMARASINGHE, S. *Neural networks for applied sciences and engineering:* from fundamentals to complex pattern recognition. Crc Press, 2016.

SCHMIDHUBER, J. Deep learning in neural networks: An overview. Neural Networks, v. 61, p. 85-117, 2015.

SCHWARTZ, E. S. The stochastic behaviour of commodity prices: Implications for valuation and hedging. *Journal of Finance*, v. 52, n. 3, p. 923–973, 1997.

SEABOLD; SKIPPER; PERKTOLD, J. statsmodels: Econometric and statistical modeling with python. *Proceedings of the 9th Python in Science Conference*, 2010.

SHANNO, D. F. Conditioning of quasi-Newton methods for function minimization. *Mathematics of Computation*, v. 24, n. 111, p. 647-656, 1970.

SHAO, Z.; JI, X. Fuzzy multi-product constraint newsboy problem. Applied Mathematics and Computation, v. 180, n. 1, p. 7-15, 2006.

SILVER, E. A.; PYKE, D. F.; THOMAS, D. J. *Inventory and production management in supply chains*. New York: Taylor and Francis, 2017.

SMITH, T. G. *et al*. pmdarima: ARIMA estimators for Python, 2017. http://www.alkaline-ml.com/pmdarima [Online; accessed Aug. 22, 2020]

STATISTA, Least profitable industries in the United States as of January 2020, 2020. https://www. statista.com/statistics/468436/least-profitable-industries-usa/ [Online; accessed Oct. 2, 2020]

SWAMIDASS, P. M. *Encyclopedia of Production and Manufacturing Management*. Springer US, 2000.

TAYLOR, S. J.; LETHAM, B. Forecasting at scale. *The American Statistician*, v. 72, n. 1, p.37-45, 2018.

TUKEY, J. W. *Exploratory data analysis*. Addison-Wesley Publishing Company, 1977.ULLAH, M.; KHAN, I.; SARKAR, B. Dynamic pricing in a multi-period newsvendor under stochastic price-dependent demand. *Mathematics*, v. 7, p. 520, 2019.

QIN, Y.; WANG, R. The newsvendor problem: review and directions for future research. *European Journal of Operational Research*, v. 213, n. 2, p. 361-374, 2011.

VIRTANEN, P.; GOMMERS, R.; OLIPHANT, T. E.; HABERLAND, M.; REDDY, T.; COURNAPEAU, D.; BUROVSKI, E.; PETERSON, P.; WECKESSER, W.; BRIGHT, J.; VAN DER WALT, S. J.; BRETT, M.; WILSON, J.; MILLMAN, K. J.; MAYOROV, N.; NELSON, A. R. J.; JONES, E.; KERN, R.; LARSON, E.; CAREY, C. J.; POLAT, I.; FENG, Y.U; MOORE, E. W. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, v. 17, n. 3, p. 261-272, 2020.

WHITIN, T. M. Inventory control and price theory. *Management Sciences*, v. 2, p. 61–68, 1955.

YAO, L.; CHEN, Y. F.; YAN, H. The newsvendor problem with pricing: Extensions. *International Journal of Management Science and Engineering Management*, v. 1, n.1, p. 3-16, 2006.

ZHANG, G. P. An investigation of neural networks for linear time-series forecasting. *Computers & Operations Research*, v. 28, n. 12, p.1183-1202, 2001.

ZHANG, G. P. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, v. 50, p. 159-175, 2003.

ZHANG, Y.; GAO, J. Assessing the performance of deep learning algorithms for newsvendor problem. In: INTERNATIONAL CONFERENCE ON NEURAL INFORMATION PROCESSING, 24., Guangzhou, 2017. *Proceedings of the International Conference on Neural Information Processing*. Guangzhou: Springer, 2017. Vol. LNCS 10634, p. 912–921.

ZHANG, Y. et al. Online ordering rules for the multi-period newsvendor problem with quantity discounts. *Annals of Operations Research*, p. 1-30, 2020.