LUT SCHOOL OF ENGINEERING SCIENCE
INDUSTRIAL ENGINEERING AND MANAGEMENT

Utilizing machine learning in data-driven pricing

Koneoppimisen hyödyntäminen datavetoisessa hinnoittelussa

Bachelor's thesis

Joonas Niinistö

ABSTRACT

Author: Joonas Niinistö **Title: Utilizing machine learning in data-driven pricing**	
Year: 2020	**Place:** Lappeenranta
Bachelor’s thesis. LUT University, Industrial Engineering and Management 45 pages and 14 figures Examiner: Antti Ylä-Kujala	
Keywords: Dynamic pricing, revenue management, machine learning, data-analytics, data-driven, algorithms	

The rapidly increasing volume and variety of data and the continuous development of the technologies used for data processing have enabled the use of dynamic data-driven pricing. While pricing can be considered as one of the most important factors influencing profitability, many companies fail to take advantage of well-implemented pricing. This bachelor’s thesis aims to determine how data from different sources and the use of machine learning can be utilized in pricing.

At the beginning of the study literature related to pricing, machine learning and data analytics is examined. Study then explores the added value of dynamic data-driven pricing achieved through machine learning and large amounts of data. Based on these, dynamic data-driven pricing process is built at the end of the work.

The biggest benefit of data-driven pricing can be seen in reasonable reactions to changes in external and internal operating environment. In the best case, data-driven pricing allows to offer the right price to the right customer at the right time. Machine learning makes these price changes even more dynamic and allows to consider multiple influencing factors simultaneously. The benefits of dynamic data-driven pricing are significant, but especially if poorly implemented, it can cause major problems.

TIIVISTELMÄ

Tekijä: Joonas Niinistö

Työn nimi: Koneoppimisen hyödyntäminen datavetoisessa hinnoittelussa

Vuosi: 2020 **Paikka:** Lappeenranta

Kandidaatintyö. LUT-yliopisto, tuotantotalous.

45 sivua and 14 kuvaa

Tarkastaja: Antti Ylä-Kujala

Hakusanat: Dynaaminen hinnoittelu, koneoppiminen, tuottojohtaminen, data-analytiikka, datavetoinen, algoritmit

Data volyymien kasvu ja sen käsittelyyn käytettävien teknologioiden kehitys ovat mahdollistaneet viime vuosina dynaamisen datavetoisen hinnoittelun. Vaikka hinnoittelua voidaan pitää yhtenä tärkeimmistä liiketoiminnan kannattavuuteen vaikuttavista tekijöistä, monet yritykset jättävät hyvin toteutetun hinnoittelun tuomat hyödyt käyttämättä. Tässä kandidaatin tutkinnossa pyritään määrittämään, miten dataa ja koneoppimista voidaan hyödyntää hinnoittelussa ja mitä haasteita niiden käyttäminen ja käyttöönotto asettaa.

Työn alussa tarkastellaan hinnoittelua, koneoppimista ja data-analytiikkaa käsittelevää kirjallisuutta. Tämän jälkeen syvennytään koneoppimisen ja suurien datamäärien avulla saavutetun dynaamisen datavetoisen hinnoittelun tuottamaan lisäarvoon. Teoriaosuuden pohjalta työn lopussa määritellään dynaamisen hinnoittelun prosessi.

Datavetoisen hinnoittelun suurimpana hyötynä voidaan nähdä perustellut reaktiot ulkoisen ja sisäisen toimintaympäristön muutoksiin. Parhaassa tapauksessa datavetoinen hinnoittelu mahdollistaa oikean hinnan tarjoamisen oikealle asiakkaalle oikeaan aikaan. Koneoppiminen tekee hinnan muutoksista entistä dynaamisempaa ja mahdollistaa yhä useamman hinnoitteluun ja kysyntään vaikuttavan tekijän huomioimisen. Dynaamisen datavetoisen hinnoittelun tuomia hyötyjä voidaan pitää varsin kattavina, mutta varsinkin huonosti implementoituna se voi aiheuttaa suuriakin ongelmia.

TABLE OF CONTENTS

1 INTRODUCTION

1.1 Background

New technologies and new ways of collecting data are constantly evolving, which in recent years has led to a huge increase in the amounts of data. This data can be collected from almost all things and phenomena happening in the world, as a result nearly anything can be tracked and analyzed. (Wamba et al. 2017) According to Statista's (2020) predictions the growth in the amount of data is by no means declining, and the amount is expected to almost double over the next two years and increase fivefold over the next five years (Figure 1).

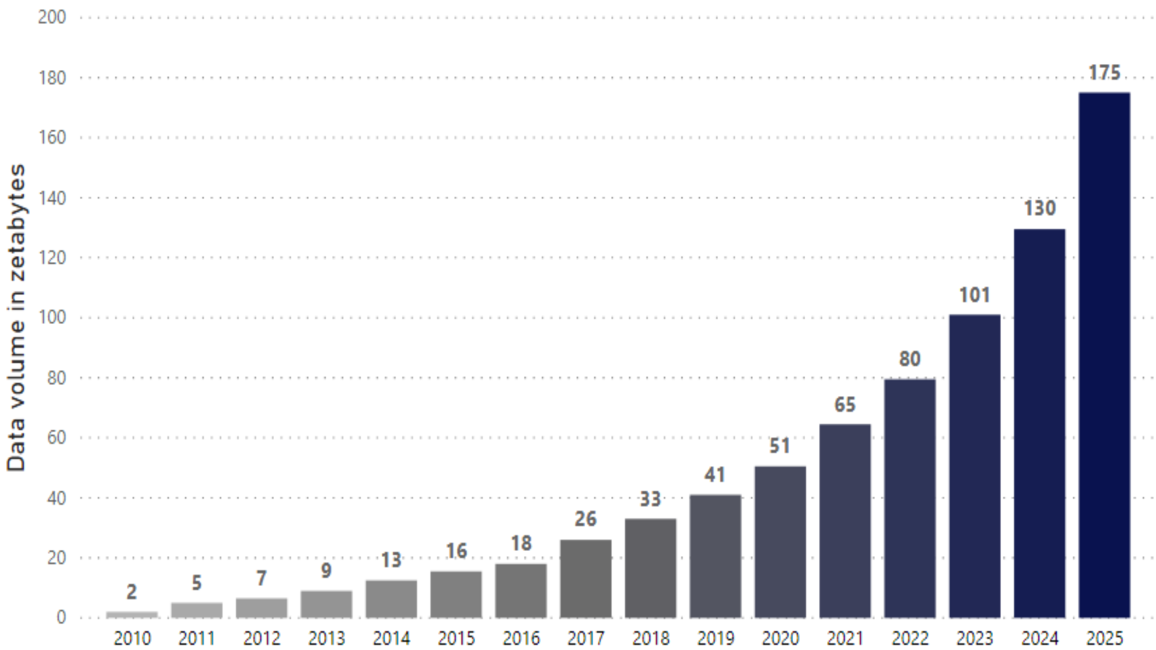

Figure 1 Volumes of data in zetabytes (Statista 2020)

At the same time, the world is becoming more globalized at an accelerating pace. As a result, competition is increasing in almost every industry (Bang & Markeset 2017). This is why companies have to be innovative and make efficient use of their resources, including data. By combining data from different sources, companies can make well-reasoned decisions and improve their competitiveness. It can be even said that data is becoming one of the key measures of whether a company will remain relevant through this revolution. (Bartosik-Purgat, & Ratajczak-Mrożek 2018)

One of the best uses for this data is in pricing. Pricing is one of the most important means of competition, because it simultaneously controls customer consumption as well as company's revenues. Data-driven pricing is not a new invention, but lately its use has become tremendously widespread (Roine 2019). Due to larger amount and easier access of data, companies can create pricing systems that take several relevant factors into account. In the best case these pricing systems can be fully dynamic, which means that the price is constantly personalized according to the situation and buyer. (Oliveira & Rana 2014)

In order for pricing to be fully dynamic, systems are needed to analyze and combine these pieces of information automatically. One answer to this problem is machine learning. Machine learning is one of the top trending research topics and its applications are constantly developed (Yao 2020). It is also a technology that many companies want to use without even knowing what it really means and what can be done with it. The potential of it is enormous, and for example fully dynamic data-driven pricing can be achieved with the use of machine learning. In pricing context machine learning can be used for customer segmentation, price staging, demand forecasting and automating different workflows. (Mupparaju et al. 2018)

1.2 Research questions and objective

This study is examining data-driven pricing and how machine learning can be used to make it more efficient and dynamic. Because the different types of markets differ greatly, the work focuses mainly on retail of consumer goods and business to business (B2B) markets are not widely considered. The study aims to give a good overview of how machine learning can be used to achieve dynamic data-driven pricing and what companies have to do to before gaining the biggest benefits of it.

As the basis of the thesis the basic concepts of machine learning, data-analytics and pricing theories are defined. Pricing theory is mainly limited to going through the pricing process and basic analysis. Other issues related to the pricing decision, such as the pricing strategy, are not carefully addressed. The purpose of machine learning theory is to give a sense of what it can achieve and what different types it involves, rather than going into mathematical and technical details. These delimitations have been made in order to keep the purpose of the work as clear as possible.

To support and guide the study following research questions are formed:

1. What are the benefits of machine learning in data-driven pricing and what challenges / risks it poses?
2. What kind of data is needed to create dynamic data-driven pricing system and where it can be collected?
3. What dynamic data-driven pricing requires of the company?

At the end of this thesis I present my own synthesis of the process required to create a dynamic pricing system as a process chart.

1.3 Research methods and structure

The literature review is done by using qualitative methods. Theory section is focused on pricing theory and understanding basic concepts of data, analytics and machine learning. At the end information from different sources used in the theory section are combined to outline the process of dynamic data-driven pricing.

Study begins in chapter two with theory about pricing process, models and factors affecting price in general. After pricing theory in chapter three prerequisites for dynamic pricing are introduced. This chapter defines how data can be gathered, what can be done with it and what is relevant in pricing context. Also, machine learning is introduced at a general level; basic concepts are defined, and the operating logic is briefly outlined.

In chapter four the key use cases of data in pricing are outlined. These concepts and definitions provide a general understanding of what data-driven pricing actually means and how it can be achieved. Report then examines how machine learning adds value to previously presented data-driven pricing and how it can make it fully dynamic. On the basis of these in chapter five the process of creating dynamic pricing system is outlined. This chapter also defines the requirements for creating a dynamic pricing system, as well as the things that need to be considered to make it work smoothly. Sixth and the final chapter recaps the main findings of the research and further discusses what research opportunities and areas for improvement the topic offers in the future.

2 PRICING

2.1 Pricing factors

It is clear that pricing plays a very important role in business, because it has direct and significant effect on profitability (Brennan et al. 2007). Assuming there is no loss on volume, on average one percent increase in price translates to almost nine percent increase in operating profits (Baker et al. 2014). Still most companies use only little time on their pricing and do not invest in finding the best possible price. According to a survey made in 2019, up to 78% of companies are aware that their pricing could be significantly better and based on more dimensions. (Caine et al. 2019) This is most likely because old manual practices for pricing make it really hard to find different pricing patterns that can create added value. It is basically impossible, especially for large companies, to manage the complexity of different constantly changing pricing variables, for thousands of products. (Baker et al. 2014)

Typically, price is set based on supply and demand (Figure 2), and for most products and services it is a critical order-winning and -qualifying factor (Sanders & Reid 2002, p. 34). This means that when there is high demand for a specific product with little supply, price can be set relatively high and vice versa. The point where supply and demand meet is called equilibrium, and it is also the point where price is usually set. (Salanié 1998, p. 2-3)

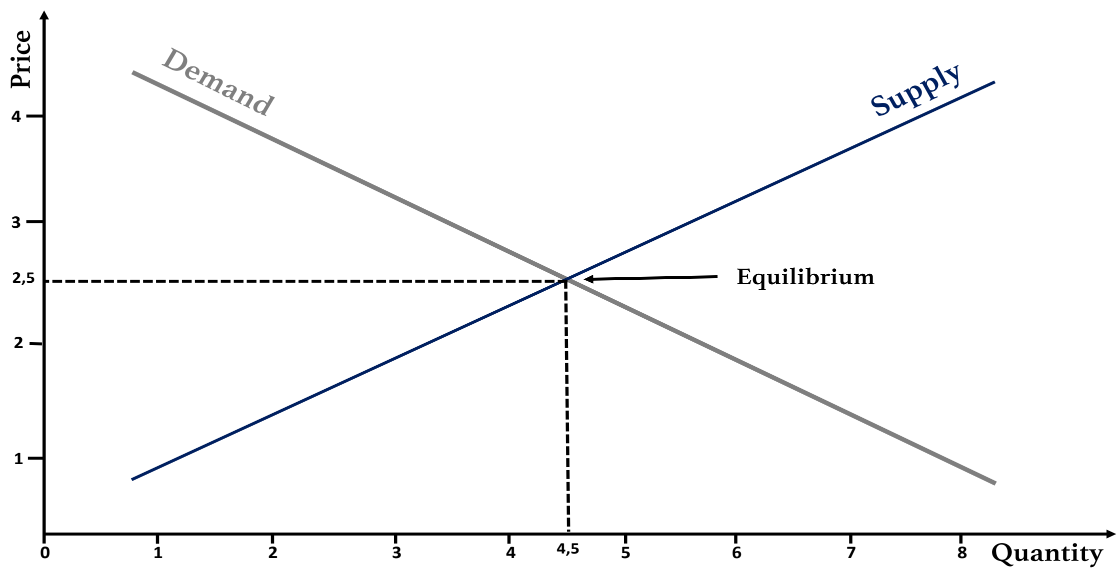

Figure 2 Law of supply and demand (adapted: Salanié 1998, p. 2-3)

Because both demand and supply are affected by a variety of factors that are constantly changing, pricing can be seen as a challenging process. There are also multiple variations of the same products that all have their own supply and demand, so the equilibrium can be really hard to find. (Mankiw 2006, p. 104 -107).

In simple terms, price defines the value of the product. It is also a factor affecting competitiveness, profitability and the positioning of the product. (Coles et al. 1999, p. 56-59) Because price can be seen as a measure of multiple things, companies have to precise when deciding price and pricing strategy (Barfield 2008, p. 32-33). Factors affecting pricing can be roughly divided into four categories: Customers, Company, Competition and Collaborators (Figure 3). (Brandt 2018; Fisher et al. 2017; Smith & Nagle 1994)

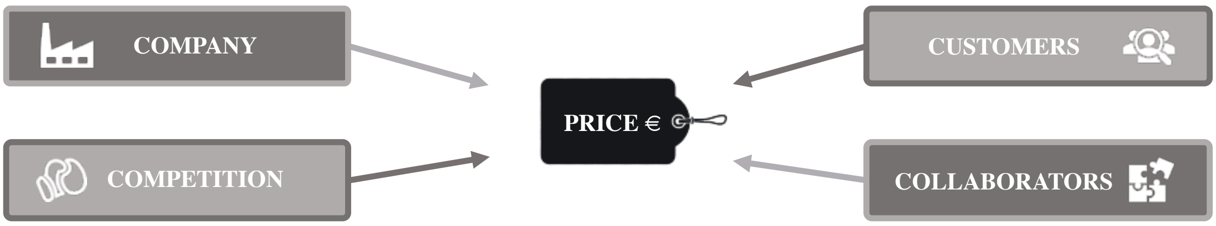

Figure 3 Factors affecting price

Company is naturally a thing affecting price. Company has costs, goals and strategy that all have strong influence on the final price (Kotler & Keller 2012, p. 411). Considering costs is especially important, because the difference between costs and price creates profit. Because profit margins can vary quite radically between different products, it is also very important to calculate costs correctly. (Smith & Nagle 1994)

The goal and strategy of the company also have a great affect to the price. If a company, for example, wants to achieve high quality brand, it is likely that the price is also in line with it. This is called value-based pricing. The basic idea in value-based pricing is that pricing is based on how valuable the product is perceived to be. (Hinterhuber, 2008) The opposite is a pricing strategy based entirely on production costs. This is called cost-based pricing and it is common in companies that aim for cost leadership with high production volumes and low margins. (Amaral & Guerreiro 2019)

Because price is one of the key order-winning and -qualifying factors, it is clear that **competition** has radical affect to prices. If there are multiple competitors with products of the same type, price plays an even more important role. (Kotler & Keller 2012, p. 430-432) This is why it is important to analyze competitor pricing constantly. Companies also have to analyze the quality of their services compared to other similar products or substitutes. This allows to place specific products in a certain price range compared to competitors. (Fisher et al. 2017)

Customers have different qualities depending on geographic location or socio-economic status. It is important to identify the customers characteristics, for example how much they are willing to pay, and whether they are looking for high quality brand or just the cheapest option. (Brandt 2018) Customers are also the only stakeholders that generate profits, so analyzing them is especially important. In addition to this, consumers' consumption patterns are constantly changing, so their analysis must be very dynamic. (Sharma 2018).

Collaborators and especially their prices contribute to price formation. For example, when selling products or services through third parties, these collaborators influence prices, because the margins of collaborators have to be made large enough to ensure good functionality, but at the same time these costs must be factored into the company's own prices (Keast et al. 2017). Similarly, raw material suppliers and subcontractors influence prices, as changes in their prices are directly reflected in the costs of the company. (Bradt 2017)

The product itself also has a major impact on how price changes affect its demand. Price elasticity is a measure to responsiveness of the demand changes when nothing but the price changes. It often depends on the importance and characteristics of the product or service; whether it is a necessity or a luxury. (Kotler & Keller 2012, p. 414) Products with high price elasticity are more sensitive to price changes and even a small increase in price can significantly reduce demand. This also means that when prices drop, the demand also increases significantly.

Price elasticity is calculated by dividing percentage of change in quantity demand by percentage of change in price. It is used to forecast how different price changes affect demand and for trying to find the best possible price with the largest profits. (Rissel 2011) It can be calculated quite easily based on historical sales and price data (Labandeira et al. 2017). Therefore, it also plays a key role in the data-driven pricing.

2.2 Pricing process

Because pricing factors vary quite much and they are constantly changing, it is important to have a clear process and infrastructure for pricing that takes these issues into account (Florian & He 2010). General pricing process (Figure 4) consists of three main stages: basic analysis, strategy matching and final price setting (Järvenpää et al. 2010, p. 198). This model describes the pricing process in its broadest framework and does not, for example, take into account the links between these stages. The goal of this process is to form a price that takes a variety of external and internal factors into account. The pricing process may vary depending on the pricing strategy, product or service, but the basic concept remains the same. (Frain 1999, p. 255-258)

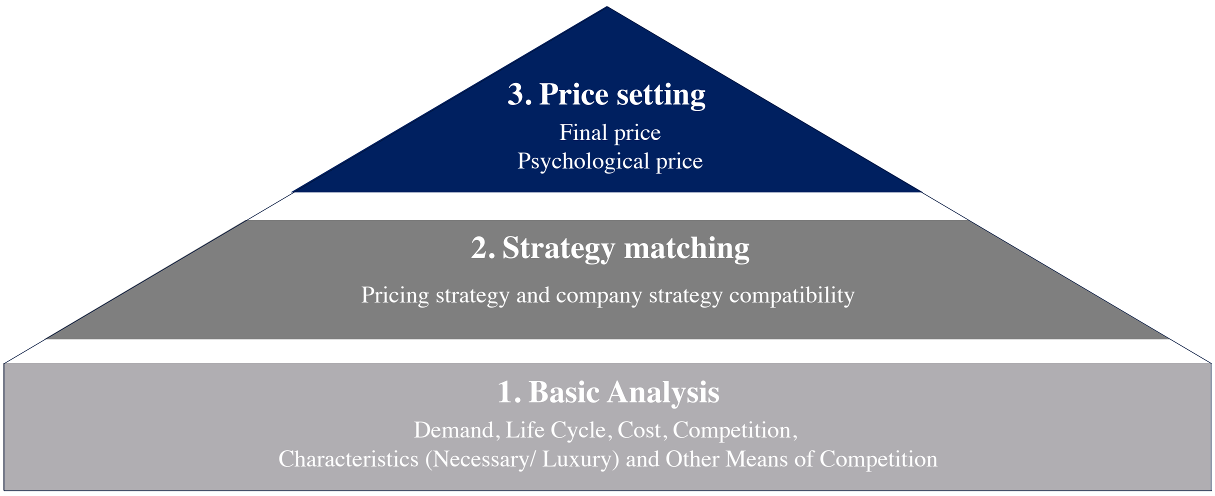

Figure 4 Pricing process pyramid (adapted: Järvenpää & Partanen 2010, p. 199)

The first stage of pricing process is **basic analysis.** In this stage different factors mentioned in chapter 2.1 are analyzed. The goal is to analyze factors affecting demand, cost and competition in the operating environment. These factors include all things affecting demand and costs, but also things such as price elasticity of demand and quality of service compared to other similar products or substitutes (Järvenpää & Partanen 2010, p. 197-198). At this stage, it is also important to analyze the factors associated with the product itself. This includes analyzing the whole life cycle of the product, and the predicted income and expenses incurred throughout it. (Grimmer et al. 2015)

In the **strategy matching** stage company's strategy and overall goals are considered from a pricing perspective. These should be in line with each other, because depending on the goals of the company, pricing can be very different. If a company is looking for the fastest possible growth, its pricing is very different from, when for example, trying to maximize margins (Järvenpää & Partanen 2010, p. 198). This step mainly creates certain thresholds for later discussed data-driven pricing systems and machine learning algorithms. For example, corporate strategy and goals can set the price not to exceed or fall below certain points. (Kotler & Keller 2012, p. 411**)** On the other hand, the choice of dynamic data-driven pricing is reflected at this stage, as it is also a strategic pricing decision.

Final stage of the process is the **price setting**, where the price for the product is determined. Successful pricing requires extensive knowledge of customers and their attitude towards the price and price changes. (Gorodnichenko & Talavera 2017) The price or the value that customer experiences is often different from the objective price of the product. Companies can significantly improve their price setting, if they know what the customer's view of the price is. Due to this it is really important to analyze what customers really want and what they are willing to pay, before making any pricing decisions. (Hinterhuber 2008)

At a more operational level, the pricing process can be divided into several stages, that go in either directions. Because the factors affecting demand and prices of competitors are constantly changing, the process of pricing needs to be continuous and iterative. In fact, the pricing process is an ongoing process that is repeated virtually all the time. (Fisher et al. 2017) Therefore, pricing should be dynamic.

3 PREREQUISITES FOR DYNAMIC PRICING

3.1 Data

Data is the most important prerequisite for dynamic data-driven pricing, because without it there is no empirical evidence of the operating environment. As mentioned on the previous chapter, pricing decisions are often the results of extensive analysis. Still most companies do not use their data at all to manage pricing. (Baker et al. 2014) This often leads to a situation where the same kind of products are sold with completely different prices, even within the same customer base. The profit margins can then lead to a situation where some products are sold at a significant loss. This is why it is crucial that companies start to use data analytics for pricing at an early stage to maximize profits. (Davenport et al. 2012)

Data is a collection of numbers, words, measurements, observations or even just descriptions of things (Cambridge Dictionary 2020). Like mentioned before new technologies make it possible to collect data from almost anywhere. This data can be in many different forms, but can roughly be divided into three types (Figure 5): structured, unstructured and semi-structured. (Marr 2019; Giudice et al. 2018)

Figure 5 Different data types

Structured data is the easiest to analyze, because it is in normalized form of columns and rows (Baars & Kemper 2008). Structured data is often managed using Structured Query Language, better known as SQL, which is a programming language for relational databases (Chatham 2012, p.7). Its normalized form makes it easy to analyze and study. This type of data is often encountered in databases or Excel spreadsheets. Examples of structured data include for example sales data, address details, demographic information or location data from smart devices. (Demetrovics et al. 2016)

Unstructured data does not have an associated data model so it cannot be stored in rows and columns. Examples of unstructured data include audio files, photos, social media contents, PDF files, written survey responses and websites (Giudice et al. 2018). Because unstructured data does not have a consistent structure, it is really difficult to manage and analyze. This is why most companies totally discard analysing unstructured data, and do not use it all. (Baars & Kemper 2008)

Semi-structured data is a third category, which in simple terms is a mix between structured and unstructured data. To make it easier to organize, semi-structured data has some structured properties, but it still does not conform to a structure that could be stored in a relational database. (Robb 2017) Good example of this could be an email, where the content is totally unstructured, but it still contains some structured data like email address and time stamp. (Marr 2019).

Unfortunately, no matter how much data there is, raw data in any form is usually just numbers or digits and does not really tell anything, if it is not processed. Because the amount of data is so large, not everything can be utilized and in fact most of it is irrelevant (Whitler 2018). So therefore, the real challenge is not the lack of data, it is the process of finding and processing relevant data that can help pricing decision-making. When data is somehow processed and prepared into a more understandable format, it turns into information. This information can be delivered for example in the form of dashboards or reports. (Baye et al. 2007) Data in this information that brings added value and can help decision-making is called insight (Dykes 2016). In other terms insights are solid conclusions made from information provided by data.

Insights are the important bits of information gained from analyzing large sets of raw data. These insights give companies direct and meaningful instructions or rules, that make decision making easier and rational (Hopkins et al. 2011). For example, data could show that some group of customers tend to shop significantly more at the beginning of each month. The insight from this would be that this group of customers tends to have more money available at the beginning of the months and hence majority of their shopping is done at that time. (Shim & Taylor 2019) These insights can also be used to improve pricing. Based on the example, pricing could be directed so that demand is more evenly distributed, or margins would be raised during higher demand. In the best scenario the process of finding these important insights is automated (Hopkins et al. 2011).

The thing that makes finding these insights even more challenging is the fact that not all data is created equal. Because the result of analytics is only as good as the input data, companies need to be careful what data they are dealing with. (Shim & Taylor 2019; Coursera Blog 2019) If the data is somehow bad, analyzing it also leads to bad reporting and bad decision-making. There are multiple examples in history where usage of bad or corrupted data has created astronomical problems. (Haug et al. 2011)

3.2 Machine learning

Machine Learning (ML) is a set of different algorithms and statistical models used in data science to automate, predict and solve a variety of problems and workflows (Louridas & Ebert 2016). In this chapter the basic concepts of machine learning are explained especially from the perspective of dynamic pricing. In general machine learning algorithms build mathematical models based on training data, in order to make predictions and automated decisions without being explicitly programmed to perform any tasks. (Shai & Shai 2017, p. 19 - 22) Although machine learning is a relatively difficult subject, it is based on fairly simple and logical mathematics (Vickery 2019). There are three main types of machine learning (Figure 6): supervised learning, unsupervised learning and reinforcement learning (Neller 2017).

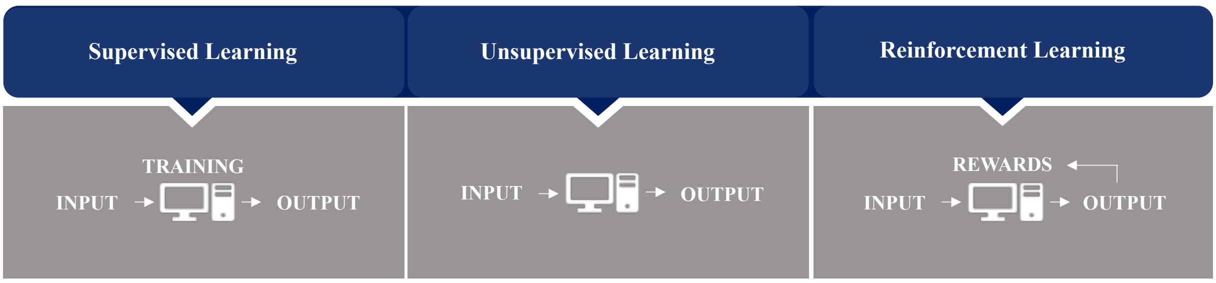

Figure 6 Machine learning types

Supervised learning is method where output (o) and input (i) are both known. Based on these pairs (i and o) machine learning algorithm learns which features and combinations of features produce a specific result. (Zhou 2018) When these algorithms have been trained with large sets of data, they can create really accurate forecasting models, where an output can be predicted based on any given input. (Norvig & Russell 2009, p. 695-697) In the context of pricing this can be used for example to predict how price changes affect demand. This is achieved by giving the model output and input pairs of price and corresponding demand. In this way, the model learns to interpret the relationship between them. (Lin 2006)

Unsupervised learning is method where only input is known, and it does not try to predict or give any specific output. The goal of the method is to understand the data: find patterns and structures. Unsupervised learning can be used for, for example, customer segmentation. (Vickery 2019) This is achieved by giving the algorithm multi-dimensional customer data. Based on this data, customers can be divided into multidimensional clusters. This is really helpful in pricing, because these algorithms can find specific segments of customers and offer the optimal price for them. (Toit et al. 2016)

Reinforcement learning is a method where the goal of the algorithm is undefined. It is the most technologically advanced of these and at the same time the most difficult to understand and interpret. (Farias 2010) It is used in continuous decision making in an unknown environment. The algorithm gets input and at first it randomly generates and output to it. Based on the quality of the output, algorithm gets rewarded. Reinforcement algorithm is programmed to seek as many rewards as possible, so it is constantly evolving in real time as it learns what kind of actions are good and bad. (Norvig & Russell 2009) This is particularly good in pricing, because like stated multiple times the factors affecting price and demand are changing constantly.

The biggest advantage of reinforcement learning is that it does not need a pre-specified model of the environment. Instead these relationships between states are learned through dynamically interacting with the environment. (Oliveira & Rana 2013) Reinforcement learning is particularly well suited to situations where external and internal issues are changing rapidly. These algorithms are able to dynamically optimize the price by considering the time remaining and, for example, the seats available on the aircraft. (Oliveira & Rana 2013)

Deep learning is even more advanced than these previously discussed models (Ning & You 2019). It is a subset of machine learning and it can contain supervised, unsupervised or reinforcement learning. This is why it can be used in quite a variety of ways to solve different types of problems. Deep learning is basically a mathematical function that tries to mimic human brain and make decisions in the same way. (Norvig & Russell 2009, p. 750-755)

Neural networks consist of multiple layers. These layers can be divided into three: input layer, hidden layer and output layer (Figure 7). (Nielsel 2019, p. 39-41) The first layer of the neural network processes input data and passes it on to the next layer as output. The second layer, called hidden layer, processes the information from previous layers and gives a new output based on them. This is repeated until a final result, such as a final price, is obtained. (Yiu 2019) A good thing from a pricing point of view is that use of deep learning and neural networks have allowed some use unstructured data, which is great because only 20% of data is structured (Giudice et al. 2018).

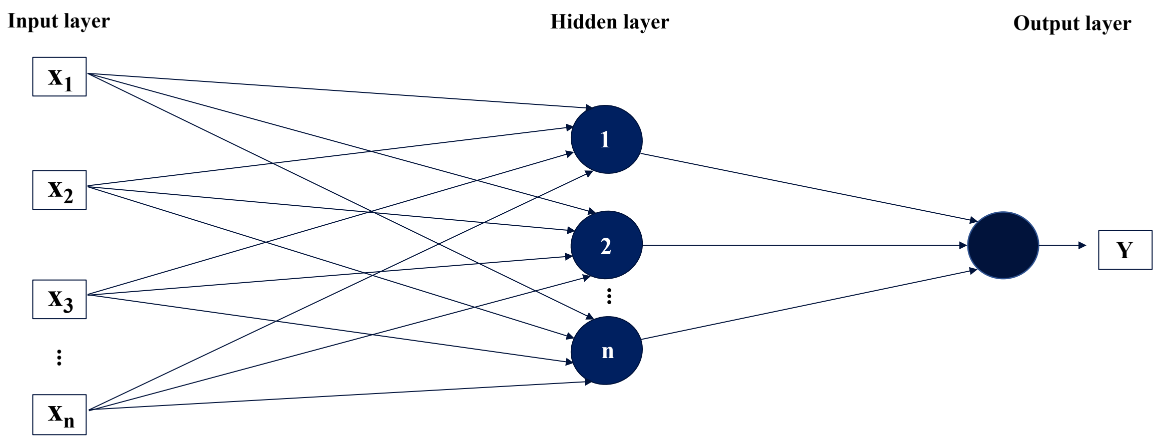

Figure 7 Neural network (adapted: Yiu 2019)

One common factor across all machine learning types is that the results of these algorithms are only as good as the input data they receive (Ahmadi-Jamid 2012). Secondly, all of these models improve as the amount of data they receive increases. Too little data makes predictions inaccurate because they are based on a small sample of the empirical world. (Géron 2019, p. 23-25) Therefore building a machine learning algorithm of any type is also a process where the importance of data collection, storage and processing is emphasized.

3.3 Data storing and collection

The more data companies have about customers and competitors, the more easily they can predict the future demand and interests of their customers (Baker et al. 2014; Lin 2006). Even though there is huge amount of data moving around, processing relevant data from all this can be surprisingly challenging. The sheer amount of available data creates a problem where it is hard to identify what data is relevant. Because collection and storing data is not free, companies have to decide what data to retain. This creates a problem between quality and quantity. (Dykes 2016)

Relational databases are great for storing structured data, but they cannot handle unstructured and semi structured data. This becomes a problem because to maximize data utilization, it should be centralized to a place where almost all data can be analyzed as a whole, not as separate compartments (AWS 2020). In the recent years data lakes have provided solution for this problem (Morrison & Stein 2014). Data lakes are repositories of data which store the data in their natural format. They can handle all kinds of data and for this reason they are great for storing all of company's data in one place (Giudice et al. 2018). By using data lakes companies have a place where all data can be accessed and analyzed. This enables data scientists to use this lake for broader analytics and machine learning. (Llave 2018)

Although data lakes are good for storing and analyzing large amounts of data, they have received quite a lot of criticism. This criticism is due to that some companies just dump everything into the lake and hope they will somehow benefit from it (Morrison & Stein 2014). As a result, companies lose knowledge of what data exists, where it is stored and in what form. In order to maximize the use of this data, only the necessary data should be stored in these lakes. (Llave 2018)

In order to use data for pricing, it must be first made available. This process of gathering and measuring information is called data collection. There are multiple channels where this data can be collected and usually the best results come by combining information from different sources. (Kabir 2016, p. 202-205) This data can be collected directly from customers, other companies or for example from company's own production lines. Additionally, to collecting data themselves, companies can acquire data from professional third-party data platforms that collect data from various sources. (Brandão et al. 2019)

By collecting for example data from closest competitors' companies can adjust their prices to the right level (Martinez-de-Albeniz & Tallur 2011). On the easiest case this can be done simply by scanning through the competitor websites, but in some cases, companies have to be more creative, because this data is not straightforwardly available (Fisher et al. 2017). At best, this data is easily accessible via APIs (Application User Interfaces), so it does not take much work to access (Mon 2018).

Internet of Things (IoT) solutions make it easier to measure internal issues and predict for example production costs quite dynamically. IoT sensors are usually small devices that can used to track different factors on real time. (Han et al. 2016.) They can, for example, have pressure, temperature, humidity or acceleration sensors (Stack 2018). Particularly good about IoT- sensors is the fact that the data it creates is usually quite easy to transfer and collect to almost any type of data warehouse (Kumar 2018). They can be used to collect data from the production lines, where they can be used, for example, to predict future maintenance costs. (Chabal 2018).

Customer data can be collected by directly asking customers, by tracking customer behavior indirectly or by appending other sources of customer data to company's own data (Dolfsma et al. 2018). These other sources of data include, for example, data from data marketplaces and open source data, that is freely available to all. Most obvious and efficient places for collecting customer data include activity and click data from websites and social media pages. (Deshpande 2019) Click-data data from websites is valuable in predicting consumption behavior, while data collected from social media can provide quite personal information about, for example, a person's hobbies or profession. There are a lot of tools available, like Google Analytics and Hotjar, that can collect this data automatically. It is also possible to generate personal location data with the same tools. (Google 2020; Hotjar 2020)

Old sales data is also thing that can be tracked and used to forecast future purchases. This data can easily be tracked at the whole company or item level, but tracking individual customer purchases requires some form of identification for the purchase (Doppsen 2020; Choy et al.

2018.). This identification can be achieved, for example, by requiring the use of a customer profile in an online store or the use of a loyalty card at a brick-and-mortar store (Ryan 2017). Note, however, that requiring authentication may, in the case of some customers, interrupt the purchase process, so it should be carefully weighed whether the benefits of authentication outweigh any potential drop in sales.

By collecting data from various sources companies can build personal customer profiles based on data and give customized prices to different customers or customer groups (Figure 8). The use of these profiles is not limited to pricing and they can be also used for variety of other purposes, like personalized marketing and communication. (Deshpande 2019) These customer profiles can be developed even further by using machine learning to be more dynamic and accurate. (Florenz-Lopez 2009)

Figure 8 Customer profiles (adapted: Doppsen 2020; Deshpande 2019)

It is also important to be aware that old data is not supposed to be deleted and it does not have an expiry date. In time series, especially in case of structured data, the structure of the data remains fairly constant, making older and newer data comparable. Because most forecast and machine learning algorithms are built by analyzing past data, old data should be retained to ensure the quality of these predictions. (Shim & Siegel 2007, p. 255-257)

4 DYNAMIC PRICING

4.1 Data-driven pricing

Data-driven means that almost all decisions and processes are controlled by data. This is most efficient when there is large amounts of data available. (Waller 2020) When companies have enough data from the right places, data can be the basis of almost all actions. Having much data gives powerful insight into the world, allows companies to make better pricing decisions and predict future demand. The recent growth in volumes of data and the development of tools for analyzing it, have made data-driven a part of everyday strategic and operational decision-making for many companies. (Ballou et al. 2018)

Once the right data is accessible, it can be used in pricing. Like stated on the chapter 2.2, pricing is a process of multiple steps that include analyzing different factors. Because data can give unambiguous answers, it is clear that it should be used in all of the pricing process steps. (Baker et al. 2014) Data-driven pricing means that most of the pricing process steps are compelled by data, rather than by personal experience, gut feel or old practices. The goal of data-driven pricing is that the right price is offered to the right people at the right time. (Oliveira & Rana 2014) This enables higher profits, higher demand and improved customer experience (Roine 2019). Modern methods of data collection and analysis allow retailers to consider factors such as competition, season, ERP-data, exchange-rates, operating cost, demand, company objectives and even weather. These factors give good insights of the operating environment and internal constraints. (Mon 2020)

Competition is further emphasized in the business to customer (B2C) markets, especially in retail, because there are many competitors who offer exactly the same products. Also, customers can easily compare prices, and there are even ready-made websites that do the comparison automatically for them. (Wang et al. 2018) As a result, certain retail chains have even switched to pricing almost entirely based on competitors' pricing data. However, pricing on the basis of competitors alone is really risky because it does not take into account the company’s own costs or strategic goals. (Fisher et al. 2017, p. 1-4; Mago & Pate 2009)

Demand for certain products is very season dependent. In the case of such goods, demand data from earlier seasons should be utilized and priced at the highest possible level permitted by price elasticity. Certain local events can also cause huge spikes in demand. (Wang 2014) In addition to taking such events into account when ordering and increasing inventories, products must be priced in line with this growing demand. (Saha & Basu 2010) As companies aim to balance demand on an annual basis as much as possible, off-season pricing should also be lower in accordance with the law of supply and demand.

In addition to seasonal fluctuations, on a local level weather is a major factor affecting demand, so therefore it should also play a role in pricing. As a simple example, sunny weather can boost demand significantly for certain products (Figure 9), such as sunscreen or ice cream, while rainy weather is seen as demand peak for umbrellas. (Ylinen 2014) The rise in demand and, in particular, the acute need will also allow the price to be raised (Chenavaz et al. 2019).

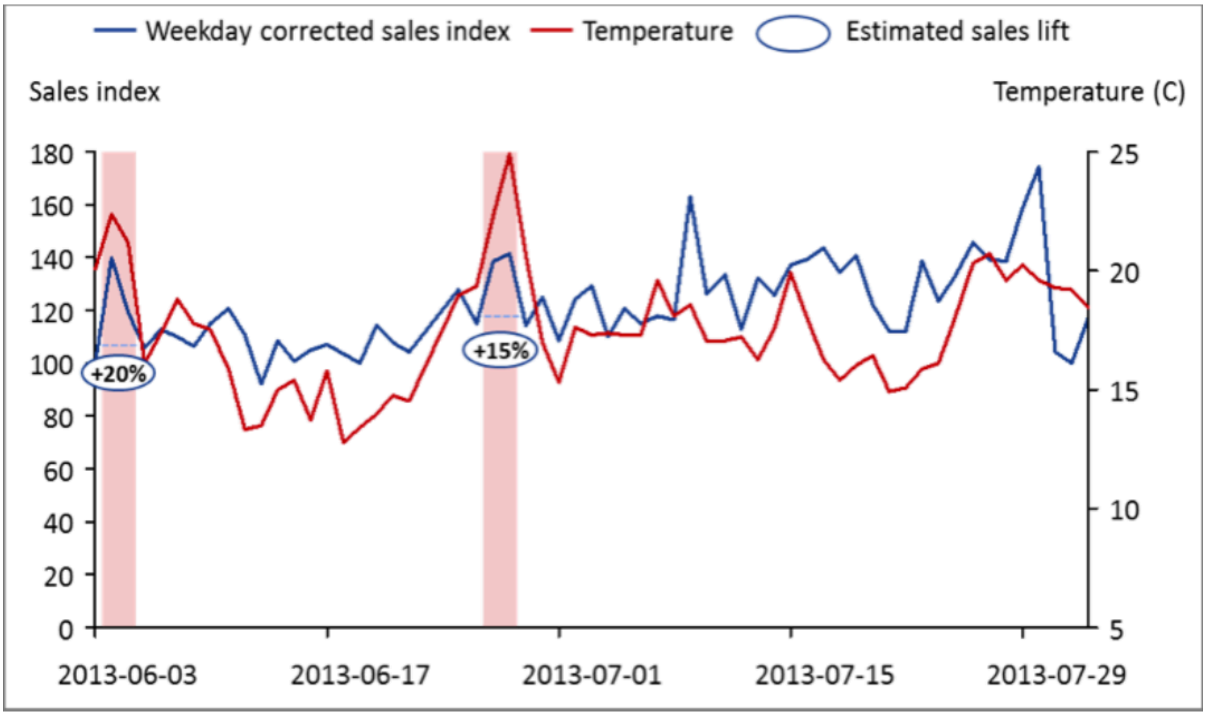

Figure 9 Example of weather impact in sales of all fresh products. (Ylinen 2014)

Data from inventory levels can also be easily utilized in pricing. Companies' own Enterprise Resource Planning systems (ERPs) usually provide a real-time view of inventory levels and inventory turnover rates. (Oracle 2020) On a general level, it can be stated that stocks are desired to move as quickly as possible. Businesses can easily analyze this data and for example lower their prices when the stock levels are rising for a specific product. This also works the other way; when inventory levels are too low and there is too much time for the next delivery, prices can be raised to reduce demand and increase profits. (Tiwari et al. 2018) The same type of optimization works also for companies in the service sector that don not have physical stocks, but a fixed deadline by which the products are forced to be sold. This includes business from a variety of industries, including for example airlines and hotels. (Oliveira & Rana 2013)

Since the difference between the selling price and costs brings profit to the company, these costs should be monitored really closely. Company's data on variable and fixed costs can be easily utilized. However, this requires good groundwork in cost accounting to allocate the right costs to the right products (Kotler & Keller 2012, p. 414-415). Production costs can be monitored quite dynamically with the help of IoT -sensors and, for example, cost-causing equipment failure can be predicted. (Chahal & Prabu 2018). Own ERP systems also provide very real-time data on suppliers' prices, which of course are also reflected in to the company's own costs. (Oracle 2020; Bradt 2017)

Exchange rates affect the company's income significantly, especially for companies strongly engaged in international trade. Exchange rate data is available really easily and does not require much analysis. By considering this data, companies can build pricing systems that systematically update prices according to exchange rates. (Roine 2019) It should also be noted that when making purchases, companies must also take exchange rate changes into account, as they have a direct impact on the company's costs.

As a whole demand is one the most critical things affecting price (Kotler & Keller 2012, p. 413-414). According to the law of supply and demand, when there is higher demand businesses can easily charge higher prices. (Sanders & Reid 2002, p. 34) By actively analyzing demand and different factors affecting it, data-driven pricing can lead to much higher profit margins or increased sales quantities. Excessive price increases can, of course, lead to a fall in demand in accordance with price elasticity, so companies must closely monitor all of these factors collectively, not individually. (Labandeira et al. 2017)

4.2 Utilizing machine learning in data-driven pricing

Like mentioned in chapter two, pricing is affected by multiple factors that are constantly changing. When the world is becoming increasingly globalized, companies have to be dynamic to changes happening all around the globe all around the clock (Bang & Markeset 2017). Because it is quite impossible to price products manually depending on various features, companies need automated systems for pricing. Such a dynamic data-driven pricing system is possible to achieve through the use machine learning. Dynamic data-driven pricing differs from just data-driven pricing in the sense that the systems are able to perform the pricing completely independently and with constant price changes. (Roine 2019; Javanmard & Nazerzadeh 2019)

Data collected from customers can be used to create multi-dimensional customer segments (Kelly 2003). Machine learning enables to dynamically update these segments, but also the use of these segments in pricing. This promotes understanding of customers which in turn could be used to increase revenue of the company. (Hess & Story 2006) Predictive models can be used for predicting the future consumption behavior and purchase patterns of these different groups (Lin 2006). This gives a relatively accurate picture of what kind of demand is expected and what kind of prices should be set. This chapter introduces a few machine learning techniques that can be used to make pricing dynamic.

One effective way to segment customers is k-means clustering algorithm. First the number of clusters needs to be decided. The decided number of clusters is variable *k* and these clusters all have a center point called *mean*, thus the name k-means. K-means algorithm then divides all the inputs into clusters, so that every input belongs to the cluster with nearest mean. (Shai & Shai 2017, p. 313) Algorithm then starts to optimize itself so that the locations of the means change so long that the total distance of each input to the mean is as short as possible. (Zhang et al. 2018)

The figure below (Figure 10) is an example result of k-means clustering. In this case the number of k:s is 3, which result in 3 different cluster. In pricing context these could be 3 different customer groups that all have different consumer habits and ability to pay. The two dimensions, in this case, could be as follows: x-axis = income and y-axis = activity on online store. In good k-means algorithms, there are far more of these dimensions to achieve the widest possible customer understanding. Customer segmentation like this is actually one of the most used applications of k-means clustering. (Sagar 2019)

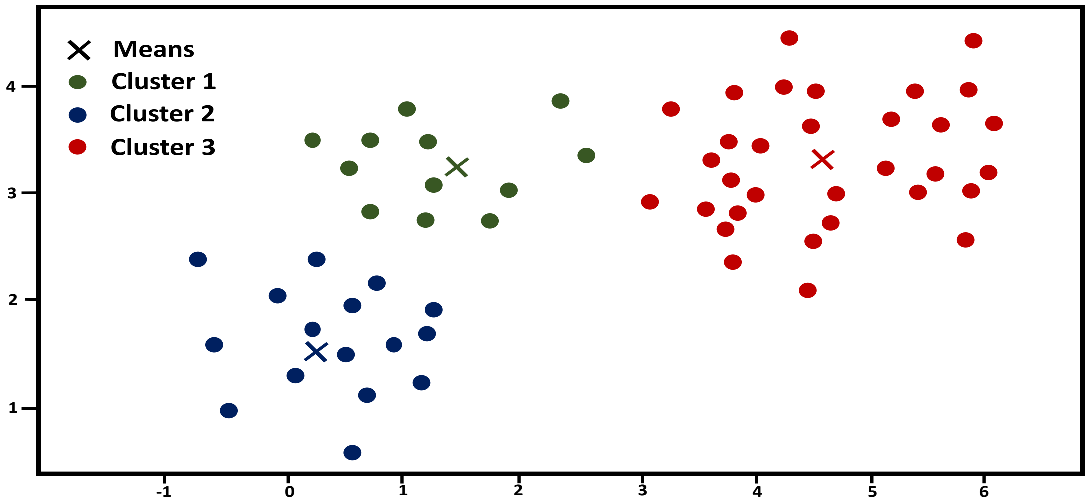

Figure 10 K-means clustering example (adapted: Sagar 2019)

Machine learning also allows forecasting future purchases of specific customers based on historical data. Decision tree learning algorithm is one of the predictive models widely used to predict future outcomes of X based on historical input and output data. In decision tree the outcome comes from X travelling from a root node of a tree to a leaf which contains the output (Gupta 2017). At each node on the decision tree the successor result is selected based on splitting. This splitting usually happens based on one of the features of X or in some cases on a predefined set of splitting rules. (Shai & Shai 2017, p. 250 - 255)

The figure (Figure 11) looks at an example of whether a certain customer is buying a product at a certain price or not. In this case the outcome leaf contains value of "SALE" or "NO SALE". To check if a given customer made a purchase or not, the decision tree first examines the sex of the customer. At this point, the tree splits into two different paths where the same questions are asked, yielding different results. (Jin et al. 2018) Next, the tree turns to examine the age of the customer. In the female branch when the age of the customer is under 20 years, the decision tree predicts that the customer does not make a purchase. If the age is over 20, the tree turns to examine whether the customer is student or not. If the customer is a student, no purchase is made, and if the customers is not a student purchase is made.

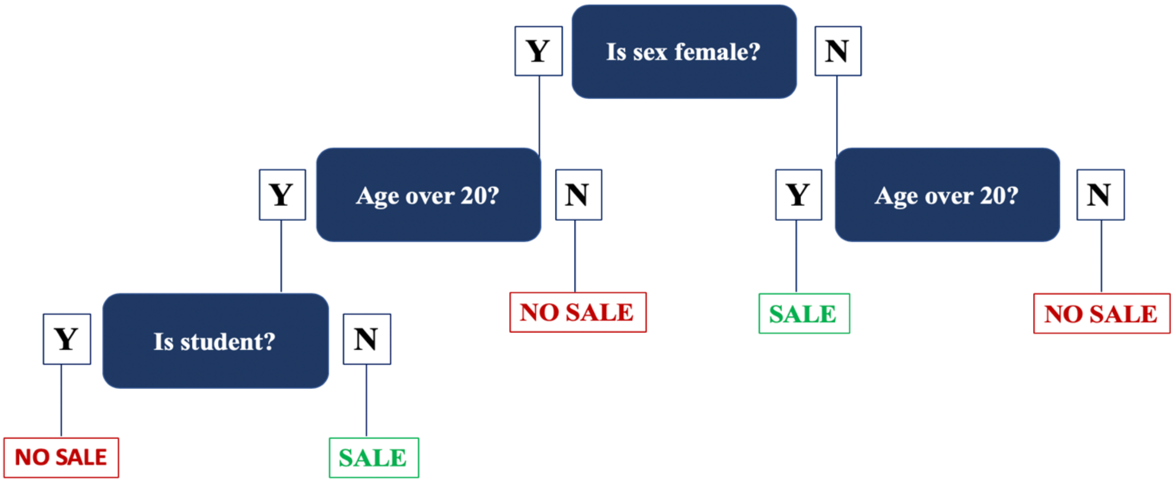

Figure 11 Desicion Tree Learning Example

The most important thing that machine learning can achieve is dynamic price fluctuation. This can be achieved through reinforcement type machine learning. Because reinforcement algorithms are designed for an environment where many external and internal factors are constantly changing, it is just right for pricing (Oliveira & Rana 2013). Such algorithms can be trained with a huge amount of history data and optimized to set the right price. (Yiu 2019) These algorithms can consider many different dimensions on the basis of which a decision is made. Because it is a machine, it is able to make these decisions with a continuous feed and take into account more dimensions than a human brain could even handle. (Jin et al. 2018)

The figure (Figure 12) looks at an example of dynamic price setting. This figure presents a simplified neural network that takes four inputs and sets the price based on them. Prior to deployment of this model, it is trained with training data to make the best possible pricing decisions. For example, based on the resulting margin and demand, it has received rewards and improved. In this particular network the inputs are estimated demand, customer segment, weather and inventory level. In a real model there should be considerably more of these factors. What is particularly good about this model is that when the factors change, it is able to offer a new optimized price immediately. (Oliveira & Rana 2013)

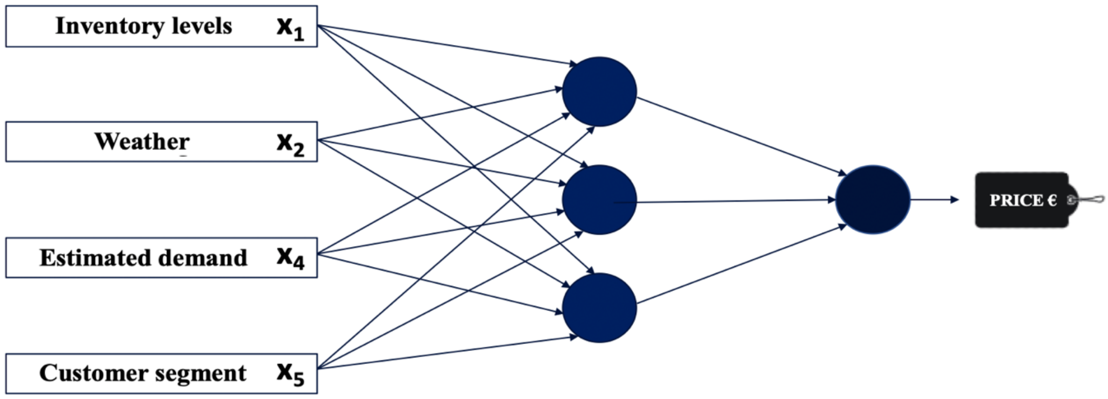

Figure 12 Pricing neural network

When these algorithms are combined with company's pricing systems, pricing becomes truly dynamic. Dynamic pricing is not limited to online shopping, and some companies have introduced electronic price tags in brick-and-mortar stores, which are updated as dynamically as online (Marcarelli 2019). However, dynamic pricing is still much more common in online stores, especially in the sale of transportation such as airline tickets. (Fiig et al. 2018)

4.3 Benefits and risks of dynamic data-driven pricing

The benefits of machine learning in data-driven pricing are clear: prices can be controlled dynamically and efficiently; the right price can be found for the right people and demand forecasting becomes more accurate. These price changes and dynamic forecasts can also be used to lower inventory levels and speed up their turnover. All of these together make company's revenue management more efficient and productive.

Finding the right price for right people at the right time is important, because it enables maximation of both revenue per customer and overall demand. Because machine learning algorithms can provide very accurate predictions, the price can be set exactly at the level where the purchase is still made. At the same time, dynamic segmentation enables people to be divided into customer groups with different consumption behaviors. Even between these groups, it is possible to regulate the price to just the right level. At best, these prices can be adjusted to a fully personalized level based on customer profiles and individual customer price elasticity. The biggest advantage is that things that would not be taken into account without machine learning can be included in the pricing decision, as well as the combination of many different dimensions that all together determine how much is bought, when and at what price. (Oliveira & Rana 2014; Martinez-de-Albeniz & Tallur 2011)

However, dynamic pricing does not bring only positive things to company's operations. Especially in situations where it is poorly implemented, the results can be really bad. Poor pricing decisions can lead to significant revenue losses, which is why companies cannot really afford big mistakes. The clear risks of dynamic pricing are at least: forecasting errors, pricing errors in general and excessive price changes. Poorly implemented demand forecasts are simply bad, which is why the decisions made on the basis of them are not good either. Pricing based on a poor forecast can therefore lead to far too low a margin, or a far too high price, which is reflected in even lower demand. The big risk here is also the fact that the company could also lose its future customers, if customers start boycotting the company because of its excessive prices (Akman & Garrod 2011).

Pure pricing errors are also a risk created by dynamic pricing. While this risk also exists in the same way when pricing in traditional ways, it is accentuated in dynamic pricing systems driven by machine learning, because they are often "black boxes" whose operation is challenging to interpret (Gabel et al. 2014). Particular important is how the various factors are weighted in the pricing decision. For example, a customer's past buying behavior is likely to have a greater impact on his or her future purchases than exchange rates, but algorithms may mis weight these, resulting in predictions that are far from correct. Thus, the price created by the dynamic pricing model can be completely wrong, resulting in multiple customers not making a purchase at all.

Excessive dynamics in pricing can also cause problems. Although in theory the best result is obtained when the price is constantly changing according to the situation, this can cause great irritation to customers. If prices change too fast, up to several times an hour, some customers may find the pricing unfair or postpone the purchase decision by waiting for lower prices. (Dholakia 2015) This, of course, is not a desirable situation because companies want demand to be as even as possible and sales prices to bring in the highest possible profit margins. Demand as a whole may also suffer over a long period of time as customers move to other competitors whose prices remain more stable.

5 CREATING DYNAMIC PRICING SYSTEM

5.1 Dynamic pricing system suitability

A dynamic pricing system is particularly well suited, and may even be necessary, for companies that have so many products on sale at the same time that their pricing would be completely impossible or at least really poorly executed without been dynamic and data driven. Some retailers like Amazon sell over 12 million unique products at the same time, so it is understandable that automation and machine learning are necessary (Dayton 2020). It can also be said that the benefits of dynamic pricing are especially emphasized when the same company has several products, so smaller players in particular should consider whether dynamic pricing is profitable.

Achieving dynamic pricing also requires the company to be systematic and willing to implement it. The system will not succeed on its own and without careful analysis of the business environment, pricing strategy planning and understanding of the factors influencing demand and price, good results will not be achieved (Florian & He 2010). In order to be successful, a company must have a lot of know-how on many different issues related to pricing, machine learning and data. Alternatively, company can hire a consultant to handle the matter on their behalf.

The importance of data volume is also emphasized as pricing becomes more dynamic. For large players who also have thousands of products, this data is more readily available. Companies need to collect huge amounts of external data to understand customers and the operating environment. (Géron 2019, p. 23-25) Internal data must also be available so that pricing can also be done according to inventory levels and costs. For this reason, it is conceivable that the creation of a model is profitable or even possible only for companies whose sales volumes are large enough. The predictions of several machine learning algorithms are largely based on historical data, so without this, at least prediction-based pricing model, it is not even worth considering.

As often emphasized, data quality is of great importance in whether it is possible to achieve successful dynamic data-based pricing. Without good data and its systematic collection and preservation, the creation of a dynamic data-driven pricing model is not possible. The quality of the data should also be constantly monitored, not just when implemented in the system. So, requirements for the company include diverse and planned data collection and creation of a functional data storing environment where multiple sources and types of data can be analyzed. A good choice for this is a data lake. (Llave 2018)

The figure below (Figure 13) summarizes the applicability of dynamic data-based pricing in different types of organizations on the basis of previously defined requirements. In summary, the more data a company has from different products, as well as expertise and resources, the more likely it is that dynamic pricing is a good option.

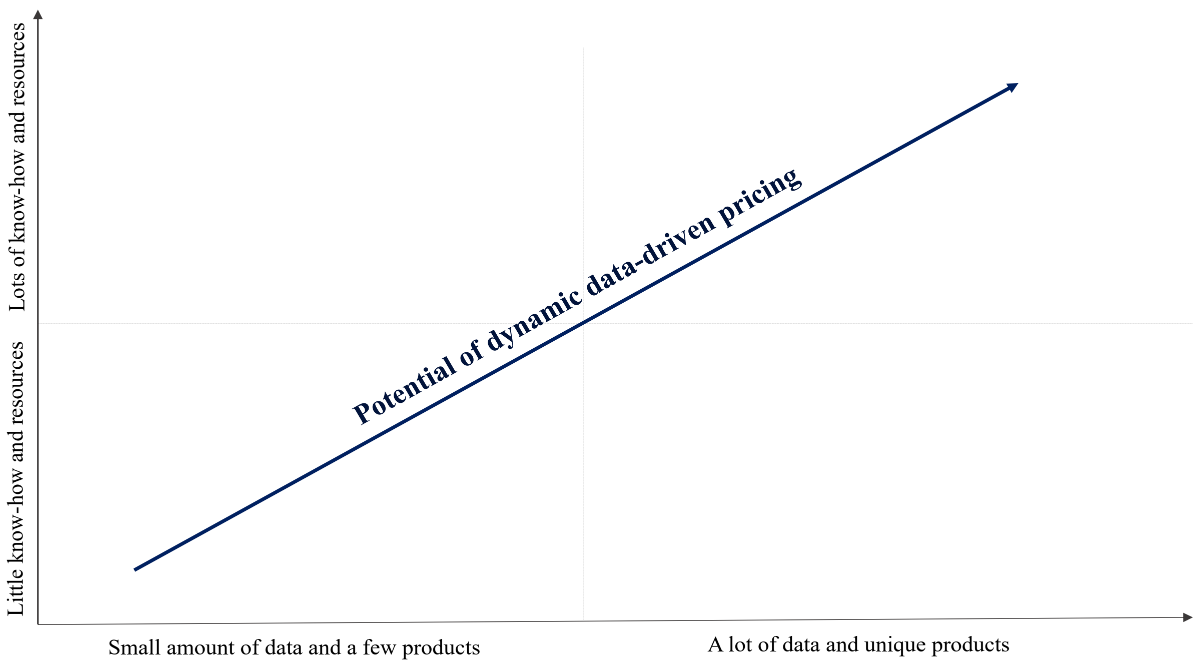

Figure 13 Potential of dynamic data-driven pricing

5.2 Dynamic pricing infrastructure

In order to implement dynamic data-driven pricing system as efficiently as possible, its implementation process should be systematic. For this reason, a process description is created, intended to guide the process. This process, called dynamic pricing infrastructure, is composed of 11 steps (Figure 14). These steps are generally performed in sequence but, if necessary, some of them can be iterated in reverse.

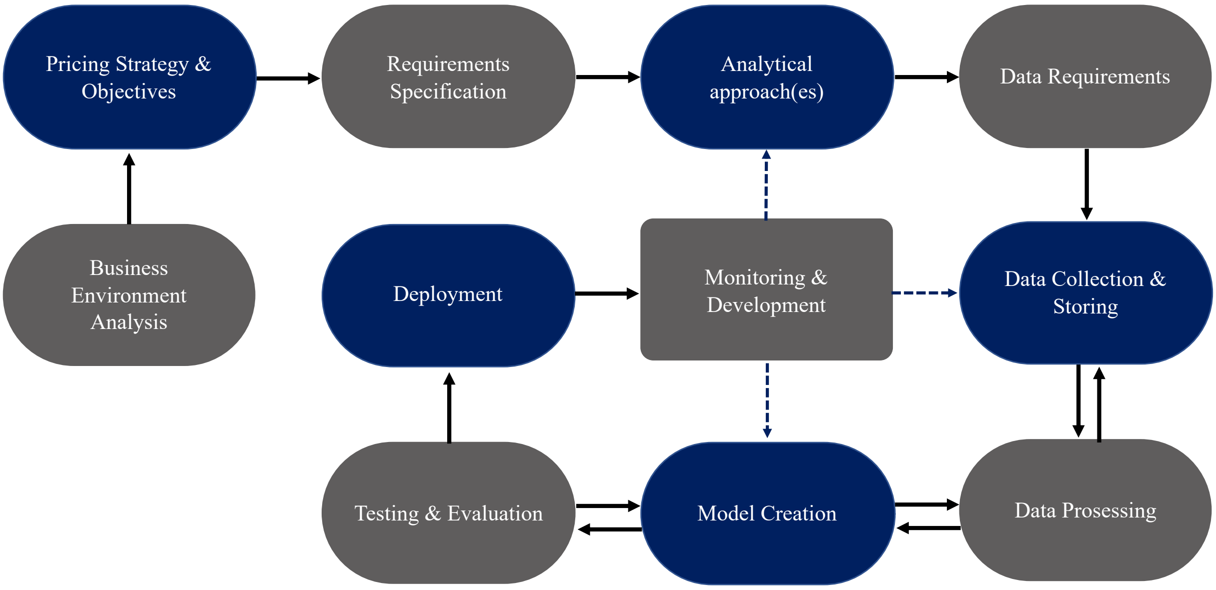

Figure 14 Dynamic pricing infrastructure

First step of the process is **business environment analysis.** At this stage, the purpose is to conduct a comprehensive analysis of the company's operating environment as well as internal matters. The phase analyzes the operating environment, competition, products offered by competitors and the company's own strengths and weaknesses. The purpose of the analysis is to produce comprehensive information, that is useful to the company's pricing strategy and other stages of the entire process. The analysis is very important because it largely guides what kind of strategy a company should start planning and what kind of goals to pursue.

The **pricing strategy and objectives** is the stage that guides the pricing of the entire company. It defines what the company's pricing is intended to achieve and how it will be achieved. In general, the choice of dynamic pricing is a strategic pricing decision that is made at this stage. On the basis of the business environment analysis, it must be decided whether there is a need for dynamic pricing. The strategy also sets certain thresholds for pricing, such as minimum and maximum margins.

The third step is **requirements specification**. The purpose of the step is to create a list of all objectives and requirements to be achieved by the final output. The phase includes both functional and non-functional requirements. Functional requirements describe what the system should be able to do. This can be, for example, producing forecasts every hour or optimizing the price at least 2 times a day. Non-functional requirements can be qualitative or resource related. Qualitative requirements define how reliable, secure and efficient a dynamic pricing system should be. Resource related requirements on the other hand define how much time and money company is willing to invest in implementing and maintaining the pricing system.

The **analytical approach** stage defines what kind of statistical and machine learning methods are used. The purpose is also to find out what are the main problems that need to be solved, because the problems define what kind of patterns need to find. This must be done judiciously because methods used greatly affect the outcome. If the goal is to find customer segments and price based on them, a descriptive unsupervised approach may be required. But if the goal is to predict future purchases and demand, a predictive model like the decision tree might be better. However, the best results are obtained using several different models to, for example, classify customers and predict future demand at the same time.

When analytic approaches are decided, **data requirements** should be defined. Based on the selected analytical methods, it can be determined what data is needed. At this stage, the quality requirements and formats of the data are defined, but also the sources from which it is collected. It is important that all the necessary data is identified and found because without data machine learning algorithms are completely useless. Once the data requirements have been defined, **data collection and storing** can start. However, it should be noted that this phase in particular is continuous as new data is generated continuously. Since dynamic pricing requires as much real-time data as possible, it is natural that its collection should be continuous. Step thus essentially means automating the collection of data for a particular data warehouse or data lake.

In the **data processing** stage data is processed and prepared for the machine learning algorithms. At this point, the key is to make sure the data is of high quality and in the right format. Data exceptions are deleted because of their misleading nature, appropriate column names are given, and data is generally cleaned. This step is continuous like the previous step because the data must always be in the correct format, not just when creating the model. The importance of this step cannot be stressed because poor data quality is directly reflected in poor results in machine learning.

After data is processed **model creation** can start. At this stage, machine learning algorithms and the resulting pricing system will be developed. Machine learning algorithms based on defined analytical approaches are created to track different factors and events in the operating environment as well as within the company. For example, in the case of predictive model such as decision tree, the model is trained and optimized using historical training data. Algorithms monitor these things in real time and make pricing decisions dynamically. Based on these decisions, the price is regulated as often as specified in the requirements definition.

After model is created it is **tested and evaluated**. At this stage, the results produced by the machine learning algorithms are tested and the results produced by the pricing system are evaluated. The results produced by the predictive models can be compared with real-world empirical data and, in general, the pricing decisions produced by the system and the resulting margins can be estimated. At this stage, the added value generated by the model can be easily assessed and, for example, it can be compared whether the sales revenues are higher at the prices generated by the model. However, it should be recognized that at this stage the model has not yet been implemented but will be run as a test alongside existing systems. This is important because, as mentioned earlier, poor pricing decisions can lead to really poor results, so testing should take a lot of time, to see the impact of various changes. If the results are poor and do not meet the objectives set by the requirements specification, the process will have to be iterated backwards.

Once testing has shown good results the model can be **deployed**. This means an implementation process where the model is integrated into an existing pricing system. In this situation, the machine learning model begins to control prices dynamically. In online stores, price fluctuation is easy, but in brick-and-mortar stores, this requires digital price tags, as manual dynamic pricing of each product would be impossible.

Once the model is in use, a continuous **monitoring and development** phase begins. The results can be easily tracked, for example, by creating a dashboard. Issues to be monitored include, for example, product or product group margin, number of price fluctuations, demand forecast in relation to actual demand, and general comparison of results to non-dynamic pricing. If problems arise during monitoring and, for example, the forecasts are not correct at any level, the process will have to be developed. In general, the model needs to be continuously developed as better technologies evolve and new data sources emerge.

6 CONCLUSIONS

The object of the thesis was to study how companies can utilize data and machine learning in their pricing. In addition, the benefits that companies can derive from dynamic data-driven pricing systems and the challenges that machine learning and dynamic pricing can pose, were explored. The purpose of the theory was to provide a comprehensive scientific basis and help outline the perspective used in the work. Based on the theory, a process description of dynamic pricing was created. The study answers the three research questions as follows:

1. What are the benefits of machine learning in data-driven pricing and what challenges / risks it poses?

The benefits of machine learning in data-driven pricing are clear: prices can be controlled dynamically and efficiently; when demand forecasting for specific customers becomes more accurate, the right price can be found for the right people. When these are integrated into a company's pricing systems to make dynamic changes, the need to vary prices manually is also reduced. This will make the company more efficient and, if the pricing is successful, it will also increase margins and overall sales. On the other hand, dynamic pricing brings its own problems, because some customers may get annoyed by excessive price fluctuations and find it unfair. Machine learning algorithms are also not 100% reliable and they can produce bad predictions and bad decisions based on them. The challenge is also to find the right weightings between the factors on which pricing is based. If these are wrong, pricing will most likely fail completely.

2. What kind of data is needed to create dynamic data-driven pricing system and where it can be collected?

For machine learning algorithms to be effective, a lot of data is required. The data must be in the correct format and of good quality in order for the results to be good. As demand is affected by many different factors, data must also be collected on a number of different issues so that almost everything can be taken into account. Pricing in particular consists of the customer's understanding, so the data collected from customers is of great importance. This data can be collected, for example, from the company's own websites, directly from open source sources or purchased from the data marketplaces. Company’s own systems also provide a comprehensive amount of data on inventory levels and production costs that should be utilized in pricing.

3. What dynamic data-driven pricing requires of a company?

In addition to collecting good and diverse data, dynamic data-driven pricing requires the company to do careful pricing strategy planning. A good strategy is the result of extensive business environment analysis and understanding of the factors influencing demand and price. This requires the company to have a good understanding of the business environment, its own operations and competitors. Company must also have a huge amount of knowledge and understanding of information technology intensive issues such as machine learning, data analytics and system integrations. In general, dynamic data-driven pricing is not easy and fast to achieve, so the implementation also requires a lot of resources, commitment and time.

The pursuit of dynamic pricing is likely to become more common in the future, because the comprehensive benefits it brings, such as higher margins and growing demand, are really attractive. The rapidly growing amount of data and evolving technologies such as machine learning make it even easier to achieve dynamic pricing. It is therefore very likely that in the future dynamic pricing will not be uncommon and its use will be rather normal. As machine learning and artificial intelligence develop further, price optimization will become more accurate and dynamic. In the future, the role of neural networks and deep learning in particular will be emphasized in pricing. Particular efforts should be made to examine these and their potential from a pricing perspective. It is also certain that in the future there will be more software for automated dynamic pricing, because the market potential is huge.

REFERENCES

Ahmadi-Javid, A. 2012. Application of entropic value-at-risk in machine learning with corrupted input data. 2012. *11th International Conference on Information Science, Signal Processing and their Applications (ISSPA).* Pp. 1104-1107.

Akman, P. & Garrod, L. 2011. WHEN ARE EXCESSIVE PRICES UNFAIR? *Journal of Competition Law & Economics.* Vol. 7, no. 2, p. 403.

Amaral, J. V. & Guerreiro, R. 2019. Factors explaining a cost-based pricing essence. *Journal of Business & Industrial Marketing*. Vol. 34, no. 8, pp. 1850-1865

AWS. 2020. What is a data lake? [WWW-document]. [accessed 26.2.2020]. Available: https://aws.amazon.com/big-data/datalakes-and-analytics/what-is-a-data-lake/

Baars, H. & Kemper, H. 2008. Management Support with Structured and Unstructured Data- An Integrated Business Intelligence Framework. Information Systems Management. Vol. 25, no. 2, pp. 132-148.

Baker, W., Kiewell, D. & Winkler, G. 2014. Using big data to make better pricing decisions. McKinsey & Company. [WWW-document]. [accessed 26.2.2020]. Available: https://www.mckinsey.com/business-. functions/marketing-and-sales/our-insights/using-big-data-to-make-better-pricing-decisions

Ballou, B., Heitger, D. L. & Stoel, D. 2018. Data-driven decision-making and its impact on accounting undergraduate curriculum. *Journal of Accounting Education*. Vol. 44, pp. 14-24.

Barfield, B. 2008. Price optimization. *Multi - Housing News*. Vol. 43, no. 4, pp. 32-33.

Bartosik-Purgat, M. & Ratajczak-Mrożek, M. 2018. Big Data Analysis as a Source of Companies' Competitive Advantage: A Review. *Entrepreneurial Business and Economics Review*. Vol. 6, no. 4, pp. 197- 215.

Baye, M., Gatti, J., Kattuman, P. & Morgan, J. 2007. A Dashboard for Online Pricing. *California Management Review.* Vol. 50, no. 1, pp. 202-216.

Brandão, A., Mamede, H. & Gonçalves, R. 2019. Trusted Data's Marketplace. *Advances in Intelligent Systems and Computing*. Vol. 930, pp. 515-527.

Brennan, R., Canning, L. & McDowell, R. 2007. Price-setting in business-to-business markets. *The Marketing Review*. Vol. 7, pp. 207-234.

Caine, S., Kermisch, R., Mewborn, S & Morningstar, T. 2019. The Pricing Is Right: Lessons from Top- Performing Consumer Companies. Bain & Company. [PDF-Document]. [accessed 5.2.2020]. Available: https://www.bain.com/globalassets/noindex/2019/bain_brief_the_pricing_is_right_lessons_from_top_performing_consumer_companies.pdf

Cambridge Dictionary. 2020. Data. [WWW-document]. [accessed 13.3.2020]. Available: https://dictionary.cambridge.org/dictionary/english/data

Chahal, V. 2018. Predict equipment failure using IoT sensor data. IBM. [WWW-document]. [accessed 26.2.2020]. Available: https://developer.ibm.com/patterns/predict-equipment-failure-using-iot-sensor-data/

Chatham, M. 2012. Structured Query Language By Example - Volume I: Data Query Language. Morrisville, Lulu. 206 p.

Chenavaz, R., Escobar, O. & Rousset, X. 2019. An analytical framework for retailer price and advertising decisions for products with temperature-sensitive demand. *Applied Economics*. Vol. 51, no. 52, pp. 5683-5693.

Choy, K., Lam, H. & Luk, C. 2018. Design of an Intelligent Customer Identification Model in e- Commerce Logistics Industry. *MATEC Web of Conferences.* Vol. 255, pp. 1-8.

Coles, M., Dransfield, R., Harris, R., Needham, D. & Rawlinson, M. 1999. Business for Higher Awards (Second Edition). Portsmouth, Heinemann. 696 p.

Coursera Blog. 2019. Not All Data Is Created Equal. [Blog Post]. [accessed 13.3.2020]. Available: https://blog.coursera.org/ds-academy-not-all-data-is-created-equal/

Davenport, C., Norkus, J. & Simonetto, M. 2012. Capturing the value of pricing analytics. *Advances in Business Marketing and Purchasing*. Vol. 19, pp. 299-333.

Dayton, E. 2020. Amazon Statistics You Should Know: Opportunities to Make the Most of America's Top Online Marketplace. [WWW-document]. [accessed 28.2.2020]. Available: https://www.bigcommerce.com/blog/amazon-statistics/#a-shopping-experience-beyond-compare

Demetrovics, J., Son, H. & Guban, A. 2016. A Formal Representation for Structured Data. *Acta Polytechnica Hungarica*. Vol. 13, no. 2, pp. 59-76.

Dholakia, U. 2020. The Risks of Changing Your Prices Too Often. Harvard Business Review. [WWW-document]. [accessed 26.2.2020]. Available: https://hbr.org/2015/07/the-risks-of-changing-your-prices-too-often

Deshpande, I. 2019. What Is Customer Data? Definition, Types, Collection, Validation and Analysis [WWW-document]. [accessed 26.2.2020]. Available: https://www.martechadvisor.com/articles/data-management/customer-data-definition-types-collection-validation-analysis-martech101/

Dolfsma, W. & Van Der Eijk, R. 2018. How info-firms use big data to target customers. *Journal of Business Strategy.* Vol. 39, no. 6, pp. 50-55.

Dopson, E. 2020. Customer Profiles: Benefits, Process and Examples. [WWW-document]. [accessed 26.2.2020]. Available: https://www.superoffice.com/blog/customer-profiles/

Dykes, B. 2016. Actionable Insights: The Missing Link Between Data and Business Value. Forbes. [WWW-document]. [accessed 28.2.2020]. Available: https://www.forbes.com/sites/brentdykes/2016/04/26/actionable-insights-the-missing-link-between-data-and-business-value/#314701e751e5

Farias, V., Moallemi, C., Van Roy, B. & Weissman, T. 2010 Universal Reinforcement Learning. *IEEE Transactions on Information Theory*. Vol. 56, no. 5, pp. 2441-2454.

Fiig, T., Le Guen, R. & Gauchet, M. 2018. Dynamic pricing of airline offers. *Journal of Revenue and Pricing Management.* Vol. 17, no. 6, pp. 381-393.

Fisher, M., Gallino, S. & Li, J. 2017. Competition-Based Dynamic Pricing in Online Retailing: A Methodology Validated with Field Experiments. Management Science. Vol. 64, no. 6, pp. 2496-2514.

Florez-Lopez, R. & Ramon-Jeronimo, J. M. 2009. Marketing Segmentation Through Machine Learning Models: *An Approach Based on Customer Relationship Management and Customer Profitability Accounting. Social Science Computer Review*. Vol 27, no. 1, pp. 96-117.

Florian, M. & He, S. 2010. Changing assignment algorithms: Price of better convergence. *Transportation Research Record.* Vol. 2176, no. 2176, pp. 67-75.

Frain, J. 1999. Introduction to Marketing. Boston, Cengage. 320 p.

Gabel, J., Desaphy, J. & Rognan, D. 2014. Beware of Machine Learning-Based Scoring Functions - On the Danger of Developing Black Boxes. *Journal of Chemical Information and Modeling.* Vol. 54, no. 10, p. 2807.

Géron, A. 2019. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. Seabastopol, O’Reilly. 818 p.

Giudice, P., Mustarella, L., Sofo, G. & Ursino, D. 2018. An approach to extracting complex knowledge patterns among concepts belonging to structured, semi-structured and unstructured sources in a data lake. *Information Sciences*. Vol. 478.

Google, 2020. Google Analytics. [WWW-document]. [accessed 26.2.2020]. Available: https://support.google.com/analytics/#topic=3544906

Gorodnichenko, Y. & Talavera, O. 2017. Price setting in online markets: Basic facts, international comparisons, and cross-border integration. *American Economic Review*. Vol. 107, no. 1, pp. 249-282.

Grimmer, M., Miles, M. P., Polonsky, M. J. & Vocino, A. 2015. The effectiveness of life-cycle pricing for consumer durables. *Journal of Business Research*. Vol. 68, no. 7, pp. 1602-1606.

Han, Z., Hoang, D., Kim, D., Luong, N., Niyato, D. & Wang, P. 2016. Smart Data Pricing Models for the Internet of Things. *IEEE Network*. Vol. 30, no. 2, pp. 18-25.

Haug, A., Zachariassen, F. & Liempd, D. 2011. The costs of poor data quality. *Journal of Industrial Engineering and Management*. Vol. 4, no. 2, pp. 168.

Hess, J. & Story, J. 2006. Segmenting customer-brand relations: beyond the personal relationship metaphor. *Journal of Consumer Marketing*, Vol. 23, no. 7, pp. 406-413.

Hinterhuber, A. 2008. Customer value-based pricing strategies: why companies resist. *Journal of Business Strategy*, Vol. 29, no. 4.

Hotjar, 2020. Hotjar vs Google Analytics: it’s not an either/or (and why you should use both). [WWW-document]. [accessed 26.2.2020]. Available: https://www.hotjar.com/blog/hotjar-vs-google-analytics/

Hopkins, M., Kruschwitz, N., LaValle, S., Lesser, E. & Shockley, R. 2011. Big Data, Analytics and the Path from Insights to Value. *MIT Solan Management Review.* Vol.52, no. 2.

Javanmard, A. & Nazerzadeh, H. 2019. Dynamic Pricing in High dimensions. *Journal of Machine Learning Research*. Vol. 20.

Jin, M., Wang, H., Zhang, Q. & Luo, C. 2018. Financial Management and Decision Based on Decision Tree Algorithm. *Wireless Personal Communications*. Vol. 102, no. 4, pp. 2869-2884.

Järvenpää, M., Länsiluoto, A., Partanen, V. & Pellinen, J. 2010. Talousohjaus ja kustannuslaskenta. Porvoo, WSOYpro Oy. 452 p.

Kabir, S. 2016. Basic Guidelines for Research: An Introductory Approach for All Disciplines. Chittagong, Book Zone.

Kelly, S. 2003. Mining data to discover customer segments. *Interactive Marketing*. Vol. 4, no. 3, pp. 235.

Kotler, P & Keller, K. 2012. Marketing Management. London, Pearson. 679 p.

Kumar, R. 2018. 4 Steps to Select the Right Database for Your Internet of Things System [WWW-document]. [accessed 5.2.2020]. Available: https://thenewstack.io/4-steps-to-select-the-right-database-for-your-internet-of-things-system/

Lin, K. 2006. Dynamic pricing with real-time demand learning. *European Journal of Operational Research*. Vol. 174, no. 1, pp. 522-538.

Llave, M. R. 2018. Data lakes in business intelligence: Reporting from the trenches. *Procedia Computer Science.* Vol. 138, pp. 516-524.

Louridas, P. & Ebert, C. 2016. Machine Learning. *IEEE Software*. Vol. 33, no. 5, pp. 110-115.

Labandeira, X., Labeaga, J. M. & López-Otero, X. 2017. A meta-analysis on the price elasticity of energy demand. *Energy Policy*. 102, no. C, pp. 549-568.

Mago, S. D. & Pate, J. G. 2009. An experimental examination of competitor-based price matching guarantees. *Journal of Economic Behavior and Organization*. Vol. 70, no. 1-2, pp. 342-360.

Mankiw, G. 2006. Principles of Microeconomics, 2 edition. Sandiego, Harcourt. 493 p.

Marcarelli, R. 2019. Why Retailers Are Turning on to Electronic Shelf Labels. [WWW-document]. [accessed 30.3.2020]. Available: https://www.winsightgrocerybusiness.com/technology/why-retailers-are-turning-electronic-shelf-labels

Marr, B. 2019. What's the Difference Between Structured, Semi-Structured and Unstructured Data? [WWW-document]. [accessed 5.2.2020]. Available: https://www.forbes.com/sites/bernardmarr/2019/10/18/whats-the-difference-between-structured-semi-structured-and-unstructured-data/#7fbc6a452b4d

Martinez-de-Albeniz, V. & Tallur, K. 2011. Dynamic Price Competition with Fixed Capacities. *Management Science*. Vol. 57, no. 6, pp. 1078-1093

Mon, J. 2018. How Machine Learning is reshaping Price Optimization. [WWW-document]. [accessed 5.2.2020]. Available: https://tryolabs.com/blog/price-optimization-machine-learning/

Morrison, A. & Stein, B. 2014. The enterprise data lake: Better integration and deeper analytics. PricewaterhouseCoopers. [PDF-Document]. [accessed 5.2.2020]. Available: https://www.pwc.com/us/en/technology-forecast/2014/cloud-computing/assets/pdf/pwc-technology-forecast-data-lakes.pdf

Mupparaju, K., Soni, K., Gujela, P. & Lanham, M. 2017. A Comparative Study of Machine Learning Frameworks for Demand Forecasting.

Neller, T. 2017. AI education: Machine learning resources. *AI Matters*. Vol. 3, no. 2, pp. 14-15.

Ning, C. & You, F. 2019. Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming. *Computers and Chemical Engineering.* Vol. 125, pp. 434-448.

Nielsel, M. 2019. Neural Networks and Deep Learning. [E-Book]. [accessed 26.2.2020]. Available: https://neuralnetworkanddeepLearning.com

Norvig, P. & Russell, S. 2009. Artificial Intelligence: A Modern Approach. Upper Saddle River, Prentice Hall. 1152 p.

Oliveira, F. & Rana, R. 2013. Real-time dynamic pricing in a non-stationary environment using model-free reinforcement learning. *Omega.* Vol. 47, pp. 116-126.

Oracle, 2020. What Is ERP? [WWW-document]. [accessed 26.2.2020]. Available: https://www.oracle.com/applications/erp/resources.html

Rissel, B. 2011. Price Elasticity of Demand. *Credit Union Management*. Vol. 34, no. 10, pp. 22-23.

Robb, R. 2017. Semi-Structured Data. [WWW-document]. [accessed 30.3.2020]. Available: https://www.datamation.com/big-data/semi-structured-data.html

Roine, A. 2019. How to start dynamic data-driven pricing? [WWW-document]. [accessed 5.2.2020]. Available: https://www.bds-bynfo.com/blog/how-to-start-dynamic-data-driven-pricing

Ryan, T. 2017. To Stay Relevant, Stores Need to ID Customers as They Come Through the Door. Forbes [WWW-document]. [accessed 26.2.2020]. Available: https://www.forbes.com/sites/retailwire/2017/03/08/to-stay-relevant-stores-need-to-id-customers/#6b5a1e05487c

Salanié, B. 1998. Microeconomics of Market Failures. Cambridge, MIT Pres. 238 p.

Saha, S & Basu, M. 2010. Integrated dynamic pricing for seasonal products with price and time dependent demand. *Asia-Pacific Journal of Operational Research (APJOR).* Vol. 27.

Sagar, A. 2019. Customer Segmentation Using K Means Clustering. [WWW-document]. [accessed 26.2.2020]. Available: https://towardsdatascience.com/customer-segmentation-using-k-means-clustering-d33964f238c3

Sanders, N. & Reid, R. 2002. Operations Management: An Integrated Approach. Hoboken, Wiley. 704 p.

Shai, B. & Shai, S. 2017. Understanding Machine Learning: From Theory to Algorithms. Cambridge, Cambridge University Press. 449 p.

Sharma, P. 2018. Consumer behavior and determinants of consumption pattern. *International Journal of Advanced Research and Development.* Vol 3.

Shim, J. & Taylor, R. 2019. Purchase-Based Analytics and Big Data for Actionable Insights. *IT Professional.* Vol. 21, no. 5, pp. 48-56.

Shim, J. & Siegel, J. 2007. Handbook of Financial Analysis, Forecasting and Modeling. Riverwoods, CCH. 478 p.

Smith, G & Nagle, T. 1994 Financial Analysis for Profit-Driven Pricing. *Sloan Management Review*. Vol. 35, no 3, pp. 74.

Tiwari, S., Cárdenas-Barrón, L. E., Goh, M. & Shaikh, A. A. 2018. Joint pricing and inventory model for deteriorating items with expiration dates and partial backlogging under two-level partial trade credits in supply chain. *International Journal of Production Economics*. Vol. 200, pp. 16-36.

Toit, J., Davimes, R., Mohamed, A., Patel, K. &Nye, J. 2016. Customer Segmentation Using Unsupervised Learning on Daily Energy Load Profiles. *Journal of Advances in Information Technology*. Vol. 7, no. 2.

Vickery, R. 2019. Beginners Guide to the Three Types of Machine Learning. [WWW-document]. [accessed 26.2.2020]. Available: https://towardsdatascience.com/beginners-guide-to-the-three-types-of-machine-learning-3141730ef45d

Waller, D. 2020. 10 Steps to Creating a Data-Driven Culture. Harward Business Review. [WWW-document]. [accessed 26.2.2020]. Available: https://hbr.org/2020/02/10-steps-to-creating-a-data-driven-culture

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J., Dubey, R. & Childe, S. J. 2017. Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*. Vol. 70, pp. 356-365.

Wang, C. & Huang, R. 2014. Pricing for seasonal deteriorating products with price- and ramp-type time-dependent demand. *Computers & Industrial Engineering*. Vol. 77, pp. 29-34.

Wang, J., Yang, Z. & Brocato, E. D. 2018. An investigation into the antecedents of prepurchase online search. *Information & Management*. Vol. 55, no. 3, pp. 285-293.

Yao, M. 2020. What are Important AI & Machine Learning Trends for 2020? [WWW-document]. [accessed 28.3.2020]. Available: https://www.forbes.com/sites/mariyayao/2020/01/22/what-are--important-ai--machine-learning-trends-for-2020/#31735fd23239

Yiu, T. 2019. Understanding Neural Networks. [WWW-document]. [accessed 28.3.2020]. Available: https://towardsdatascience.com/understanding-neural-networks-19020b758230

Ylinen, T. 2014. How to Factor in Weather Impacts in Demand Forecasting. [WWW-document]. [accessed 28.3.2020]. Available: https://www.relexsolutions.com/resources/how-to-factor-in-weather-impacts-in-demand-forecasting/

Yu, H. & Zhang, M. 2017. Data pricing strategy based on data quality. *Computers & Industrial Engineering*. Vol. 112, pp. 1-10.

Zhang, G., Zhang, C. & Zhang, H. 2018. *Improved K-means algorithm based on density Canopy. Knowledge-Based Systems.* Vol. 145, pp. 289-297.

Zhou, Z. 2018. A brief introduction to weakly supervised learning. *National Science Review.* Vol. 5, no. 1, pp. 44-53.