VILNIUS GEDIMINAS TECHNICAL UNIVERSITY

FACULTY OF BUSINESS MANAGEMENT

DEPARTMENT OF BUSINESS TECHNOLOGIES AND ENTREPRENEURSHIP

Giedre Stanelyte

INVENTORY OPTIMIZATION IN RETAIL NETWORK BY CREATING DEMAND PREDICTION MODEL

Master's degree Thesis

Business Management study programme, state code 6211LX058

International Business specialization

Business study field

Vilnius, 2021

VILNIUS GEDIMINAS TECHNICAL UNIVERSITY

FACULTY OF BUSINESS MANAGEMENT

DEPARTMENT OF BUSINESS TECHNOLOGIES AND ENTREPRENEURSHIP

APPROVED BY
Head of Department

(Signature)
prof. dr. Vida Davidavičienė
(Name, Surname)
2020-12-18
(Date)

Giedre Stanelyte

INVENTORY OPTIMIZATION IN RETAIL NETWORK BY CREATING DEMAND PREDICTION MODEL

Master′s degree Thesis

Business Management study programme, state code 6211LX058

International Business specialization

Business study field

Supervisor prof. dr. Aleksei Iurasov 2020-12-17
(Title, Name, Surname) (Signature) (Date)

Consultant
(Title, Name, Surname) (Signature) (Date)

Consultant
(Title, Name, Surname) (Signature) (Date)

Vilnius, 2021

VILNIUS GEDIMINAS TECHNICAL UNIVERSITY
FACULTY OF BUSINESS MANAGEMENT
DEPARTMENT OF BUSINESS TECHNOLOGIES AND ENTREPRENEURSHIP

APPROVED BY
Head of Department
____
(signature)
prof. dr. Vida Davidavičienė
(Name, Surname)
2019-10-31_____________
(date)

Business study field

Business Management study programme, state code 6211LX058

International Business specialisation

OBJECTIVES FOR MASTER THESIS

2019-10-31No ....

Vilnius

For student: Giedre Stanelyte

Master Thesis title: Inventory Optimization In Retail Network By Creating Demand Prediction Model

Approved on October 31, 2019 ' 240vv
(day, Month) (year)

The Final work has to be completed by December 17, 2020.
(Day, Month) (Year)

THE OBJECTIVES:

The objective of this study is to provide a structured source of information regarding the process of inventory optimization, to compare the application of different methods and by using the real time data to create a demand prediction model for inventory optimization.

: ....

....
(Title, Name, Surname)

Academic Supervisor .... prof. dr. Aleksei Iurasov............
(signature) (Title, Name, Surname)

Objectives accepted as a quidance for my Master Thesis
....
( )
….......Giedre Stanelyte………
(' ,)
2019-10-31
(Date)

Vilnius Gediminas Technical University
Faculty of Business Management
Department of Business Technologies and Entrepreneurship

ISBN ISSN
Copies No. ....
Date ..........-.....-.....

Master Degree Studies **Business Management** study programme Master Graduation Thesis

Title **Inventory Optimization in Retail Network by Creating a Demand Prediction Model**

Author **Giedrė Stanelytė**

Academic supervisor **Aleksei Iurasov**

Thesis language: English

Annotation

Today when various social and economic factors change established consumer habits, demand forecasting is becoming one of the key tools that can help maintain or even lead in the retail market. A properly designed demand forecasting model provides an opportunity to improve the inventory management and marketing processes those are important to the company and those have a direct impact on the company's financial results. There are many different ways and opinions on how to create the right model. The aim of this study is to analyze the different techniques using to create demand prediction models and based on the insights gained to develop a demand forecasting and inventory optimization model. The theoretical part of the work deals with research on inventory management systems, analytical programs and possible mathematical methods applied in different studies. The formed insights allow choosing fixed time ordering system, KNIME analytical program and Bayesian Additive Regression Trees (BART) mathematical method those will be used to develop a practical model for determining demand. In the research part, detailed sales data for January-June 2019 is selected, processed and adjusted at different stages so that the model developed using the Bayesian Additive Regression Trees mathematical method would be as accurate as possible. Based on available information and developed demand forecasting model, the need for replenishment is assessed on 11.07.2019. The results reveal that 399 products in ten different stores need replenishment because the available stock level cannot meet the projected demand. According to the author, replenishment model created could be improved by applying the economic order quantity model that with the optimal number of inventories defined may even better assist in company's inventory management processes. Thesis structure: introduction, three parts, conclusions, references and appendices. Thesis consists of 79 pages text without appendices, 10 tables, 15 pictures and 87 bibliographical entries.

Keywords: inventory optimization, Bayesian Additive Regression Trees method, retail network, demand prediction

Vilniaus Gedimino technikos universitetas
Verslo vadybos fakultetas
Verslo technologijų ir verslininkystės katedra

ISBN ISSN
Egz. sk. ....
Data ..........-.....-.....

Antrosios pakopos studijų **Verslo vadybos** programos magistro baigiamasis darbas
Pavadinimas **Prekybos tinklų atsargų optimizavimas pasitelkiant paklausos prognozavimo modelį**
Autorius **Giedrė Stanelytė**
Vadovas **Aleksei Iurasov**

Kalba: anglų

Anotacija

Šiandien, kai įvairūs socialiniai ir ekonominiai veiksniai keičia nusistovėjusius vartotojų įpročius, paklausos prognozavimas tampa viena pagrindinių priemonių, galinčių padėti išsilaikyti ar net pirmauti mažmeninės prekybos rinkoje. Tinkamai parengtas paklausos prognozavimo modelis suteikia galimybę pagerinti svarbius įmonei atsargų valdymo ir rinkodaros procesus, turinčius tiesioginės įtakos įmonės finansiniams rezultatams. Yra daug skirtingų būdų ir nuomonių, kaip sukurti tinkamą modelį. Šio tyrimo tikslas – išanalizuoti skirtingas technikas, naudojamas paklausos modeliams kurti, ir, remiantis gautomis įžvalgomis, sukurti realiais pardavimų duomenimis grįstą paklausos prognozavimo ir atsargų optimizavimo modelį. Teorinėje darbo dalyje nagrinėjami moksliniai darbai apie praktikoje taikomas atsargų valdymo sistemas, analitines programas ir galimus pritaikyti matematinius metodus. Suformuotos įžvalgos leidžia pasirinkti konkrečius įrankius – fiksuoto laiko užsakymų sistemą, KNIME analitinę programą ir BART matematinį metodą, kuriuos naudojant bus kuriamas praktinis paklausos nustatymo modelis. Tiriamojoje dalyje detalūs 2019 m. sausio–birželio mėnesių pardavimų duomenys atrenkami, apdorojami ir skirtinguose etapuose koreguojami taip, kad taikant BART matematinį metodą sukurtas modelis kaip įmanoma tiksliau nustatytų būsimą paklausą. Remiantis turima informacija ir sukurtu paklausos prognozavimo modeliu įvertinamas 2019-07-11 atsargų papildymo poreikis. Rezultatai atskleidžia, jog 399 produktams dešimtyje skirtingų parduotuvių reikalingas papildymas, nes turimas atsargų lygis negali patenkinti prognozuojamos paklausos. Pasak autoriaus, modelis galėtų būti plėtojamas pritaikant ekonominio užsakymo kiekio modelį, taip suteikiant efektyvią pagalbą įmonės procesams gerinti. Darbo struktūra: įvadas, trys dalys, išvados, literatūros šaltiniai ir priedai. Darbą sudaro: 79 puslapių teksto be priedų, 15 paveikslų, 10 lentelių ir 87 bibliografinis šaltinis.

Prasminiai žodžiai: atsargų optimizavimas, BART matematinis metodas, mažmeninė rinka, paklausos prognozavimas

Vilniaus Gedimino technikos universiteto egzaminų sesijų ir baigiamųjų darbų rengimo bei gynimo organizavimo tvarkos aprašo
2 priedas

(Baigiamojo darbo sąžiningumo deklaracijos forma)

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS

Giedrė Stanelytė, 20190654

(Studento vardas ir pavardė, studento pažymėjimo Nr.)

Verslo vadybos fakultetas

(Fakultetas)

Verslo vadyba, TVfmu-19

(Studijų programa, akademinė grupė)

BAIGIAMOJO DARBO (PROJEKTO)
SĄŽININGUMO DEKLARACIJA

2020 m. gruodžio 17 d.

Patvirtinu, kad mano baigiamasis darbas tema „Prekybos tinklų atsargų optimizavimas pasitelkiant paklausos prognozavimo modelį“ patvirtintas 2019 m. spalio 31 d. dekano potvarkiu Nr. 240vv, yra savarankiškai parašytas. Šiame darbe pateikta medžiaga nėra plagijuota. Tiesiogiai ar netiesiogiai panaudotos kitų šaltinių citatos pažymėtos literatūros nuorodose.

Mano darbo vadovas profesorius daktaras Aleksei Iurasov.

Kitų asmenų indėlio į parengtą baigiamąjį darbą nėra. Jokių įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs (-usi).

(Parašas) Giedrė Stanelytė

(Vardas ir pavardė)

CONTENT

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

The world is going through the changes. Digitalization, globalization and various security risks are just a several of the macro developments impacting industries and supply chains all around the world. Moreover, it is important not to forget customers behaviour as one more significant aspect, which is changing faster than anyone thought possible just a few years ago as modern technologies make it possible to source cheaper and better goods from all around the world. Inventory is the most significant investment in any industry and specifically in retail sector. The effect of inventory management and its optimization in retail is phenomenally high. In order to maximize the Return on Investment, each and every retailer must ensure that they are familiar with the changes and their inventory investments are kept to the optimal level. Optimal, or in the other words – best, is not an easy goal to reach. Retailers face huge challenges with the management and storage of their products. And even for the retailers, those have a big warehouses and the storage problem for them is not so high, they could feel the effect of the storage on the totally different perspective, as they have to manage the products in a way the most demanded ones would not be replaced by non-purchased ones. So the entire phenomenon of inventory in retail sector is dependent upon the customers and their demand. The potential of firms to succeed at the stage they hit today relies on their ability to forecast future demand. Studies on demand forecasting are an important area of research for many fields, but the retail sector is more complicated environment than other sectors. The most significant explanation for this is the variables affecting the demand those are not known exactly. There could be many variable such as habits, beliefs, seasonality or trends possible. The retail industry generally tends to sell fast. Only in the limited situations, the customer will wait for the replenishment of the product they want in the specific store. In most of the cases they will enter the competitor's stores and will fulfill the desires there. That in a short time perspective means the loss of sales for the store. But also in a long time perspective affects the loyalty of the customers and reputation of the store in whole. Therefore it is essential to accurately forecast the demand of the customer in order to create the precise inventory replenishment model.

In the practice, there are many different methods applying to forecast the demand. Besides classical methods such as time series, artificial intelligence based methods that expand the application areas with the developing technology are also often used in studies of prediction (Güven & Şimşir, 2020). The variety of studies, different methods and even the combination of those methods applying emphasize the existing relevance of the topic on the demand prediction in the process of inventory optimization. In order to supplement and refine the studies conducted so far by other researchers, the work of thesis is focus on the study of significant parts of inventory optimization by extending it with the practical model of demand prediction created based on the real time data of retail company. In the first part of work the theoretical foundation of inventory optimization is overviewed for the need and values to emphasize. The different inventory optimization techniques in retail – Fixed Time Period, Fixed Order and Vendor Managed Inventory systems are compared in order to see the options possible and to clarify the best approach for the case analysis in practical part. As we are living on time when consumption increases, IT applications are changing the traditional ways of analysis and the big data is growing its role in the field of customers data processing, the most often used IT applications for inventory analysis – Microsoft Power BI, IBM SPSS Modeler, KNIME and Apache Hadoop are also compared in the first part of the study. Each are different, therefore the main functionality, the benefits and areas to improve from the users perspective are delivered. However, only IT application by itself is not making the right decisions. It is no less important to choose the mathematical method that will be applied to determinate the future demand. Six different mathematical methods – Linear regression, Logistic regression, Probabilistic Neural Network (PNN), Bayesian Additive Regression Trees (BART), Random Forest and Fuzzy Logic for demand prediction modelling in various situation in retail are disclosed and compared according to established sections. Review of the first chapter provides the reader with the specifics of inventory optimization and application of possible IT developments together with mathematical models for demand prediction. In the second part of the work BART algorithm for estimation of non parametric functions using regression trees in demand prediction is analysed in more specific. The analysis of the main assumptions and understanding of mathematical path of algorithm enables better application of method on a real time data in practical part. As in the research part, the real sales data from grocery retailer is conducted, before the practical study the grocery retail market situation in Lithuania is overviewed in order to clarify the main assumptions for the features in the practical part to add. In the practical part of the work, with the help of KNIME analytical platform and application of BART algorithm demand prediction model for inventory optimization with the flag for replenishment is created. Half year period sales data of one retail company is used to train the model to predict the need for product replenishment for upcoming periods. Flexible to use and to modify model may be integrated as a powerful tool for the management in the inventory decision making process. As model created based on a real time data of company, correct application of the model in the processes may lead to the processes optimization, better coordination with cost reduction and possible increase of the profit.

Topicality. The topic studying is relevant from both – theoretical and practical point of view. Today when the world is facing increasing amounts of consumption, the role of analysis in retail sector inventories management becoming increasingly important. Theoretical review in each of the situation emphasize that authors are still seeing the area of demand prediction as an open for further researches with possible improvements. Therefore comparison of different studies and methods applied, enriched with new personal empirical research results are valuable as provide new insights and raise the new questions for the further studies to perform.

Problem. There is no tool in retail network for stock replenishment that may work based on up to date mathematical methods in demand prediction and existing algorithms of economic order quantity.

Research object. Enterprise inventories optimization on behalf of demand prediction modelling.

Aim of the thesis. The aim of this study is to provide a structured source of information regarding the process of inventory optimization, to compare the application of different methods and by using the real time data to create a demand prediction model for inventory optimization.

Tasks:

1. To analyse different inventory systems, IT applications and mathematical methods for inventory analysis, optimization and demand prediction in retail network. Study its application in practice.
2. To analyse the tendencies in retail market of Lithuania for possible assumptions applicable in practical part to distinguish.
3. To systemize the real time data of sale in the first stage of data collection and pre-processing in order to use data for demand prediction.
4. To optimize the model by applying parameters optimization and feature elimination procedures for the minimal error and highest accuracy to receive.
5. To apply the demand prediction model created to deploy the data for products replenishment to predict.

Methods of the Research. Scientific methods applied consist of comparative analysis and systematization of scientific literature performed in theoretical part and demand prediction modelling in practical part. Modelling prepared by using Bayesian Additive Regression Trees method. For modelling assessment of results KNIME workflow was used.

1. THE OVERVIEW OF INVENTORY OPTIMIZATION TECHNIQUES IN RETAIL

1.1. Theoretical foundation of inventory optimization

Everyday improvements, competition and new challenges force the companies strongly focus on inventory optimization. Companies have to manage the stock in the way, the business and the customers would gained the highest value possible. Well organized inventory optimization process ensure that the company would have the right inventory to meet the target service levels while at the same time would ensure an optimal sum of capital to invest in inventory to buy and hold. Moreover well managed process could even lead to growth of profit. Therefore the studies of inventory optimization here take an important meaning. Research performed by Huang (2019) of inventory optimization in fresh food based supermarket, have showed the results that after the inventory optimization, the profit margin of various kinds of meat selected has been improved to different degrees with grow till 40% achieved. An increase in profit here shows the right decisions taken, involving multiple parties – customers, suppliers, producers, and different levels of company's departments. The key elements of optimizing the inventory are as follow (Haling J., 2019):

1) Demand forecasting
2) Inventory policy
3) Replenishment

The authors San-José, Sicilia, González-De-la-Rosa, & Febles-Acosta (2019) in their work also emphasizing three main elements of inventory management by saying that those are: carrying inventories, incurring shortages and replenishment. In inventory system, the shortages could arise due to the lack of stock to satisfy the need of the customers. Demand according to authors are inconstant and depends on demand rate varies on time. All the inventory systems are constrated in the way of variables included those are the subjects to control in the system. Inventory costs are influenced by controllable variables and the problem consists of finding the particular values of certain variables that may minimize the overall cost of inventory. In our practical work we integrated elements highlighted above as important parts in the way we first analysed the different features and forecast the demand. After that under the inventory policy existing we made decisions of replenishment with quantities evaluated as needed. However, it is important to mention – deciding how much inventory to carry is not a simple task, first of all requiring considering not only the stock needed for daily operations, but also the way inventory will be added to warehouse. The process usually is determinate by system chosen and is defined in company inventory management policy.

The issue also important to mention especially when analysing retail food market inventory systems is perishable inventory routing problem. From farmers to final customer, fresh food items undergo many different procedures starting from planting, harvesting and after moving to proceeding. At different points of the supply chain fresh items are carried in the warehouse and after that delivered to retailer. The period of time counting from the beginning of product cycle and delivery in stores may differ based on the inventory management system applied. The increase of variety of fresh food in the shelves of the store also brought more complexity to this problem from the inventory management perspective as if in the past only one or two different suggestions of one item exist, today there are number of different forms of fresh food for demand to satisfy. However, the pressure from the government regulatory and customers regarding the issues of food safety, freshness, conformance and traceability has increase together with variety that increase. Therefore for perishable inventory there are even more important to find the best suitable inventory system to apply. The authors (Onggo, Panadero, Corlu, & Juan, 2019) reviewed the literature under this topic and find out that there is a lack of studies integrating inventory and routing decisions for fresh products supply chains. The authors in their study analysed supply chain of fresh agro-food production where there is a single food supplier who owns a central warehouse that serves several retail centres with stochastic demands. The decisions are taken by supplier on the amount of inventory and fresh products transportation to each retail centre. The case was moderate to be analysed as a multi-period inventory routing problem with stochastic demands and perishable products in the contexts of supply chain of agro-food. The authors have modelled the problem as a mixed integer problem and developed a simheuristic algorithm, which is based on a local search metaheuristic, in order to solve it. Even those the limitation highlighted by authors here were a narrow extension of products and suppliers, the work of authors propose simheuristic algorithm which can produce an optimal solution for inventory management in particular case that can minimize the spending of inventory, routing operations and costs related to food wasting, which for the business means the loss of the profit. Although the case was on agro-food supply chains and was analysed from the perspective of product supplier, the retail market as the final receiver of the profit from the perishable products should also include the perishable inventory routing problem into the consideration of selecting inventory system used by their own. It is important as most of the companies are selecting and using one inventory system for the whole inventories to managed, therefore it is essential that the system chosen would suit to the whole variety of inventories holding.

1.1.1. Fixed Time Period inventory system

Systems to manage the stock can be divided into single period and multi period inventory models. Single period systems are only for one time ordering decision and are mostly used when there are one time events or sales. If the stock required to be filled constantly, the multi period inventory system here is considered as needed (Carr P., 2011). A multi period inventory model can have two variations – Fixed Time Period and Fixed Order Quantity systems. Also it is important to mention that in classical inventory models the demand rate and holding cost are assumed to be constant. However looking into realistic situation of inventory in market conditions the demand and holding cost both are dependant and vary according to the time (Mishra, 2013). The later studies inventory management system by authors Prasad & Mukherjee, (2016) reveal that demand rate of consumer is dependent on current stock level and time. Moreover, mathematical model created also observed that growth in the stock level of the production usually has a positive impact on its demand and in such case retailer display each of item in large quantity to encourage the demand of the product (Prasad & Mukherjee, 2016). So, time is playing an important role in the inventory system, especially in fixed time period inventory system.

Fixed Time Period inventory system could be described as stock system where stock is counted and reordered at pre-determined intervals – days, weeks or months. The order quantity here would depend on how much inventory is needed. Regularly scheduled shipments are common practice for the efficient use of transportation resources and the procurement of transport services at least cost. Here, in the situations with a multi – item inventory distribution network, where the replenishment of each item takes a small place of a truckload, transportation economies dictate a replenishment schedule with consolidation of item shipments. Most large retail chains, such as Walmarts, replenish most of their items in such way. (Graves, 1996).

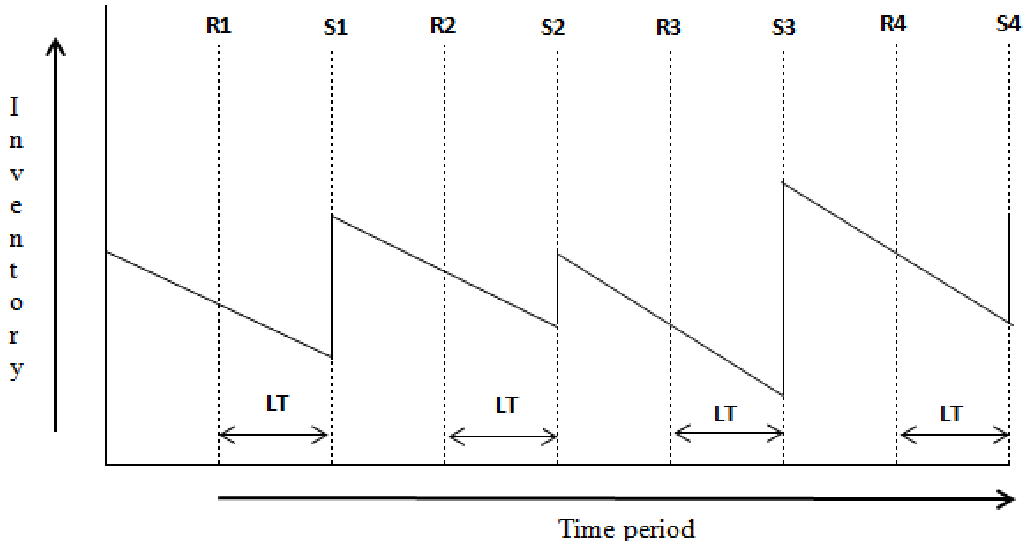

Fig. 1.Fixed time period system

Source: prepared by author

Figure 1 illustrating the process of Fixed Time Period inventory system, where R1, R2, R3, R4 – are the review periods, S1, S2, S3, S4 – are the supply points and LT – is Lead time. Lead time is the delay applicable for inventory control purposes. Delay structure consists of two components – delay in delivery, that is, the time it takes for a supplier to deliver the products after an order has been placed, and the delay in reordering the products, which is the time until an opportunity to place an order comes up again. Lead-time is normally fuzzy or stochastic in nature. Replenishment lead time according to the authors Pan & Hsiao, (2002) is considered to be one of the important factors for decision making in inventory management and has always been a popular research subject. Even today, it has been included in researches as subject together with topics of inventory management quality improvements, cost minimization and variation controlling (Malik & Sarkar, 2018). This is mostly due to today's demand uncertainties and possible shortage in stock those control and coordination become a challenge for the business. The study in 2013 performed a research to understand the customers' actions on missing products that they were willing to buy. Results showed that around 15% of all customers delayed the purchase if they faced with a stock-out for the item they were looking for in a particular store. The other 85% of the customers decided to buy product substitution, to buy the item in another store or not to buy the item at all (Van Donselaar & Broekmeulen, 2013). Only the situation of substitution here would give a profit for the retailer, while in all the remaining situations demand for the preferred item – store combination would lead to loss. The reason of such losses usually is caused by uncertainty. Here are three main sources of uncertainty in supply chain: manufacturing, suppliers and customers (Rahdar, Wang, & Hu, 2018). Uncertainty of suppliers contributes to inconsistency of lead time while uncertainty of customers affects in order time or quantity, both of which could cause unplanned costs. To avoid the situation of possible losses and uncertainties, the lead time or the lead time demand is calculated for different situations. Different authors have analysed this topic and gave their insights and models. Rahdar et al., (2018) inventory control models classified into four types:

a) deterministic;
b) stochastic demand and fixed lead time;
c) fixed demand and stochastic lead time;
d) stochastic demand and lead time.

Based on authors majority of inventory management systems studies concentrated on deterministic models or discussed uncertainties on either the supply or demand side. However, deterministic model methodology focused only on the premise that all parameters and variables associated with inventory stock are defined and the demand on replenishment is known. In real time environment such situation where all the parameters are known are unlikely, therefore it is better to focus and analyse the studies of other types where the lead time is the most important. In their study the authors (Rahdar et al., 2018) suggested a new understanding of uncertainty in a manufacturing facility which is ordering production to meet demand. Study has important parts of: takes into account both sources of uncertainty (demand and lead time), proposes two-stage tri-level optimization model for problem of inventory control and designs an exact algorithm for the tri-level optimization model to deploy. Results of the study resume that inventory optimization model created works more adaptively in response to the range of cost parameters. The authors Sajadieh, Jokar, & Modarres, (2009) suggested a model to reduce the overall spending of an integrated supply chain for the vendor and buyer when the lead time is stochastic. Like before discussed authors Barnes-Schuster, Bassok, & Anupindi, (2006) in their study also analysed the big companies and their possibility to composed the system of inventory by both suppliers and their buyers (in our case – retailer as inventory buyer). The main argument that encouraged authors to perform the study was a real event by Compaq Company placed in 1994. The company was able to reduce its supply cost by 600 million dollars, by reorganizing its supply chain. Reorganization included their inventory movement from their warehouses to nearby premises leased by their main 35 suppliers. By giving their suppliers moving compensations Compaq as a company was able to received inventory when and where they were needed with minimized transportation cost and reduced delivery lead time. Based on the example mentioned authors prepare the study to calculate whether is it optimal for the system to reduce the delivery lead time to zero and what are the impact of delivery lead time on different cases. For the purpose of comparison authors prepare and considered two different cases – two-stage supply system consisting of one supplier and one buyer and compare it with the more complex with one supplier and multiple buyers. Result of the study determinate one more important component of stock – safety stock (Fig. 2) and its relation with delivery lead time, which differs based on case. One more inventory model was developed by three authors Maiti, Maiti, & Maiti, (2009) where the lead time was a random variable that followed normal or exponential distribution. The lead time was considered by author (Hoque, 2013) to be an independent random variable from a normal distribution. Further study on polynomials function for lead time demand to calculate was prepared by Cobb, Johnson, Rumí, & Salmerón, (2015). The lead time demand determinate the total demand between particular moment and the anticipated time for the delivery after the next one if a reorder is made at particular time to replenish the inventory. For the parameters of model to calculate authors used Bayesian criteria. Formulized function provides an opportunity to find an optimal solution for inventory policy in systems. Other authors (Ignaciuk & Bartoszewicz, 2009) develop a new supply model for periodic inspections on inventory level using strict control-theoretic methodology. The aim of the inspection action was the ability to satisfy the entire demand from the readily available stock capacity in combination with currently arriving shipments, thus ensuring no backorders or changes in the need of the customers. For this purpose authors applied regulation technique of discrete-time sliding-mode with linear-quadratic optimization and performance index for determination of parameters and coefficients. As quantity computation in model involves only simple operations, with review of limitations, the model could be efficiently implemented into real not complex inventory management systems Further the short summary of the studies performed on the topic of Lead Time in inventory management is provided for comparison of different ideas of authors.

Table 1. Summary of studies on Lead Time

Authors	Main ideas of studies
(Barnes-Schuster et al., 2006)	Analysed the possibility to composed the system of inventory by both suppliers and their buyers (retailer as inventory buyer);
(Ignaciuk & Bartoszewicz, 2009)	Developed a new supply model for periodic inspections on inventory level using strict control-theoretic methodology;
(Sajadieh et al., 2009)	Suggested a model to reduce the overall spending of an integrated supply chain for the vendor and buyer when the lead time is stochastic;
(Maiti et al., 2009)	Analysed lead time as random variable that followed normal or exponential distribution;
(Hoque, 2013)	Considered lead time as an independent random variable from a normal distribution.
(Cobb et al., 2015)	Study on polynomials function for lead time demand to calculate. Formed function provide an opportunity to find an optimal solution for inventory policy in systems;

Table 1. Summary of studies on Lead Time (Continuation)

Authors	Main ideas of studies
(Rahdar et al., 2018)	Suggested a new understanding of uncertainty in a manufacturing facility which is ordering production to meet demand. Inventory optimization model created works more adaptively in response to the range of cost parameters.

Source: prepared by author

The overview of studies disclosed different views and problems in the lead time perception. The main concepts where studies analysed were focused on – uncertainties, satisfaction of demand, reduction of costs, analysis in lead time as random variable and new models composition for improvement of entire concept mentioned. None of the studies created a perfect all the sectors and situations suitable model, therefore the studies on this issue is ongoing.

1.1.2. Fixed Order inventory system

Another important multi period inventory system is Fixed Order Quantity systems. A Fixed Order Quantity system is the settlement in which the level of inventory is continuously monitored and replenishment stock is ordered in previously-fixed quantities whenever at-hand stock falls to the set reorder point. Many enterprises follow the Fixed Order Quantity system as it helps to reduce failures in reordering, to efficiently manage the storage capacity and prevent the unnecessary funding blockage. Moreover, Fixed Order Quantity system is popular among manufacturers as method ensures the regular replenishment of inventory items, those are needed in process of production.

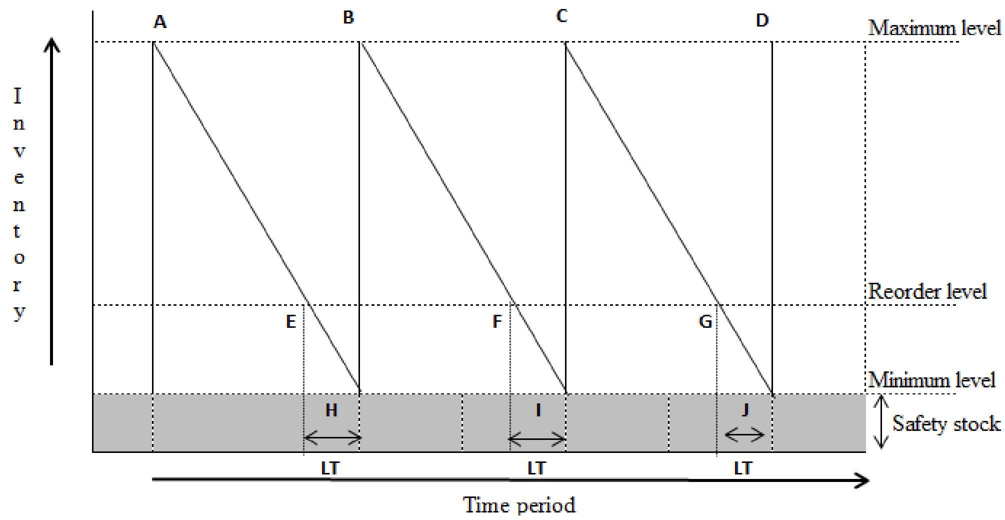

Fig. 2. Fixed Order Quantity system

Source: prepared by author

The Figure 2 illustrating the process of Fixed Order Quantity inventory system, where A, B, C, D – are the supply points, E, F, G – reorder points and H, I, J (LT) – are Lead time periods. When analyzing Fixed Order Quantity inventory system, it is also important to mention Economic Order Quantity (EOQ). Economic order quantity is a model that defines the optimal quantity to order that minimizes total variable costs of ordering and holding inventory (Kostić, 2009). Formula of EOQ, determine the inventory reorder point when the certain level of inventories in stock is reached. By determining reorder point, company is able to prepare therefore can avoid to face running out of stock. Due to its ease to use, robustness and simplicity to define EOQ is one of the most common and efficient optimization models in supply chain management. There are various studies suggesting different applications on EOQ. Some authors suggest their own approaches, while others applied already existing mathematical ones. Author Minner (2007) suggest that the most popular approach to create the EOQ formulation is to provide an analytical expression for the average cost per unit and then to adapt differential calculus, that is formulizing sufficient conditions by taking of the cost function the first and second orders derivatives. In his own study the author suggests a different approach in order to receive EOQ without different calculus and algebraic manipulations. When obtaining the local cost minimum, author compared the cost coefficients, took the time horizon to infinity to get the EOQ and then established that it is a global minimum. His colleague Teng (2009) after 2 years further created, as he named, even simpler method by using the arithmetic-geometric-mean-inequality theorem to compute the global optimal order quantity without including a hard calculus, by applying tedious algebraic manipulations and by comparing cost coefficients. However as time goes, together with it changes and new assumptions in models have to be considered all the time. Even though, a constant unit purchase assumption considered being of the main assumption in the classic economic order quantity model, in today practice, providers sometimes are offering special prices either to reduce inventories of certain items or to increase the cash flows. The buyer is such situation has two different choices: to create a special order or to remain at a constant quantity of good ordered. If the special order to be chosen at particular time, then it is important clearly determinate the size of the order to avoid possible losses of unsold goods. This is the most relevant to food retail sector that is our key sector to analyse as most of the products there are fresh and have a limited lifetime. Arcelus, Shah, & Srinivasan (2001) already investigated the benefits and drawbacks the retailers and suppliers are faced when they are getting decrease in a purchase price of inventory. Later on, some of the authors Arcelus, Pakkala, & Srinivasan (2003) continues their study by analyzing the effect on retailer if the promotion would have a set minimum duration time, however uncertain termination date. The study identified the need of special sales model under uncertainty, which lead to profit maximization for retailer in the periods of both deterministic and stochastic periods during the special sales terms. As study performed identified many uncertainties in the model formed, Arcelus, Shah, & Srinivasan (2003) prepared one more study on generalization of retailer special sales problem. Authors were able to formulize retailer profit maximization strategy in conditions of credit or price discount gained from provider on a purchase of regular and perishable production with the flexibility for retailer to modify its own selling policies based on managerial implications and to better respond to competitive pressure. Later on, the studies of authors have been supplemented by ideas of more authors. Sarker & Al Kindi (2006) analysed the optimal ordering policies in response to a discount offer for five different cases:

a) coincidence of period of sale with time of replenishment;
b) non-coincidence of period of sale with time of replenishment;
c) sale period longer than a cycle;
d) price with promotion as a function of the special ordering quantity;
e) incremental promotion.

Comparison of different sales scenarios where created to sense the effect of different parameters of possible ordering policies. In each different case authors were looking for the optimal ordering quality by maximizing the difference among regular EOQ spending and spending of special promotion during the set sales with promotion period. A remarkable increase in the efficiency of the inventory system when a special order is issued during the selling time was demonstrated by numerical analysis of authors. Taleizadeh, Pentico, Aryanezhad, & Ghoreyshi (2012) in their model of EOQ in a special price conditions included also a partial backordering. Backordering is the strategy closely related to emergency replenishment and satisfaction of demand. Nowadays more and more businesses are implementing the strategies of emergency replenishment in case to hold unsatisfied demand by providing an opportunity of backordering (Li & Ou, 2020). Even though, application of backordering may be applicable only in online retail marketing, as consumer in the real shop will not wait till the product they are willing to buy will be replenish and delivered to shelves of the store, however, it is needed to include these features into consideration of inventory systems and EOQ as after adding the features inventory stocking decision making for the different products become much more challenging task. Mentioned above Taleizadeh et al. (2012) have modelled EOQ for three different scenarios with the special sales promotion and partial backordering opportunity: the sales price is accessible at the normal time to place an order, it is accessible only if there is still existing inventory or it is accessible only when there is a stock out. The result of the study shows that for the three different scenarios optimal solution variable values are the same. The authors considered the results of their study where decision variables are the cycle length and the percentage of demand filled from stock as the advantage of model comparing with other authors, where the decision variables are the level of stock out and order quantity, which differ for each of three scenarios. Adding to the further ideas of inventory models and backordering, an interesting results on EOQ was also provided by Zhang, Kaku, & Xiao (2011) where the correlation in developed two products inventory model was recognized. Authors identified that the demand for the minor item is correlating to demand of the major item as of cross-selling and backordering solution that was applied equally for both items. On the other hand, there are studies on EOQ which also analysed inventory orders modelling when the growth in price is known. In such situation, when the provider declares either a temporary or permanent increase in the unit acquisition cost of the product, the buyer has to decide – whether to minimize the overall acquisition cost by placing a larger than usual special order or to remain the quantity of order the same as usual. Taleizadeh & Pentico (2013) analysed such case in their study where EOQ model was created for two scenarios with a declared increase in price and partial backordering during the period of stock out. Authors improved inventory model prepared by Tersine (1996) as in their two scenarios model decision variable of order quantity and stock out level are in both same, what leads to better results of the model. As in addition to topic of growth, an interesting study was performed by Rezaei (2014) regarding the optimizing inventory model for growing items. The author considered the case where an enterprise is buying and growing the newborn animals and after that selling final product to the markets. The sensitivity analysis was conducted to calculate the effect of the main model parameters on two decision variables of the model (EOQ at period start and optimal period ending day) and on the objective of the model that is profit. The study of author emphasized the variety of studies on EOQ modelling existing and results those could be used by decision-makers in planning, maintaining and controlling operations. Further the short summary of the studies on the topic of EOQ in inventory management is provided for comparison of different ideas of authors.

Table 2. Summary of studies on Economic Order Quantity

Authors	Main ideas of studies
(Arcelus et al., 2001)	Investigated the benefits and drawbacks the retailers and suppliers are faced when they are getting decrease in a purchase price of inventory;
(Arcelus, Pakkala, et al., 2003)	Analysed the effect on retailer if the promotion would have a set minimum duration time, however uncertain termination date;
(Arcelus, Shah, et al., 2003)	Formed retailer profit maximization strategy in conditions of credit or price discount gained from provider on a purchase of regular and perishable production with the flexibility for retailer to modify its own selling policy based on managerial implications and to better respond to competitive pressure;
(Sarker & Al Kindi, 2006)	Analysed the optimal ordering policies in response to a discount offer for five different cases;
(Minner, 2007)	Suggested a different approach in order to receive EOQ without different calculus and algebraic manipulations. When obtaining the local cost minimum, author compared the cost coefficients, took the time horizon to infinity to get the EOQ and then established that it is a global minimum;
(Teng, 2009)	Suggested the method by using the arithmetic-geometric-mean-inequality theorem to compute the global optimal order quantity without including a hard calculus, by applying tedious algebraic manipulations and by comparing cost coefficients;
(Taleizadeh et al., 2012)	Modelled EOQ for three different scenarios with the special sales promotion and partial backordering opportunity;
(Taleizadeh & Pentico, 2013)	Created EOQ model for two scenarios with a declared increase in price and partial backordering during the period of stock out;
(Shi, Zhou, Wang, Xu, & Xiong, 2014)	Created the EOQ model where the correlation in developed two products inventory model was recognized;
(Rezaei, 2014)	Formulated the optimizing inventory model for growing items.

Source: prepared by author

The overview of studies disclosed different problems in the EOQ perception especially when there is fixed order quantity system of inventories applied. The main concepts studies reviewed were focused on – price changes and effect to ordering quantities, inventory policies, inventory optimization and profit maximization. Even the studies analysed different problems on the same topic, none of the studies created a perfect model suitable for the all different situation therefore the studies on this issue are ongoing.

The other important method to prevent from running out of stock is safety stock (Fig. 2). In supply chain inventory management it is generally accepted that strategy of safety stock could help to deal with stochastic supply and demand uncertainties, to save the situation of lower production output, to protect against delayed shipments, or guard against unpredictable issues such as natural disasters. (Hallin J., 2019). When analyzing retail, safety stock is principally kept to reduce a risk of stock outs, to carry on with seasonal supply and demand fluctuations, or possible lost sales and customers due to before mentioned reasons. The companies those have a safety stock is one step ahead the competitors. If five or ten years ago the customers could wait for the product they are searching, today the market is filled with different suggestions, thus only the best prepared business could remain or win the competition. However locating safety stock lead to additional cost therefore safety stock does not have to be placed in every stage and by every group of the products. It is important that safety stock would be optimally deployed to protect the supply chain from any kind of uncertainties and additional costs. The issue of safety stock position consisting of determination of amount of safety stock that has to be stored at each different stages in such a way that the supply chain's overall service level would be accomplished. In the face of numerous business and technological uncertainties, the output of supply chain is typically calculated by service level, that is estimated fraction of demand that can be fulfilled by the supply chain within a predefined permissible delivery time window (Jung, Blau, Pekny, Reklaitis, & Eversdyk, 2008). Lesnaia (2004) identified the problem of the amount of stock to hold at each stage of the supply chain for the company that is facing uncertainty in demand. According to the author the amount of stock held should be at such level that the storage cost would be the minimum, while at the same time, the amount of stock has to fulfilled demand and the need of the customers on the highest level possible. Study analysed the problem by creating the framework of deterministic service time models and in general-network supply chain forming an algorithm for safety stock placement. The study had some limitations with the most important of assumption on infinite production capacity left by author for further researches. The assumption after two years was included into the studies by other authors Sitompul, Aghezzaf, Chen, & Dullaert (2006). Their study analysed the safety stock placement in capacitated and uncapacitated production stages. Results of the study emphasized that well chosen production capacities in the chain will decrease the required amount of safety stock. Moreover it shows that planning of capacities has a positive influence on ordering quality and lead time. Therefore modelling of supply chains according to the authors has to start from correct planning of capacity. However, when here is a planning process, risk of estimation takes an important part. Nechval & Nechval (1999) have investigated the effect on estimation risk on decision making involving a simple one-period inventory problem where as an example Christmas tree warehousing problem was taking. Results of the study show that where the estimation risk was ignored, the level of stock was captured incorrectly therefore the level of service was inadequate and uncertain to meet the demand. Even though the model established by author has specific application on the stocking policy of highly seasonal, single period demand item that could be considered as limitation, however it highlights the need to focus on risky estimation areas and in the case where it is impossible to control – include an amount of safety stock to avoid possible estimation errors. Inderfurth & Vogelgesang (2013) contributed to the idea of safety stock held for significantly variable inventories. Authors dedicated their study to manufacturer's stochastic inventory problem under periodic review and existing methods for safety stock determination to cope with uncertainties in demand and other variables, such as yield randomness. As the case study were performed in agricultural area, the variation in manufacturing lead time there are dependent from many different assumptions therefore are even higher than in other manufacturing sectors. Caused by non-zero manufacturing lead time an approaches for calculating the dynamic safety stock for different models of inventories to decrease the risk of uncertainties and shortage in stock that may appear were suggested. Korponai, Tóth, & Illés (2017) made a research on the same topic to understand the effect of the safety stock on the probability of occurrence of the stock shortage. In the study performed authors analysed the relation between stock level and the risk of shortage, which could be reduce by safety stock. Results of the study emphasized the relation in inventory management strategy where the more strict expectations on frequency of inventory shortage lead to higher costs for prevention of shortage. In practice the best way would be to find the balance between the level of customers' satisfaction and optimal cost of inventories management, however due to some influencing unpredictable factors, authors left the modelling part for the further researches. Jung, Blau, Pekny, Reklaitis, & Eversdyk (2008) analysed the integrated safety stock management of multi-stage supply chains under production capacity constraints. The authors distinguished three components from where the complexity in safety stock management arise:

a) the non linear performance function that relate the service level, anticipated inventory with safety stock variables of management in each of sites;

b) the interdependence of various sites' performances;

c) The margin by which production capacity exceed the demand which is uncertain.

The authors highlighted that each supply chain manufacturing or warehousing stages display its specific performance functions which relate the level of service and the level of safety stock to the safety stock control variable of the target inventory or the base-stock level. Due to this reason for the result with the high assurance to receive in research authors propose linear programming formulation to evaluate the function for each product individually at each site. Idea of different safety stock levels in different stages was also analysed before by Graves & Willems (2003). In a supply chain where each stage is regulated by a base-stock policy and the presumptions exist that there is an up-per bound for the consumer demand and infinite capacity constraints, the authors create a model for determination of safety stock levels. Ideas formed were continued to study by Boulaksil (2016). Author proposes an approach which let to optimize the safety stock placement in different stages of supply chain under the assumption of demand uncertainty. The simplified schema to understand the modelling was suggested by author:

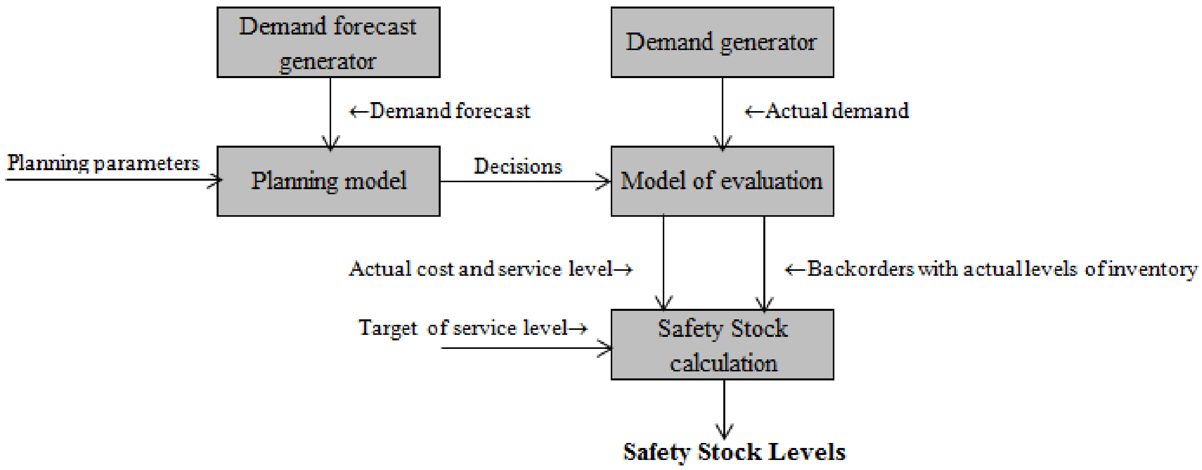

Fig. 3. Schematic review of the approach on safety stock placement

Source: prepared by author based on (Boulaksil, 2016)

The model prepared assumed that supply chain nowadays is fully control by central authorities where an Advanced Planning Systems are implemented, therefore all the needed parameters, forecast, costs and target services levels for modelling are supported by system. In the simplified scheme author emphasized three main stages of modelling for safety stock level to receive – planning, evaluation and safety stock evaluation. After the planning level which was frequently solved in a rolling horizon setting, the actual demand was indicated, based on which the all the necessary parameters of supply chain were calculated and distributed for further calculations in model of evaluation. Consumers demand was modelled by following the Martingale Model of Forecast Evaluation mathematical method, which was chosen due to better reflection on consumers demand evolvement for many products compared to when assuming it being independently and identically distributed. The study assumed that supply chain is considered with several stages and backordered opportunity of unsatisfied demand. By doing so the distribution of the actual inventory and levels of backordered may be derived, which lead to setting of safety stock levels in a way that consumers services levels targeted by Planning Systems would be achieved. The approach of the author has its own case based assumptions and limitations, which could not be applicable for all the inventory management situations, however the clearly structured way of the modelling give a beneficial insights of how the process has to be organized to receive a valuable results. Further the short summary of the studies performed on the topic of safety stock in inventory management is provided for comparison of different ideas of authors.

Table 3. Summary of studies on Safety Stock

Authors	Main ideas of studies
(Nechval & Nechval, 1999)	Investigated the effect on estimation risk on decision making involving a simple one-period inventory problem;
(Graves & Willems, 2003)	Idea of different safety stock levels in different stages. Authors create a model for determination of safety stock levels;
Lesnaia (2004)	Identified the problem of the amount of stock to hold at each stage of the supply chain for the company that is facing uncertainty in demand;
(Sitompul et al., 2006)	Study analysed an issue on safety stock placement in capacitated and uncapacitated productions stages;
(Jung et al., 2008)	Analysed the integrated safety stock management of multi-stage supply chains under production capacity constrains;
(Inderfurth & Vogelgesang, 2013)	Authors dedicated their study to manufacturer's stochastic inventory problem under periodic review and existing methods for safety stock determination to cope with uncertainties in demand and other variables;
(Boulaksil, 2016)	Author propose an approach which let to optimize the safety stock placement in a different stages of supply chain under the assumption of demand uncertainty;
(Korponai et al., 2017)	Made a research on the the effect of the safety stock on the probability of occurrence of the stock shortage.

Source: prepared by author

The overview of studies disclosed different problems under the safety stock perception. The main concepts where studies reviewed were focused on – optimization of safety stock placement in each different stage in supply chain, uncertainties in demand, stock shortage and backordering. However all the different authors faced some limitations in their studies therefore the conclusions under the safety stock and fixed order inventory systems are limited. Such limitations are faced by each of the research as there are no single perfectly suitable system for all the industries and businesses. Therefore the analysis of specific problem first require to have clear understanding of the sector in combination with review of the studies and models available for creation of the best suitable approach.

1.1.3. Vendor managed inventory system

In the practice, there are one more system which is used to manage the inventories between the retailer and his ventor – Vendor Managed Inventory system (VMI). The only different from traditional inventory management here are the responsibility, as in the traditional inventory management systems retailer makes its own decisions regarding the necessity of the order, reordering time or order size, while in vendor management system, the vendor is decision maker, who set the order size and other related parameters for both. Achabal, McIntyre, Smith, & Kalyanam (2000) stated that the system was introduced by vendors to improve both – customer service level of retailer and turnover of inventory. According to authors for VMI systems it is easier to achieve the objectives set due to more accurate sales forecasting methods and higher efficiency in allocation of inventories in the supply chain. From the perspective of vendor, the system let to manage the inventory without information distortion, by giving the opportunity to control the production replenishment of the multiple suppliers at the same time. Such kind of control give vendor an opportunity to adapt a common replenishment cycle for all the retailers in order to reduce inventory level and costs (Yu, Chu, & Chen, 2009). For the retailer VMI systems allow to expand the demand and assortment of the supplier's production by having the time to focus on marketing decision instead of inventory management. Such kind of cooperation in inventory management and inventory brand improvement may be profitable for both – vendor and retailer. Retailer as he is working alone is not capable to achieve the highest measures of productivity and profitability as some important elements such as more responsive replenishment systems could be achieved only with the involvement of vendor and more precise data of demand received (Achabal et al., 2000). Moreover, the system may prevent from stocking of undesired inventories and on the same time can lead to overall cost reduction. The support on issue of replenishment frequency there as well could be suggested to retailer. Tyan & Wee (2003) in their study on Taiwanese grocery industries identified the performance indicators those are measured by different parts when they are applying the VMI system.

Table 4. VMI system performance indicators

Dimensions	Retailer	Supplier
1. Reduction of cost	Inventory Cost	Allocation cost
2. Productivity	Retailer inventory turnover Sales-inventory ratio	Distribution/allocation centre inventory turnover Overall efficiency of production process
3. Inventory level	Purchase cost Retailer inventory level	Production cost Vendor inventory level
4. Sales increase	Products availability/stock shortage Changes in sale	Demand forecast accuracy Changes in quantity of orders
5. Consumer satisfaction	Product fill rates Sales rate	Distribution service level

Source: prepared by author based on (Tyan & Wee, 2003)

The valuation of indicators listed in combination with VMI system application in real inventory management operations may lead to great results. Success stories where the benefits from the VMI system received are widely reported. Tyan & Wee (2003) in their study mentioned high technology industries such as Dell, HP and ST Microelectronics those operated efficient supply chains through the VMI and by doing that were able to reduce inventory level and cost. Next very well known example highlights the big world companies of Procter & Gamble (P&G) and Walmart those partnership based on VMI system in 1985 led to achieve great success on improvement of P&G's on-time deliveries and sales of Walmart, with the increase of both companies inventory turnovers by 30 % (Yu et al., 2009). There are two main characteristics of VMI identified by Yu et al. (2009): cooperation with retailer, and rights of vendor as the cooperation partner in VMI system to know the information of retailer inventories and market. The second give vendor an opportunity to provide the best decision based on retailer situation. However at the same time, it is one of the main reasons, why do some retailers refuse to apply the VMI system.

As different opinions exist, in practice various studies were performed to analyse the system in order to suggest the best possible methods applicable for different cases. Sadeghi, Sadeghi, & Niaki (2014) in study considered the VMI model in a supply chain with one vendor and several retailers, in which the determination of the different machines that works in order to manufacture a single item was considered. As in VMI system, the key role is leading by vendor, the main decisions taken by vendor have always reflected the need of the entire supply chain. The vendor of study faced two limitations in the number of orders and budget together with an intension of retailer to minimize the cost and extend the profit. As the results non linear integer programming model was developed to determine the order size, the shortest possible way and replenishment cycle for both – vendor and retailer.(Sadeghi et al., 2014) Other related authors Taleizadeh, Noori-Daryan, & Cárdenas-Barrón (2015) also prepared the inventory management model for two-level chain composed of one supplier and several retailers with aim to monitor the problem of the products with different deterioration rates in the supply chain. The process of model proposed combined all the main elements of – retail price optimization, frequency of row materials replenishment, products cycle replenishment and the rate of the production in a combination with the main objective to maximize the profit and minimize the costs. For development of inventory model author applied Stackelberg approach to determinate the relationship between the supply chain members where on the leader role is vendor (supplier) and on the follower role are retailers (buyers). For the optimal solutions to find the profit function in combination with solution algorithm were developed. In the study results provided, it was confirmed the hypothesis set that the profit of the entire chain with its members increase by increasing the deterioration rates of the row materials and finished products. Akbari Kaasgari, Imani, & Mahmoodjanloo (2017) their studies also designated to model on topic in supply chain of perishable products. The authors believes that VMI is not only inventory management strategy which application can reduce the costs, increase responsiveness and improve cooperation between the members of the supply chain, but also can reduce deterioration in perishable products supply chain. In study inventory management model created at two-level supply chain with single supplier and multiple retailers. Based on assumptions of study, the product is considered no longer available to use after passing the specific time of product lifetime that is called the critical time. As after the critical time the product is no longer available to sell the authors in study added one more parameter of discount that the management system may use to stimulate the demand for the fixed period before the product lifetime is over. The final model was formed on nonlinear programming principles. The model proposed minimized the total cost of supply chain and production time needed to supply the inventory of each retailer, what resulted to better rate of deteriorating inventory. An inventory management of perishable products for sure is one of the main issue on retail inventory management and it is confirmed by numbers which shows that only in grocery industry of U.S. in 2015 perishable products sales account for 55% of total sales (Chen, 2018). In addition to data of sales presented Chen (2018) performed a study regarding the decision issue for perishable products in production inventory with promotion and pricing for a single-supplier and multi-retailer system under the just-in-time shipment policy. Author identified that the problem of integrated inventory and pricing optimization models for perishable product had been analysed frequently, while there are limited number of studies existed on integrated models of VMI under the policy of just-in-time shipment (JIT) conditions. JIT shipment conditions provide the vendor with rights to split the production lot into some smaller lots for delivery where multiple retailers share one transportation truck to deliver cyclically and by doing that may received the inventory more often. To solve the model author used a metaheuristic algorithm named the quantum-behaved particle swarm optimization method. Study provided an interesting findings those may help an enterprises to implement JIT production together with the inventory control of VMI for perishable products. The authors Tarhini, Karam, & Jaber (2020) analyzing also not so often realizable topic regarding the VMI combination with Consignment Stock policy, which according to the authors may lead to successful coordination strategy for both vendors and suppliers. The main features of Consignment Stock policy is:

a) the buyer is assured that suppliers will continue supplying them with the demand in stock between two levels of minimum and maximum, thus ensuring of no shortage in stock;
b) supplier may store the inventory in buyers' warehouse;
c) buyers pay the supplier only for the products removed from their stock;
d) suppliers have details on the patterns in the needs of their buyers.

The study focuses on the case of one supplier with multiple buyers cooperating with the possibility of transhipment between buyers through multiple redistribution hubs under VMI Consignment Stock policy. The hypothesis whether transhipments between buyers can reduce the total costs faced by supplier and buyer wants to be checked. The cost function and genetic algorithm were developed to solve the model and receive the results. The finding revealed that the implementation of buyer-to-buyer transhipments helps the joint optimum total cost per unit time to be decreased. The model developed may assist managers to make better decisions that can allow them to reduce their spending as well as to upgrade the work of the systems. Another interesting study in retailing inventory management was released by de Maio & Laganà (2020) who analysed the application of VMI system in managing of supply chain of a specific Enterprise in Eastern Asian city, focusing on last mile deliveries in the environment of urban. The topic chosen as the efficiency of last-mile logistics is a huge problem for many enterprises of urban areas in Eastern Asian, who need to adopt operating methods every day in order to reach a successful trade-off between the consumers' loyalty and expenditures in transport. The objective of the study – was by modelling to look at the potential value of VMI strategy application comparing with the traditional strategies of inventory management. In order to find a good quality solution the Multi-depot Inventory Routing problem was formulized for solvement of which the matheuristic methodology was applied. The optimization model created was tested under the real network of enterprise. Result revealed the decrease in total distance travelled in combination with evidence on potential saving cost by applying VMI strategy in the last mile deliveries. According to the authors in numerous large cities, especially in emerging markets, the scenario mentioned is quite similar, therefore introducing the VMI strategy may be a good opportunity to save money in terms of transportation costs for the last miles. Unlike the other authors who analysed supply chain network of only two-channels Han, Lu, & Zhang (2017) studied an optimal inventory planning issue on three-channels supply chain network consisting from vendor, several retailers and multiple third-party logistics companies as distributors. The paper of authors had three main findings. First, study provided a tri-level decision model to define the problem of decentralized VMI, which let to understand how the members of the supply channel cooperate with each others. Second, the study focused on analysis of the solution space of the resulting decision model and proposed the efficient vertex enumeration algorithm for an optimal solution to find. Third, the study and the model created displayed how to improve the individual performance of each member of supply chain and to balance the total cost sharing in the centralized VMI system where the vendor and buyer may minimize inventory to zero. However it is important to highlight that the model is created under the deterministic demand of buyers, therefore require further examination under uncertain environment before the full model application into the real businesses. Further the short summary of the studies performed on the inventory management strategy of VMI is provided for comparison of different ideas of authors:

Table 5. Summary of studies on Vendor Managed Inventory

Authors	Main ideas of studies
(Sadeghi et al., 2014)	Considered determination of the different machines that works in order to manufacture a single item under VMI;
(Taleizadeh et al., 2015)	Created the inventory management model under VMI to monitor the problem of the products with different deterioration rates in the supply chain;
(Akbari Kaasgari et al., 2017)	Analysed VMI modelling on topic in supply chain of perishable products;
(Han et al., 2017)	Studied an optimal inventory planning issue on three-channels supply chain network consisting from vendor, several retailers and multiple third-party logistics companies as distributors;
(Chen, 2018)	Performed a study regarding the decision issue for perishable products in production inventory with promotion and pricing under the VMI and just-in-time shipment policy;
(Tarhini et al., 2020)	Analysed inventory management under VMI combination with Consignment Stock policy;
(de Maio & Laganà, 2020)	Analysed the application of VMI system in managing of supply chain on last mile deliveries in the environment of urban.

Source: prepared by author

The overview of studies disclosed different combinations and applications of VMI possible based on situation analyzing. The main concepts where studies reviewed were focused on – inventory optimization, reduction in transportation and holding costs, profit maximization and increase of customer satisfaction. However as studies reviewed were every time based on specific assumptions with different limitations identified, the models developed required deeper analysis before an application into the real inventory management workflows of business. Summarizing all the systems of inventory management reviewed, there are many different studies and views available. Inventories may be managed by company itself by choosing fixed time or fixed order size replacement, either inventory may be managed by vendor, if both parts agree to the conditions needed. The decision as to which system to use has to be made based on specifics of the business. In retail, looking from the perspective of the daily nature of store replenishment the most applicable one probable would be the Fixed Time Period System that ensures proper time management and flexibility in order quantity. Regardless of the decision of system taken in all situations stock levels needed should be driven by analytical tools those works with big data on a basis of scientific methods incorporating most important features of inventory system used. There a different IT applications those are used by companies today. Before selection deep analysis for specific of each is needed. Some of the well known will be presented and compared in the next section.

1.2. IT applications for inventory analysis

Global changes facing every day, grow the amount of assortment in stores dramatically. If several decades ago there was only limited choices possible, today we are free to buy almost everything we want all around the world. And here almost the highest challenges are faced by information technologies and different applications those have to maintain the amount of stock at the right level to perform. There are many different IT applications available for business for various analyses to organize. The work focuses on processes of demand prediction and inventory optimization. Microsoft Power BI, IBM SPSS Modeller, KNIME and Apache Hadoop are some of the well known IT applications therefore selected by us to review.

1.2.1. Microsoft Power BI

Microsoft Power BI is a business intelligence platform which provides a comprehensive collection of analytic tools for different teams. It is the latest offer of Microsoft's Self-service BI. Analytical platform provides business user focus data analysis and visualization capabilities to enhance business insights and decision making process (Lu, 2014). Microsoft Power BI is a cloud based analytic tool, which provide an ability to easily create and deploy solutions with different data sources such as – data from cloud, on-premises data sources, various other systems and applications. According to author Lu (2014) to understand the tool, it is important to know how it is constructed. Microsoft Power BI consists of three different concepts or parts:

Self-service BI features in Excel. It consists of four features inside: Power Pivot, Power Query, Power View and Power Map. Those entire features are used to supply business with all the wide range of possible data analysis capabilities.

Power BI for Office 365, which provide ability for sharing the reports, datasets and all the other important business solutions online. In this way the tool enables the business to work proactively, responding quickly to all changes in the market and internally.

IT infrastructure for Power BI, which formulate a simply way of data management and administration of data on premises and in clouds.

Today when most of the companies are dealing with big data and cloud services, decision making without help of strong analytical tools would be hard to imagine. The services of Microsoft Power BI in decision making process are used by such companies as HP, Pepsi or Toshiba. An interested study by Abusager, Baldwin, & Hsu (2020) illustrated the usefulness of the program with the concrete case of application in the real environment. Case study was performed in one laboratory where the tendency of laboratory results with likely overestimation of true incidence with further ordering of medicines needed was identified. The system by itself required an improvement therefore specific dashboard was developed using Microsoft Power BI. The dashboard created and the root-cause database used allows analyzing diagnostic ordering trends in order to identified gaps to decrease unnecessary testing. The rates presented after the improvement disclosed a positive trend in rates reduction over the period of years. There are more studies showing the positive application and improvement with the Microsoft Power BI. Company by itself also investing a lot into continuous improvement. In August, 2020 the article comes where Microsoft was disclosed as a company, which has used a hydrogen fuel cell system to power a row of data centre servers for 48 consecutive, what shows that the company is innovative and curious into nowadays technologies application in IT ("Microsoft data centre servers run on fuel cell backup power for 48 h," 2020). From the user perspective, as the main benefits of the platform users emphasized an affordability, interactive visualization, integration, data accessibility and frequent improvements. Of course, in addition to benefits listed, there are areas for further improvements suggested by users. Those could be a need for wider options of configurations of visuals, as users found it limited in options for changing. Crowded interface also was mentioned as confusion to the platform users. Moreover consumers of tool found it difficult to learn about more complex tools of Power BI, as the main information of guiding the platform is mostly concentrated to basic reporting (Dataflair, 2019).

1.2.2. IBM SPSS Modeler

Other well known among the big data analytics is IBM SPSS Modeler analytic platform. The platform could be defined as leading, visual data science and machine learning solution. The platform works in a way to help the users to construct accurate predictive models for individuals, groups and the systems of enterprise. Leading organizations worldwide rely on

IBM for data exploration, predictive analysis preparation, model management and implementation, and machine learning to monetize data assets. The IBM SPSS Modeler empowers organizations to use complete, out-of-box algorithms and analysis techniques, including text and business analytics, decision management and optimization to deliver insights on a real-time. Moreover, platform also use different models to access data assets and modern applications, appropriate for hybrid, multi – cloud environments with robust governance and security posture, which is so actual today. Jouve, Martin, & Guerin (2012) analysed IBM SPSS Modeler application on a real time under case study of sentiment-based text analytics to better predict customers' satisfaction in calculation of customer loyalty metric of Net Promoter Score (NPS). The ratio that could be used in all the different situations of valuation in satisfaction of company, service or specific goods defines customer satisfaction on a scale from 0 to 10. Based on results received customers are segmented into three groups: Promoters, Passives, and Detractors. Such kind of the surveys is often used by marketers in order to feel the emotion of the company client database. The concern exists that NPS is extracted only from objective responses that do not represent the explanations why do the customers are thinking and reacting in the way they are. However using the free-text consumers' reviews enable business truly listens to the customers' opinion, their understandings and complains and text analytics here may help by making it easy to grasp and measure the data. By using the IBM SPSS Modeler authors of the study created relevant categorization model based on features and functions extracted from unstructured data. In this specific case studied the majority of the consumers complain were received due to poor services of the reception what was highlighted as a focus for company to improve. It is also important to emphasize that users of IBM SPSS Modeler considered the program easy for beginners to request analyses, while yet offer a fairly wide range of more advanced features. Data files can be imported from many different programs. Moreover, according to users of the platform, it offers many valuable data handling procedures, such as ability to merge files, specify value labels and compute new variables from existing variables. However there are also some areas for improvement. According to the users of the platform – documentation about algorithms is sometimes hard or impossible to find. Moreover, default graphics are far from publication quality and need to be implemented. Only the paid version providing more capabilities, yet it is expensive comparing to many others competitive alternatives.

1.2.3. KNIME

The third analytical platform we want to review is Konstanz Information Miner (KNIME). It is cloud-based solution that provides businesses with tools to configure data science workflows using predictive analytics. KNIME analytical platform is open-source, thus the original source code is made freely available and could be redistributed and implemented. KNIME is written in Java and based on Eclipse uses its extension mechanism to add additional functionality to plug-ins. Core architecture of platform allows working with large volumes of data that are limited only by the available space in hard disk. Additional plug-ins in Modeling allows integrating such methods as text and image mining, time series analysis and network. As being the open-source analytical platform KNIME integrates various others open-source projects of machine learning algorithms to use. A graphical users interface and use of JDBC provide assembly of nodes that combine various data sources for data understanding, pre-processing, modeling, evaluation and finally model deployment for real cases by using the interactive views (Nikolla, n.d., 2018). By programmers the platform is considered as a high quality alternative to Statistical Analysis System (SAS). Starting from 2006 when the analytical platform was released several pharmaceutical businesses started to apply KNIME in a number of life science studies. Moving from establishment years due to numbers of possibilities and alternatives of use an interest in platform has increased dramatically. The KNIME analytical platform today is used in various different areas of analytics such as – pharmaceutical researches, customer data analysis and predictions, business intelligence, data mining or financial data analysis. The platform is used by such worldwide companies as Comcast, Johnson & Johnson and Canadian Tire. By users analytic platform is a favorite for its high amount of opportunities in analytics and prediction capabilities available with only minimal knowledge of programming needed. As platform works under the free principle, users are free of publishing their workflows what could help the users across many different organizations to study and adapt the relevant models and workflows by themselves. Modeling in platform consists of simple ETL operations, valuable algorithms, highly usable and organized workflows, automatization in most of the manual work and easy set-ups. What is more important, the KNIME is very well integrated with other technologies and languages therefore it ensures easy application and implementation of workflows creating. The platform is free, only some further extension and use of different servers required some payments to perform, what leads to the higher popularity among small businesses and individual researchers. However, as constant users of the workflow have identified, the KNIME cannot fully replace regular reporting tools, as reporting function still require some improvements to proceed (Nikolla, n.d., 2018). Moreover, some improvements could be implemented in data handling capacities and integration with graphic databases.

1.2.4. Apache Hadoop

The last big data analytic tool to review is Apache Hadoop. Important to mentioned, that for years, although the computing ability of application servers has improved in many ways, owning to their restricted storage and speed, databases have lagged behind. However, today, when online possibilities is growing and many applications are producing large data to be processed, such frameworks as Apache Hadoop play an important role in providing the database environment with a much-needed makeovers Apache Hadoop is an open source mostly Java programming language with only some native codes in C based framework applied to store and process big data. A framework is a collection of open-source software utilities that facilitates by using a network of many computers to solve problems involving huge amounts of data and computations. By using the MapReduce programming model Apache Hadoop provides a software framework for distributed storage and processing of big data. The Hadoop was originally designed for computer clusters to build from commodity hardware, which is still in a common use. Since then framework also was found useful on clusters of higher-end hardware. The fundamental assumption that the framework provide is detection of failure. Rather than rely on hardware to deliver high-availability, the Hadoop framework itself is programmed to detect and handle failures at the application layer, so delivered a highly available service on top of the cluster of computers, each of which may have a potential to failure. The core of Apache Hadoop includes the part of storage knows as Hadoop File System and part of processing known as MapReduce programming model. Apache Hadoop partitions data into large blocks and distributes them in a cluster across nodes. To process the data in parallel it than passes packed code into nodes. This technique takes benefit of the location of data where nodes control the data to which they have access. This makes it possible to process the dataset faster and more effectively than in a more traditional supercomputer architecture built on a parallel file system where storage and data are transmitted through high-speed network. The base Apache Hadoop framework is composed of the following modules:

– Hadoop Common: consisting of common utilities that support the other Hadoop modules.

– Hadoop Distribution File System: a distributed file system providing high-throughput access to application data.

– Hadoop YARN: framework providing users' application scheduling and cluster resource management possibilities.

– Hadoop MapReduce: YARN based system providing the parallel processing of large data sets.

– Hadoop Ozone: an objects store for Hadoop.

Due to wide possibilities in application, Apache Hadoop analytic tool is well known between the companies. Some of the big names using the tool are Amazon Web Services, IBM, Facebook, Microsoft and Intel. As one of the main strength of the Hadoop users highlighting its Distributed File System allowing to hold all the different types of data including video, images, text of XML together. Moreover, the tool is highly beneficial for research and development (R&D) purposes, provides quick access to data, is highly scalable and supplies with highly-available service resting on a cluster of computers. However, users as well see some areas to improve which they expressed as issues with small files, problems with disc space, special knowledge requirement on MapReduce for usage of clusters and security risks as according to the users Hadoop lacks the level of security functionality for safe enterprise deployment, especially if it cancers a sensitive data of company. Therefore even with the high interest and number of applications, there are still places for upgrade existing where the owners of the tool have to focus on for further improvements.

Summing up the comparison of different IT applications – today companies have the wide range of choices. While each application has some strong sides, they do have some areas for improvement as well. For our further modelling we selected to use KNIME analytic platform. Decision was based on platform availability, positive reviews and possibility to apply multiple decisions and techniques while modelling the prediction without prior programming knowledge required. However it is important to know that all the inventory analytic systems works based on mathematical methods and algorithms selected. The choice of the mathematical method may have an impact to accuracy of final result. Therefore further different mathematical methods for demand prediction modelling will be analysed and compared for the best method for an application in practical part to select.

1.3. Mathematical methods for demand prediction modelling

Demand prediction is the first and very important step of inventory optimization. By itself it is the combination of two words. The first one – demand, and second – prediction. Before combining those words into phrase, and further reviewing different methods applicable, it is important to understand economical meaning and value of this phrase. Word demand – could be define as requirement of products and services, while prediction – in general, means making estimation in present for future events, at this case – future demand of products. Demand plays a vital role in the decision making of a business. In competitive market conditions, there is a need to take correct decision and make planning for future events related to business like a sale, production or inventory optimization. The better analysis of different factors are performed – the better decisions based on demand prediction results could be made. As the process by itself is difficult, there are a lot of different methods invented and studies made based on the applications of those methods. In this part – application of the methods of Linear regression, Probabilistic Neural Network, Bayesian Additive Regression Trees, Random Forest and Fuzzy Logic will be reviewed and discussed.

The first method – linear regression is the basic and ofen used type of predictive analysis. Linear regression uses only one independent variable as a predictor, which has an effect to dependent variable (outcome). The main idea of regression analysis is to determinate the strength of predictor, to forecast an effect, and to use the model formed in forecasting the further results. However, the method when there is only one predictor is not clear enough to predict possible subsequent development of the analyzing process, therefore the authors commontly choose a Multiple Linear regression. This means that linear regression can be extended and instead of one independent variable, we have a many different variables (Anghelache, 2015). As the method is treated as clearly understandable, it is widely used in various studies and researches to predict outcomes selected by different authors. Author (Anghelache, 2015) used multiple linear regression to analyse the final consumption and gross investment influence to Romania GDP. Based on the data collected authors formed an equation and based on statistical tests evaluate the accuracy of the model. Author conclude that higher number of factors in regression model allows the researcher to draw more conclusive results in macroeconomic analysis. Authors (Aghdaei, Kokogiannakis, Daly, & McCarthy, 2017) also used a regression in building of energy simulation model. The aim of the study was to predict the space heating and cooling requirements in different cities of Australia. Findings of the research show that the linear regression with simple independent variables can predict the requirements for space heating and cooling of the residential buildings in the specific climates within acceptable errors. It can even be applied in the studies where the relation between the financial news as an independence variable and the stock price of financial market is evaluating. In this case the author of the work named regression as machine learning-based approach (Ihlayyel, Sharef, Nazri, & Bakar, 2018). However, as method is quite simple, for better results it has to be used in combination with some rules, methods or algorithms. In the researches mentioned above to reduce the modelling cost of the parametric analysis authors (Aghdaei et al., 2017) use methods of Taguchi and ANOVA. Other author Ihlayyel mentioned above applied Enhanced ELR-BoW algorithm. Only after simplification of the data regression models were formed. The research of the authors (Cankurt, Subasi, 2015) also tried to compare the linear regression with neural network and agree with the idea above. Formed forecast shows that neural network present higher accuracy than the linear regression when there is no combination with linear regression and other methods.

When it comes to regression it is always important to mention – logistic regression. Unlike traditional linear regression – logistic regression is appropriate for modelling a binary variable. In wider perspective it means that results can procedure two outcomes (1 or 0), those could be considered as "positive" and "negative". Such results are useful in the practice however can't be received by using simple linear regression. This is mostly due to two main reasons. Firstly, a simple linear regression can only predict values outside the acceptable range. Secondly, as the dichotomous experiments can only have one of two possible outcomes for every experiment, the residuals will not be normally allocated about the predicted line. Performing analytics with logistic regression includes three main goals: prediction that the outcome or response variable equals to 1, categorization of outcomes and predictions and finally, access to the odds or risk associated with model predictors (Grömping, 2016). Logistic regression is considered as very important statistical procedure in predictive analytics in areas of health-care, medical analysis, social statistics and economy. The authors Joubert, Verster, & Raubenheimer (2019) included logistic regression as commonly used method to predict probability components and loss severity in study of Loss Given Default (LGD) evaluation in banks. In authors study probability components were modelled by making use of logistic regression binary outcomes (write-offs or not write-offs). Moreover in study logistic regression was used in combination with method of survival analysis, what let to increase the model's predictive power and accuracy of results.

As it is mention above, one more method often included into studies is – Probabilistic Neural Network (PNN). PNN is the method of artificial intelligence that allow to form a complex nonlinear relationship between response variables and explanatory ones. The network was introduced in late 20ths – in 1990 by Specht. The main characters after presentation of network were – easy to use and possibility to interpret the network 's structure in the form of a probability density function which is simple to understand. For those reasons the method is used in various sectors to analyse. Authors Penpece & Elma (2014) used neural networks for the purpose to forecast Sales Revenue in Grocery Retailing Industry. Based on results received authors stated that neural network method is more organic and predicting results better than the other methods. Revenue forecasts calculated based on neural networks were very close to actual data of sales revenue. However, the model also faced some problems of estimation of probability density function and high space complexity of PNN pattern layer. Moreover, model by itself has only one parameter of training and the smoothing parameter (σ), which must be optimized in order to make the network achieve the highest prediction ability. Based on different scientists' network is composed of 3 either 4 layers: an input layer, a pattern layer, a summation layer, and an output layer. Some scientists (Sun, Wu, & Li, 2017) does not count last layout and named third layout as Classes. The neurons in the input layer are simply the features of input vectors. The pattern layer consists of as many neurons as training examples. In the summation layer, the number of neurons is equal to the cardinality of classes in the data set. Finally, the output layer consists of a single neuron that provides the classification result. As the structure of the network is considered as a complex and probabilistic neural network is a frequently exploited model in the field of data mining in different researches of scientists, certain PNN reduction techniques have been established. These techniques include dynamic decay adjustment algorithm (DDA) (Berthold & Diamond, 1998) backpropagation mechanism (Sun et al., 2017), dimensionality reduction (Kusy, 2015) and also some other, presented earlier (in 1991-1994) – learning vector quantization (Burrascano, 1991) or maximum-likelihood algorithm techniques. All the techniques have the own specific parameters and the influence on the network. Further mainly parameters and the practical value of techniques are reviewed:

– Dynamic adjustment algorithm (DDA). Operation of algorithm required two phases – training and classification. New neurons are added if necessary. Less than five epochs are needed to complete training. The algorithm can be proven to terminate when a finite number of training examples is used. And finally, only two thresholds are required to be adjusted manually (Berthold & Diamond, 1998).

– Dimensionality reduction. For the reduction creation model by author (Kusy, 2015) two main steps – feature selection (methods of single decision tree (SDT) and random forest (RF)) and feature extraction (method of principal component analysis (PCA)) need to be follow. The main idea of the SDT is treated method as a predictive model – it maps the input data into desired targets. If the desired targets take the form of groups to which the data belong, SDT is treated as classification tree. After this process RF utilizes the collection of independent decision trees formed by SDT. Within the training process, the trees grow in parallel, not interacting until all of them have been grow. Once the training is completed, the need to move to the next phase appears and predictions of single trees are combined to make the overall prediction of RF. For the further step – principal component analysis is used. PCA is one of most popular feature extraction method to use. The methods combine the statistical technique which converts a set of input features into a set of new values by means of linear transformation. The results are names principal components and are linearly uncorrelated. With the help of method patterns of similarities and differences in data can be identified. Once these patterns are determined, the data can be compressed by decreasing the number of dimensions without significance loss of information. According to the author Kusy (2015), who applied this variation of method for medical data classification tasks, the results showed an increase of prediction ability and decrease in computational time needed to complete the task in every single case.

– Backpropagation mechanism (BP). The authors (Parry, Cao, & Song, 2011) define the method as feed-forward neural network and included it into the test of prediction accuracy in the first-time adoption of DVD players. The authors forecasted performance of the logic model and the three neural network models. At the end it was found that the PNN algorithm significantly outperforms the logic model and the two remaining neural network algorithms. BP was one of those methods. However in 2017 scientists (Sun et al., 2017) present the idea of new PNN model in combination with backpropagation algorithm. New BP-PNN model has two phases – learning phase, where the idea is to receive the initialized value of the variable weights and training phase – where the error function is propagated back. Based on the analyses performed by the authors comparing with PNN, BP-PNN has fewer components in the pattern layer, which helps to reduce the space complexity of the model. Moreover, comparing with PNN, BP-PNN is designed with the much clearer structure and ability to identify the importance of indicators. Nevertheless, there are still some limitation of model emphasized by the authors – the model required further studies as the number of parameters trained is higher than in other models and thus requires a long time for calculations in the model.

Review emphasized that the method of PNN can be used for forecasting. However classical PNN needs to be combining with one of the algorithms reviewed for the purpose to receive the better results.

One more popular method – Bayesian Additive Regression Trees (BART) by authors Pratola et al. (2014) were described as nonparametric method. Logan, Sparapani, McCulloch, & Laud (2019) define BART method as fully nonparametric and flexible model of prediction which can deal with complex functional forms as well as interactions among wide range of variables The authors Ajidarma & Irianto (2019) in their study agree with the mentioned definition by adding the idea that BART regression method harnesses dimensionally adaptive random basis elements. The method consisting from a prior and likelihood and is a function of an ensample of trees. As method can combine and compare a number of different factors it is widely used in the researches of different authors. Method can use different algorithms with whose help more precise search of the model space and variation across the algorithm draws can be organized. However there are some problems as publicly available version of algorithm in R package can process only the limited number of observations (Pratola et al., 2014). Moreover the authors Linero & Yang (2018) performed a study for one more problem of decision trees to analyse. According to the authors in those methods - a high possibility of deficiency in ensembles is possible and could be caused by lack of smoothness and vulnerability to the curse of dimensionality. The idea of soft decision trees was suggested and implemented on the BART method. The authors demonstrate that their methods can have meaningful improvement over existing methods. Yet as there are still a lot of limitation, further studies of idea is performing. Ajidarma & Irianto (2019) have included the BART in the analysis of the Growth of Electric Vehicle market in USA. With an application of BART method study goals to predict the growth of electric automobile industry in different states of United States. BART method was used to analyse the relationship between several factors chosen and the sales of electric vehicle. With help of the algorithms (on this case Markov chain Monte Carlo (MCMC) algorithm was chosen) four BART models were generated and fitted into the data. The models identified the top predictors those correlations with the sales of electric vehicles were confirmed in the further steps. Authors of the work concluded that to use the method was beneficial – as the method enables a full assessment of prediction uncertainty while remaining highly competitive in terms of prediction accuracy.

Moving forward to Random forest method – it is one more decision trees method which consist of a chosen number of decision trees, which are used for classification and regression analysis (Feng & Wang, 2017). Authors Gupta, Rawat, Jain, Arora, & Dhami (2017) in their study of decision methods described this random forest method as tool that form the ability of multiple varied analyses, organization strategies, supply and demand prediction modelling, perceptive variables and importance ranking on the record-by-record basis for deep data understanding. Instead of tool authors Yin, Lee, & Wong (2012) define random forest method as the algorithm with an ensemble random method. Author Ghatasheh (2014) agree with the definitions mentioned and emphasized some more positive attributes, such as an immunity to over fit, good estimation of internal errors, and high accuracy comparing with other learning algorithms. Because of these reasons' method is included in various calculations and studies performed. In 2017 authors Feng & Wang made a study, where the demand of the bicycle rental was evaluating. Two methods – multiple linear regression analysis and random forest were included into calculation. However, research shows that accuracy of linear regression model to forecast is too low even though the normal distribution of factors and good relationship between them was identified. Authors identified the high possibility of error due to specific characteristics of factors. The method was change to random forest method and this improves the accuracy of result to 82%.

The last method reviewed in this article – Fuzzy method. The authors Agápito et al., (2019) characterize Fuzzy method as logic which with a help of specific set of rules can make an associations between linquistic and numeric data of a database. Rules can be provided by two groups – artificial intelligence algorithms, which are the targer of researches in nowadays and by group of experts. The author Syahputra (2016) in her study agree with the definition about the logic, but additionally emphasized that this logical function recognizing only two parametres, aither „Yes“ or „No“ („1“ or „0“). The logic mostly used to create an expert systems and knowledge - based control settlemets. However there are some limitations, those are important to know before analysing the method in more detail. Firstly, the method is case-dependent – every time the changing scenario can have a different influencing factors, where each of them need to be evaluated as important. Secondly, contribution of domain experts is of significant importance in process of forming control settlements (Yadav, Kumar, Kumar, & Yadav, 2018). Moreover there also could be the problem of too many rules. The problem usually arise among the rule-based models, when rules are created for every single factors and the outliers, those were not identified in the begining of modelling are detected only in the process (Berthold, 2003). Despide it, the fuzzy logic method is applying in studies, yet usually in combination with other methods. The author Syahputra, (2016) performed a study to predict a vehicle fuel consumption by using the combination artificial neural networks and fuzzy logic (ANFIS). Prediction was made for different models of cars based on two criterias – weight and age. Results show that with an increase of weight of the motor vehicle, an amount of fuel needed to travel the same distance is increasing. Moreover car age also affects fuel efficiency. For the younger car, the higher efficiency of fuel consumption is calculated. The same combination of methods were also used by author Ridwan, (2019) who used adaptive network-based fuzzy inference system (ANFIS) in the study to predict the price of good – lamp.

As all methods reviewed are different, for better understanding it is important to summarize what are the advantages and disadvantages of every method and what are the areas the method could be applied (Table 6).

Table 6. Advantages, disadvantages and application of mathematical methods proposed

	Linear Regression	Logistic regression	PNN	BART	Random forest	Fuzzy logic
Basic principles of method	Determination of relationship between independent and dependent variables	Prediction and determination of relationship between variables when the dependent variable is binary	Classification of patterns based on learning from examples	Creation of sum-of-trees model and regularization prior on the parameters of that model	Combination of tree predictors where each tree depends on the values of a random vector sampled independently with the same distribution for all trees in the forest	Form of many-valued logic in which the truth values of variables may be any real number between 0 and 1 both inclusive
Authors	Anghelache; Aghdaei;Kokogiannakis, Daly, & McCarthy; Ihlayyel, Sharef, Nazri, & Bakar; Aghdaei et al.; Cankurt, Subasi	Grömping; Joubert, Verster, & Raubenheimer	Penpece & Elma; Sun, Wu, & Li; Berthold & Diamond; Kusy; Sun; Burrascano; Parry, Cao, & Song	Pratola; Logan, Sparapani, McCulloch, & Laud; Ajidarma & Irianto; Pratola; Linero & Yang.	Feng & Wang; Gupta, Rawat, Jain, Arora, & Dhami; Yin, Lee, & Wong; Ghatasheh	Agápito; Syahputra; Yadav, Kumar, Kumar, & Yadav; Berthold
Advantages	− Simple estimation procedure; − Eeasy to understand interpretation on a modular level (i.e. the weights).	– Gives not only a measure of how relevant a predictor is, but also its direction of association; – Is easy to implement, interpret and very efficient to train.	− Less time consumed to train virtually; − Relatively sensitive to outliers; − Can calculate probability scores.	− Provides a flexible approach to fitting a variety of regression models and algorithms while avoiding strong parametric assum.; − Able to represent interactions; − Can handle missing values.	− One of the most accurate learning algorithms; − Runs efficiently on large database; − Provide a reliable feature estimate; − Offer efficient estimates of the test error; -Can estimate missing data.	− Using simple mathematics for non linear, integrated and complex systems; − Is based on linguistic model; − Has rapid operations; −Can handle trouble with inaccurate data.
Dissadvantages	− Can over simplifies the problems, and make a linear relationship where there is no such; −Is sensitive to outliers; − The data using in method must be independent.	– Assumption of linearity between variables. Hard to achieve in real world; – Dependent variable is restricted; –Observations < features, lead to possible over fit.	− Required a lot of memory for data; − Usually long testing time; − Possible large camputational cost.	− In publicly available version of algorithm in R package can process only the limited number of observations (Pratola et al., 2014); − A high posibility of defficiency in ensembles due to lack of smoothness and vulnerability to the curse of dimensionality. (Linero & Yang, 2018)	− An ensemble model is inherently less interpretable than an individual decision tree; − Training a high number of trees could lead to large camputational cost and can use a lot of memory.	− Case-dependent; − Contribution of domain experts is of significant importance in process of forming control settlements (Yadav et al., 2018) − There also could be the problem of too many rules; − Based on the size, high memory and cost required.
Business processes were methods are applied	Inventory, sales prediction, evaluation of marketing effectiveness, promotions.	Supply chain management, inventory, sales forecasting.	Selection of retail stores location, prediction of sales revenue, promotions.	Supply chain management, demand and orders prediction modelling, correlated outcomes.	Prediction of retail demand, optimal retail location , pricing.	Market trend analysis, retail servise quality evaluation, consumption and demand predictions.

Source: prepared by author

The comparison of methods shows that there is no one perfectly suitable method to use. All the methods have their advantages and also areas to improve. Main advantages of the methods: understandable to use, flexible, interactive, accurate, and efficient with the large data sets. Recurring problems identified: possible deficiencies, high memory requirements, long testing time and large computational cost. For further practical application in demand prediction modelling based on the real time retail company data BART method was selected as the method with high flexibility, wide displaying options on interactions and ability to handle missing values. After the examination of inventory optimization systems and IT applications available in retail sector inventory management, for further application we also selected fixed time inventory system and KNIME analytic tool to adapt. As in the first chapter only the different studies with integration of methods and systems were analysed, in the further part of thesis mathematical method selected will be analysed in more detail to understand the specific and features of an algorithm in order to avoid mistakes in method practical application. After the systematization of BART method specularities, all the selections and insights gained through the theoretical chapter will be applied by creating the practical model for inventory replenishment on a real time situation in retail enterprise.

2. BART ALGORITHM IN DEMAND PREDICTION

BART is Bayesian approach to the estimation of non parametric functions using regression trees (Kapelner & Bleich, 2016). The method for further application in practical part was chosen due to BART model ability to capture a complex relationships between X and Y, with the goal of using it in the prediction what is related to the aim of our work (Tan & Roy, 2019). In the model regression trees depends on recursive binary portioning of predictor space in a set of hyperrest-angles for idea to approximate some unknown function f. The space for the predictors has dimension same as the number of variables, which is denoted as p. Tree based regression models are able to flexibly fit interactions and nonlinearities. The BART model includes an effective Markov Chain Monte Carlo (MCMC) algorithm, which already where mentioned above through methods practical overview. MCMC algorithm explores the complex space of an ensemble of trees without presetting the dimensions of each tree. Guided by the prior, the complexity of the model is assumed (George, Laud, Logan, McCulloch, & Sparapani, 2019). Models composed of sum of regression trees have an even better capacity to capture interactions and non-linearities as well as additive effects in f function than single trees (Kapelner & Bleich, 2016). BART can be called an ensemble of sum-of-trees, with a novel estimation method based on fully Bayesian model of probability.

In this mathematical part we found it important to discuss the BART equation in general, prior and likelihood and MCMC algorithm. Those, we think are the most important parts for model to understand.

Specifically, the BART model can be expressed as:

$$Y = f(X) + \varepsilon \approx \mathrm{T}_1^{\mathrm{M}}(\mathrm{X}) + \mathrm{T}_2^{\mathrm{M}}(\mathrm{X}) + \cdots + \mathrm{T}_{\mathrm{m}}^{\mathrm{M}}(\mathrm{X}) + \varepsilon, \quad \varepsilon \sim N_n(0, \sigma^2 I_n)\,, \tag{1}$$

where Y is the $\mathrm{n} \times 1$ vector of responses, X is the $\mathrm{n} \times \mathrm{p}$ design matrix (the predictors column- joined) and ε is the $\mathrm{n} \times 1$ vector of noise. Here we have m distinct regression trees, each composed of a tree structure, denoted by T, and the parameters at the terminal nodes (also called leaves), denoted by M. The two together, denoted as T^{M} represents a whole tree with both its form or the structure and the set of leaf parameters. The task is to be able to infer the function f(X) with nominal assumptions and high dimensional X (George et al., 2019).

A given tree T_{t} structure provides details about how every observation resurrects down the trees. There is a "spilling rule" for each nonterminal (inner) node of the tree, which taking the form $x_j < c$. Form $x_j < c$ consisting of "split variable" x_j and the "splitting value" c. If the condition set by "splitting rule" is met, a calculation moves to the left child node (it moves to the right child node otherwise). All the cycle continues until a terminal node is reached. After that, the calculation receives the terminal node's leaf value. The set of tree's leaf parameters could be marked as $M_t = \{\mu_{t.1}, \mu_{t,2}, \dots, \mu_{t_{b_t}}\}$, where b_t an amount of terminal nodes for a given tree. The estimated value of the observation is then the sum of the values of the m leaf reach by recurrence of all the m trees (Kapelner & Bleich, 2016).

It is important to mention that BART model can be differentiated from other ensemble-of-trees models because of its underlying probability model. BART as a Bayesian model is composed of a set of priors for the structure and the leaf parameters and a likelihood for the data in the terminal nodes. The prior's goal here is to provide regularization, moreover, to prevent the total fit from being controlled and dominated by any single regression tree. That is the reason why it is important to understand the value of priors and likelihood before starting to combine the model of BART.

Prior and Likelihood. According to Kapelner & Bleich (2016) and George et al., (2019) the BART model prior has three components:

The tree structure itself:

parameters of the leaf given the tree structure;

the error variance σ^2, which is independent of the three structure and leaf parameters.

The formula here could be provided as following (2):

$$P = (T_1^M, \dots, T_m^M, \sigma^2) = \left[\prod_t P(T_t^M)\right] P(\sigma^2) = \left[\prod_t P(M_t|\, T_t) P(T_t)\right] P(\sigma^2)$$
$$= \left[\prod_t \prod_l P(\mu_{t,l}|T_t) P(T_t)\right] P(\sigma^2)\,,$$

where the last equality follows from an additional assumption that the leaf parameters are conditional independent given the tree's structure (Kapelner & Bleich, 2016).

Based in the authors, looking into the components, first, there is a P (T_t) – the prior portion that effects node location within the tree. Depth of node is defined as distance from the root. Therefore, the root itself has depth evaluated as 0 its first child node has depth 1 and ect. Nodes at depth d are nonterminal with prior probability$\alpha(1 + d)^{-\beta}$, where $\alpha \in (0,1)$ and$\beta \in [0, \infty]$. This component of the prior tree structure has an ability to implement shallow tree structure, thereby reducing any single tree's complexity and resulting in further regularization of model.

The next component $P(M_t \mid T_t)$ controls the leaf parameters. Given a tree with a set of terminal nodes, every terminal node (or leaf) has a continuous parameter (the leaf parameter), which is showing the "best guess" of the response in this partition of predictor space. This parameter is the equipped value assigned to each node observation (evaluation). The prior on each of the leaf parameters is given as$\mu_l \sim N\ (\mu_\mu/m, \alpha_\mu^2)$. The expectation is that μ_μ is selected to be a range center, or $(y_{min} + y_{max})/2$. The center of the range can be affected by the outliers. However, if this is a problem, the user can log-transform or windsorize the answer/result before building the model (Kapelner & Bleich, 2016).

The variance hyperparameter $\sigma_\mu^2$is empirically chosen so that the range center plus or minus k = 2 variances cover 95% (or other determinated range) of the provided response values in the training set. If there is m trees, it will be chosen than σ_μ such that $m\mu_\mu - k\sqrt{m}\sigma_\mu = y_{min}$ and$m\mu_\mu + k\sqrt{m}\sigma_\mu = y_{max}$. The purpose of this prior is to provide regularization of the model by shrinking the leaf parameters to the center of the response distribution. The higher the value k, the smaller the value of σ_μ^2, resulting in higher regularization of model (Kapelner & Bleich, 2016).

The final prior is on the error variance and is chosen to be $\sigma^2 \sim \mathrm{InvGamma}(v/2, v\,\lambda\,/\,2)$. λ is determined from the data in such a way that there is a q= 90% a priori chance (by default) for the BART model to improve the Root Mean Square Error (RMSE) from the ordinary minimum square regression. Thus, from the least square regression, the majority of the prior probability mass lies below RMSE. Furthermore, this prior limits the probability mass imposed on low values of σ^2 to avoid overfitting. Therefore, the higher the value of q, the greater the value of the sampled $\sigma^2{}'$s, resulting in higher model regularization (Kapelner & Bleich, 2016).

Based on the authors Kapelner & Bleich (2016) it is also important to know that adjustable values are hyperparameters α, k, β, v, q. Additionally, the number of trees – m has to be chosen. In general, default values provide a good performance, however optimal tuning could be realized by cross-validations. In our practical case we applied process mentioned to our gathered retail data. We started from 20 and 3 trees chosen for each parameter as starting points, by adding in each step 1 more every time. The process for the best combination was repeating till stop values of trees was selected.

To sum up the part of priors and likelihood, BART determines the probability of responses in the terminal nodes, along with a set of priors. They are considered to be a priori normal with the mean being the "best guess" in the leaf at the present moment (in the current MCMC iteration) and the variance being the "best guess" of the variance at the specific moment.

MCMC algorithm. According to the authors George et al., (2019) given the observed data y, the BART model induces a posterior distribution

$$p\big((T_1, M_1), \ldots, (T_m, M_m), \sigma|y\big)\ , \tag{3}$$

on all the unknowns that determine a sum-of-trees model (1 and 2). Though the parameter space's sheer size precludes exhaustive calculation, the following backtting MCMC algorithm can be used to sample from this posterior. In our practical model of retail demand prediction for backtting process mentioned, parameter optimization loop was applied. The idea of the process was to extract the optimal features of the model that further could be used as important results supporting decision making.

The algorithm is a Gibbs sampler, at a general level. For notational convenience, let $T_{(j)}$ be the combination of all trees in the sum exceptT_j, and similarly with the definition of $M_{(j)}$. Therefore, $T_{(j)}$ will be a combination of $m - 1$ trees, and $M_{(j)}$ the parameters associated with terminal nodes. The Gibbs sampler here entails m successive draws of $\left(T_j, M_j\right)$ conditionally on $\left(T_{(j)}, M_{(j)}, \sigma\right)$(Kapelner & Bleich, 2016):

$$\left(T_j, M_j\right)|T_{(j)}, M_{(j)}, \sigma, y\ , \tag{4}$$

j = 1,...,m, followed by a draw of σ from the full conditional:

$$\sigma|T_1, \ldots T_m, M_1, \ldots, M_m, y\ , \tag{5}$$

Every (T, M) in formula (4) is done by letting$p(T, M|\omega) = p(T|\omega)p(M|T, \omega)$, where ω indicates all the other conditioning information. With the prior options, μ can be incorporate analytically to obtain a computationally convenient expression for $p(T|\omega)$ Metropolis-Hastings with Gibbs steps are than use for the changes in T to propose. Although in the literature many valuable steps are provided, the key steps here are "birth-death" pair

(Chipman, George, & McCulloch, 2012; George et al., 2019). A birth step suggests adding decision rule to a current tree's bottom node so that it spawns left and right child bottom nodes. While the death step suggests the elimination of a left/ right pair of the bottom nodes. This main pair of moves by birth/ death allows the MCMC to explore trees of varying complexity and size (George et al., 2019).

Summing up, only the key attributes of BART complex approach were reviewed for understanding of mathematical path of algorithm. Further application of BART algorithms moving to posterior distribution with a help of a Metropolis-within-Gibbs sampler which employed the "Bayesian backfitting". After all it moves to the prediction, classification and further application of algorithm in different situations based on the circumstances existing. In our practical part we will use BART algorithm to capture relationships between dependent (Y) and independent variables (X), with the goal of using it in the demand prediction of retail business. Data processing for the practical part of this study is conducted by using KNIME.

3. DEMAND PREDICTION AND REPLENISHMENT IN RETAIL NETWORK

3.1. Analysis of retail sector competition in Lithuania

Before starting the practical part of inventory optimization in retail network the situation in Lithuania grocery retail sector is overviewed in order to get the understanding of market distribution and specularities those may help in further modelling of demand prediction for specific retailer of this sector. Grocery retail market in Lithuania has grown significantly during the past few years. Based on statistics it has constant increase by 5-6% annually (Fig. 4).

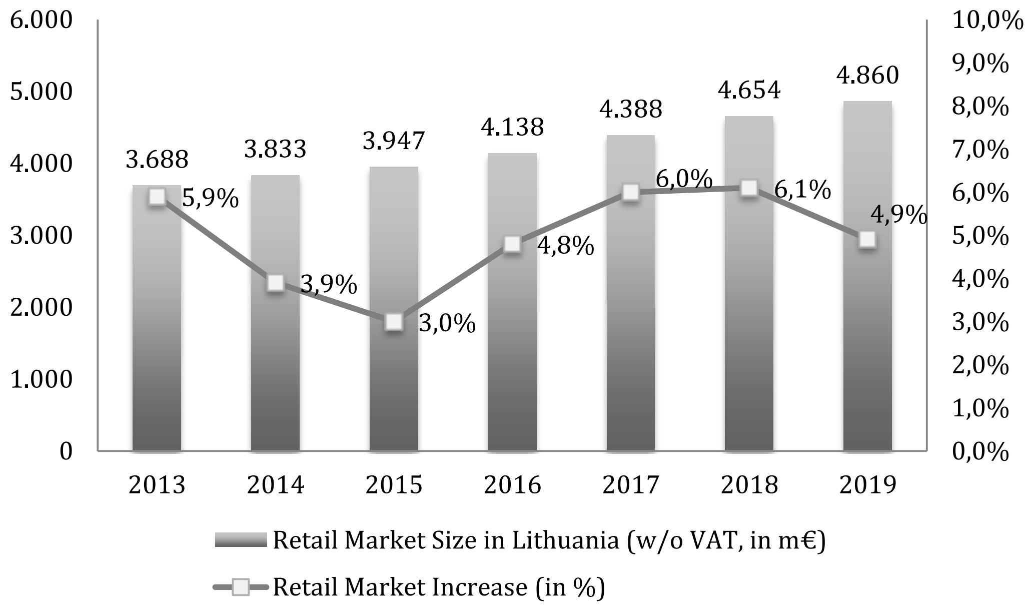

Fig. 4. Grocery retail market changes in Lithuania 2013 – 2019

Source: prepared by author based on Official Statistics Portal of Lithuania

In 2015 grocery retail faced slower increase of the sector however was still able to sustain the growth of 3%. In the following year higher consumption and purchasing power was impacted by improving economic situation in the country including decreasing unemployment, rising wages and investments (Lietuvos bankas, 2016). Today business solutions are powered by information technologies such as big data, the Internet of Things and Artificial Intelligence, which are concentrated on consumers shopping history and experience (Yuyu Yin, Zhang, Gao, & Xi, 2019). Modern technologies help better identified customers needs therefore different sales programs to attract the customer are offered globally or even individually. Lithuania together with other Baltic countries have well developed network of supermarket those may be divided into three categories (Investment & Survey, n.d.):

Hypermarkets and Big Supermarkets, those are generating the largest quantities of sales. In Lithuania examples of such grocery retail centers could be Akropolis or Mega, where an extra size space grocery retailers work in one premises in collaboration with other trending stores or selling areas.

Supermarkets, usually of medium size, operating both either on their own and accompanied by other smaller diversified non-food and food shops.

Small stores, stand-alone regional or local shops, those belonging to retail chain with a limited assortment of basic consumption products.

In Lithuania there are five main grocery retailers sharing more than 80 percent of the market (Table 7).

Table 7. Major retail chain companies in Lithuania

Retail Chain	Information		Ownership	Turnover (netto, M€) 2019	2018	2017	2016	Revenue Market Share 2019 (%)
MAXIMA	Year of Opening Number Of Stores Number of Employees	1995 247 14.334	Maxima Grupe, UAB	1.670	1.600	1.507	1.503	34,4%
PALINK	Year of Opening Number Of Stores Number of Employees	1992 228 6.540	Coopernic	678	649	643	632	14,0%
NORFA	Year of Opening Number Of Stores Number of Employees	1997 143 3.265	Norfos grupe (LT)	495	458	429	399	10,2%
LIDL	Year of Opening Number Of Stores Number of Employees	2016 48 1.764	Schwarz Group	469	368	298	165	10,1%
RIMI	Year of Opening Number Of Stores Number of Employees	1996 61 3.147	ICA Gruppen	337	325	313	314	9,7%

Source: prepared by author

The competition among these food supermarket chain members is very high. Each chain has a very dense point of sale network that provides freedom of choice for the customers. The leading retailer in Lithuania Maxima UAB has more than 30% of the market. This is mostly impact by high number of big stores around the different cities in Lithuania. All the retailers, except LIDL, are not new in the market, however LIDL as the newcomer already took a significant part of the market amounting to 10,1%. The majority of overviewed retail chains are supplied by wholesalers. As they are partners with major labels and trademarks, the strongest ones have a reasonably significant bargaining capacity. Some of most important wholesalers in Lithuania working with the well known brands of Procter & Gamble, Neste or Kinder are – Sanitex, Trojina, Bennet Distributors, Mineraliniai Vandenys and Eugesta. However, the customers' habits are changing. According to the authors Popluga, Pilvere, & Nipers (2014) research performed in Baltic countries within better economic situation in the country, customer's purchasing habits have changed – the sense of positive shopping experience is getting more and more important. Customers are looking for the new experiences in shopping, therefore grocery retailers have to focus and work every day in order to keep the interest of the customers. Well prepared marketing campaigns or the new extraordinary production offers are only some of the ways possible to keep existing or attract the new customers. Strong competition in the market force to apply innovations. To understand the customers need first it is essential to have a good inventory management system integrated to provide the wide range of possibilities and continues supply for demand satisfaction. Only after ensure the fulfillment of demand with no stockouts, the proper analysis of the customers' habits could be processed. In our practical part we will create the demand prediction model for replenishment of products to forecast. Model will be formed on behalf of historical sales data of one Lithuania grocery retailer. The model trained with past data will help the retailer in decision making of products replenishment in future. We hope that the model created with an adaption in company may help to optimize the processes, reduce the cost and at the same time ensure the right quantity of inventory needed in stores.

In order to create a model applicable for the real time situation in retail company inventories management KNIME workflow are created. With the help of this workflow Bayesian method based demand prediction and replenishment model is generated and assessed. However, before starting to analyse the modelling, it is essential first to get understanding of how the KNIME modelling works in general. For this the simplified scheme is provided below.

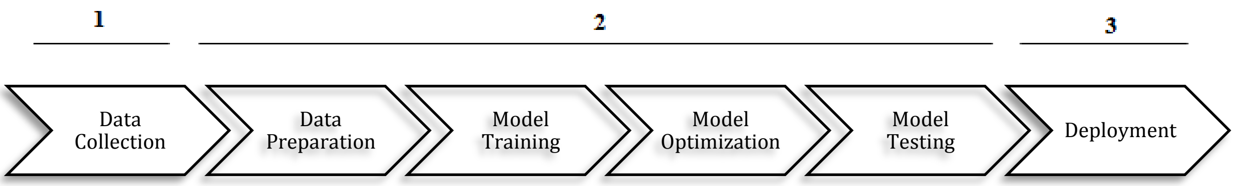

Fig. 5. The process of KNIME modelling

Source: prepared by author

From the Figure 5 it can be seen that the process of modelling has three major stages: data collection, model processing (training, optimization and testing) and model deployment. Each of particular has the unique procedures inside those are important in order to receive valuable results. Practical part of this study will also be divided into three parts regarding the system reviewed, where every procedures taken will be justified in more detailed.

3.2. Data Collection

No methods or well-known models were invented only from theory without the data. Well organized correct data is an essential basis for the modelling with significant results to receive. Moreover well organized and right selected data may work as a guidance or idea generator for the further researches to perform. For these reasons, data collection is becoming increasingly important in the design of the studies. There are different types of the data existing. Data may combine both – quantitative and qualitative types. Both of types are valuable and may be useful in modelling the procedures needed for the final results. Only the collection of quantitative and qualitative data can differs and often relies on multiple data collection methods. The quality of data is also important to mention as it was done by emphasizing that for the study well organized and correct data is needed. Data quality is the key element or assessment of data's ability to serve its purpose in a given context. The quality of data is defined by factors such as accuracy, reliability, completeness, relevance and how up to date it is. Today, as data has become more closely related to corporate practices, the focus on data accuracy has gained even greater importance. In our practical part the model is created for one Lithuanian grocery retailer. As dependant variable the sales data collected from 10 different shops over the period of 2019 in split of products sold are included. In relation with sales more independent features are selected as meaningful to include into the modelling. Selected features with the range of variation are disclosed in table below:

Table 8. Parameters of data collected for modelling

Feature	Range of Meaning	Lower Band	Upper Band
Sales (dependent variable)	Kg./Unit	0	991
Price	Eur	0,49	10,74
Delivery	Kg./Unit	1	990
Write-offs	Kg./Unit	0	541
Discount	Eur	0	1.336
Balance	Kg./Unit	0	2.219

Table 8. Parameters of data collected for modelling (Continuation)

Feature	Range of Meaning	Lower Band	Upper Band
DB Balance + Delivery	Kg./Unit	1	999
Promotion	Yes/No as 1/0	0	1
Promotion on special goods/Assortment goods	0/In out/Sortiment	0	In out/Sortiment
Advertising	0/TV/Leaflet	0	TV/Leaflet
Holiday	Yes/No as 1/0	0	1
Weekend	Yes/No as 1/0	0	1
Store placement (Vilnius/Other City)	Vilnius/Other City as 1/0	0	1

Source: prepared by author

Some additional important features are calculated and added in the specific phrase of modelling. An impact of different features will be clarified in model optimization section. Not all of the features have an equal impact to sales and in the further procedures the model will be optimize in the way to have the highest accuracy with only significant features remaining. Further procedures performed will lead to the goal accomplishment.

3.3. Demand Prediction: Model Training, Optimization and Testing

In this section the demand prediction model creation, optimization and further preparation for deployment of replenishment will be overview. All the steps are grouped, numbered and named in order to emphasize the most important procedures of the workflow performed (Fig. 6).

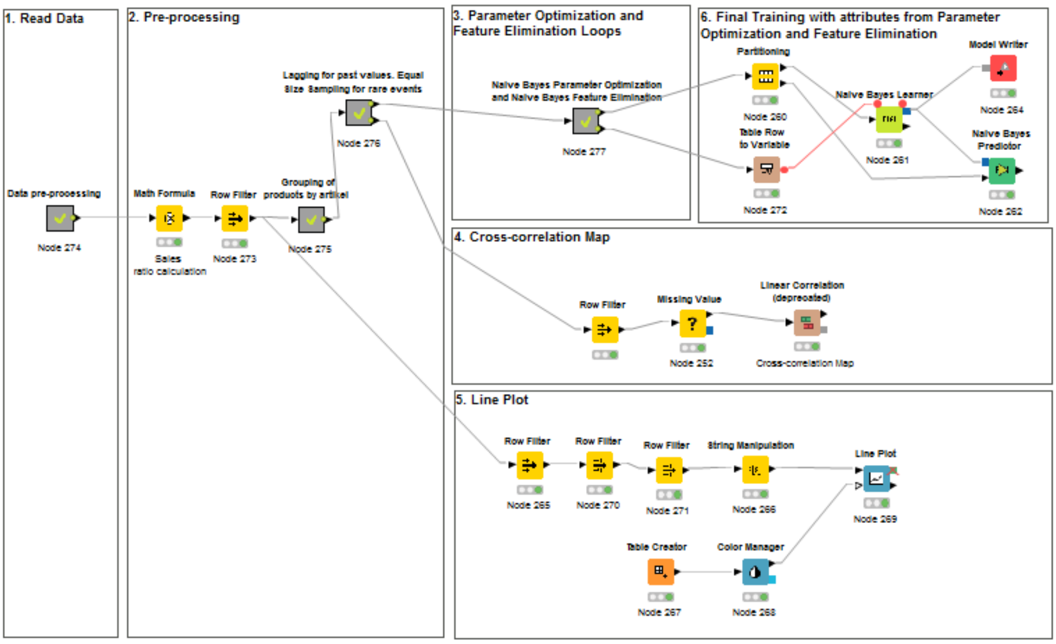

Fig. 6. Knime workflow. Stages of Model Training, Optimization and Testing

Source: prepared by author

The first group of operations in study workflow above – Read Data includes data pre-processing metanode where the procedures are constructed in the way to upload and prepare the data for further modelling.

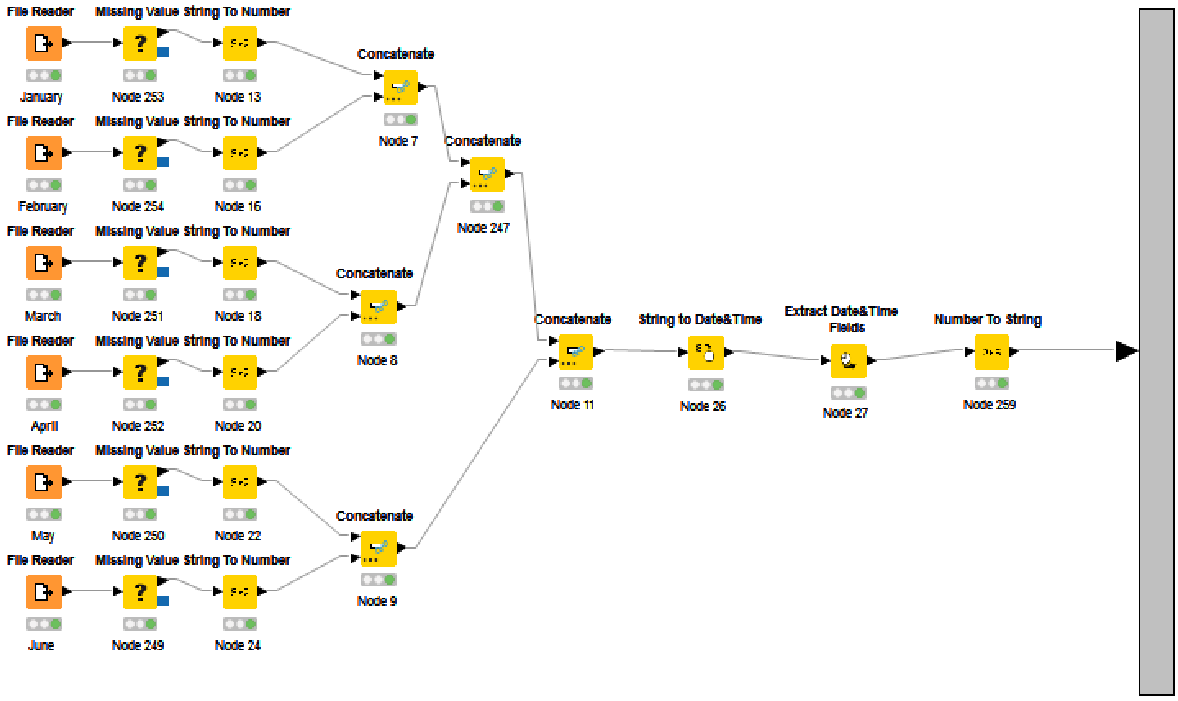

Fig. 7. Data reading and pre-processing metanode

Source: prepared by author

In the case of retail analyzing in this practical study an extracted metanode above includes six seperate csv files those are uploaded with data of sale according to the month from January to June, 2019 (Fig. 7). To upload data in seperate files was decided due to large number of rows in each month – more than 100 000 rows in each file. For data to upload File Reader node is used. File Reader node is configured in the way to read various formats of data. In data uploading step, there is important to select correct settings of data formating and type, as those settings may have an impact to further modelling of workflow. The configurations of node based on the data provided try to guess the settings of data by analyzing the content of the files. After sales data are confirmed, the Missing Value node is added close to each of the files to eliminate the missing values found in cells of input table possible due to the large amount of the data. In the first tab of the node dialog in configuration area the default handling options for all the columns of the given type are selected. In study case for three different types of data the following values for missing items are selected: most frequent value for String data columns, rounded mean for Number Integer data columns and rounded mean for Number Double data columns. Settings set extend to all the columns in the input table that are not precisely specified under the second tab in node configuration field, named „Individual Column Settings“. This tab permits individual settings available for each of the column existing in the input data. In our retail case we identified two columns as significant for futher calculations – Sales and DB Balance + Delivery, therefore set the minimum for the missing values in order to avoid value distortions in further evaluations. After deal with missing values, the logic of case requires to change the feature of price from String value to Number, in order to use it as countable element for further calculations. We did this with String to Number node that converts Strings in a column to Numbers. After execution of data seperately for the month, all the data are joint to one data output with the help of Concatenate Node. The Node concatenates two tables based on setting of rows and columns chosen. In case of our, the data is merged in a way the output table contains all the rows from the input tables. Moreover, as all the input tables have the same columns structure, the intersection of columns option that uses only the columns those appear in both input tables by ignoring any other columns is chosen. The last Concatenate node in graphic above in Figure 7 provides the final output table of sales and other relevant information for the half of the year analyzing. The following String to Date&Time and Extract Date&Time Fields nodes are applied fist, to concert date from String to Date type and second, to extract the Local Date to seperate columns of year, month, day of month and day of week in order to apply values seperately in further modelling. As final step in data pre-processing metanode Store Number value changes from Number to String, in order to express and use the type of variable value in the right way. Summarizing all the operations performed in extracted metanode, sales data was read and configurated in order to prepare it for the further modelling. Different KNIME nodes enabled the changes desired in the data, with wide range of options available. The output table of organized data for further application may be reconsider in final node (Fig. 7).

After all the data are uploaded into the workflow – the data are prepared for the further use in models creation and training (Fig. 6). In second pre-proccessing part sales ratio as an additional measure is added with the help of Math Formula node. Math Formula node evaluates a mathematical expression based on the values in rows and computes the results either as new column or as the replacement of an input column existing. Sale ratio calculated by dividing quantity sold by DB Balance + Delivery, that provides the real balance of the products in the store available (in every store, for every item, in every time frame - day). We did this because retail companies don't use demand prediction itself but as part of decision support to replenish goods. Therefore the ratio can better identify the need for replenishment than just the quantity of items sold and quantity of items currently available. Actually, as it was already mentioned the task of the retail companies is to forecast the need for replenishment. So in our particular case analyzing, we also need to forecast whether an action is needed: "No action" or "Add Products". In order to do this one more additional column, named “Flag” is added to the data table. The logic against the value assignment for each row:

- If value of Sales ratio >= 0,5, than the flag assigns “Add products”
- If value of Sales ratio < 0,5, than the flag assigns “No action”.

With two additional columns assign as variables in the further modelling we add Missing Values and Row Filter nodes to ensure the accuracy of the data with no unclear or missing values existing that would arise as a problem in further modelling. Moreover, as data source consists of many products expressed by the article numbers, the need to group the products for higher coordination to ensure was identified. For the grouping the “Grouping of products by article” metanode in workflow is created. Metanode consists of Group Loop Start, Math Formula and Loop End nodes, those ensure the right procedures of grouping. Group Loop Start node sorts the input data based on selected column in configuration dialog where each iteration process another group of rows. Math Formula node works as a body assigning selected values for each of iteration. In our case the new column appeared with the Mean of Sales Ratio of each iteration. Loop End node is used to mark the end of a workflow loop and to collect the intermediate results by row-wise concatenation of the incoming table. In our study articles were grouped in 639 iterations, each of which has assigned mean value of sales ratio. Thus means that there are 639 different product groups those sales are analysed and those demand with replenishment in different stores need to be forecasted.

In addition to pre-proccessing of data we also create a worklow for sales analysis in different periods selected with the section named – Line Plot (Fig. 6). With the help of Row Filter nodes the specific store and time period are extracted. We decided to review how the sales correlate with quantity of products available in store in 02.01.2019. Time frame is changed by changing the configurations in Row Filter node dialog. The diagram is created with the help of Line Plot node, which provides view using the JavaScript based charting library.

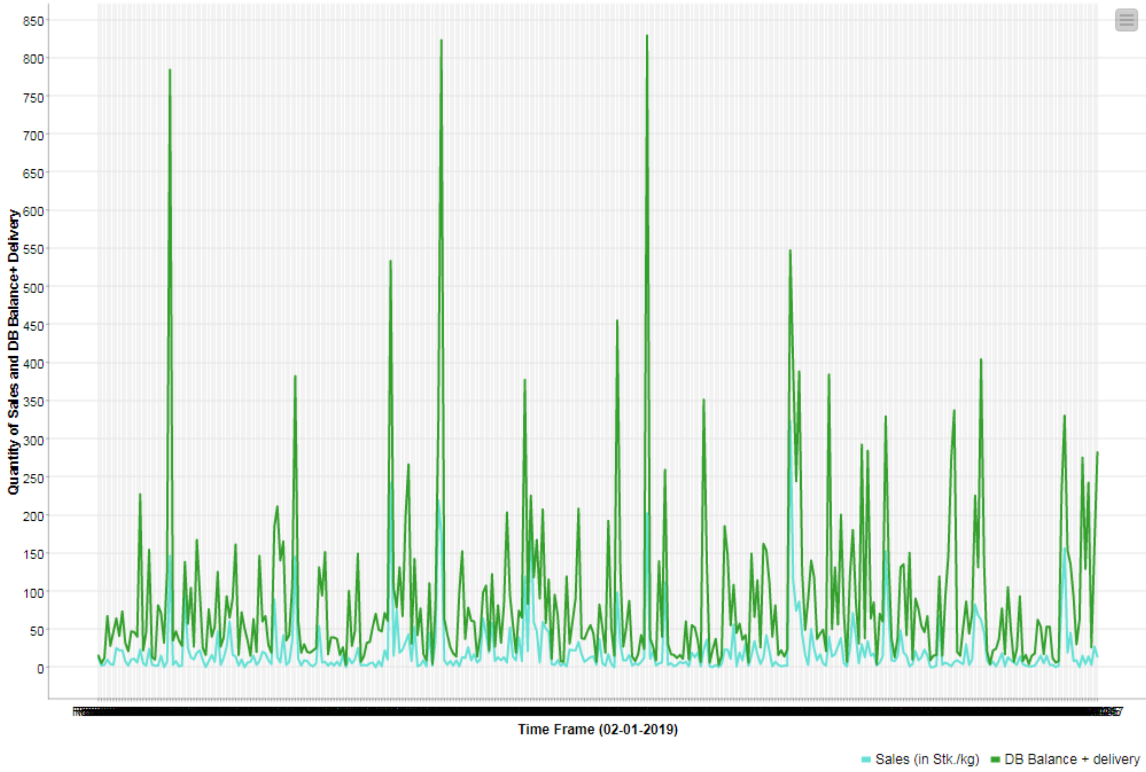

Fig. 8. Graphic of sales in relation with product amount available

Source: prepared by author

The diagram created already give us some information of company sales and strategy of inventories management. The goal of the company is to ensure the right quantity of inventory available in stores. This can be seen from the graphic above (Fig. 8) where correlation between sales and stock amount is visible. However to ensure an optimal amount of inventory without the system created in a long time perspective may be difficult to perform. There are some products, those sales during period selected reached the high peak. The stock of such products are almost or fully sold. Moreover some other lower demand products also reached the sales higher than half of the stock amount available. To ensure the demand to be satisfied the right decisions of replenishment need to be taken. In our model creating the flag with the note of „Add products“ will appear for the products with the high demand (>= 0.5 stock amount sold). For all of those products an inspection of management may be suggested with the decisions of right amount to order taken. This could improve the management of inventories as well as to save some cost occuring.

After data preparation moving forward to model trainings for time series prediction, it requires the prediction of a value at time t, x(t), given its past values, x(t-1), x(t-2), …, x(t-n). In our case 10 past Sales ratios and 1 past Flag value are added into input vector to train the model for time series prediction (Fig. 9).

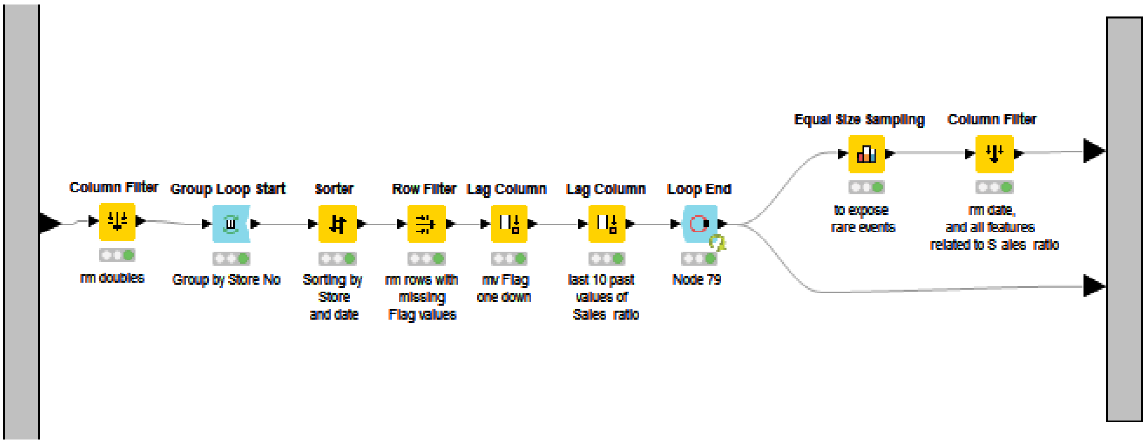

Fig. 9. Past values building metanode

Source: prepared by author

In Knime time past values creation is performed with the help of Lag Column node (Fig. 9). The Lag Column node has been introduced exactly to perform the first step, to move n past values of a time series to the same row with the current value. We just have to specify the Lag (n=10 in case of Sales Ratio and n=1 in case of Flag) in the node configuration window, to move from row x(t) to past rows down. The Lag Column node works as a body of Group Loop Start and Loop End nodes that after execution provides the model with new columns sorted by set configurations in Sorter node. To ensure the equal distribution for further training of model Equal Size Sampling node to newly generated Flag (-1) column is applied. Based on the logic of Equal Size Sampling node random rows belonging to the majority classes are removed. The rows returned by the node contain all records from the minority classes and a random sample from each of the majority classes, whereby each sample contains as many objects as the minority class contains.

After the procedures of past values performed the model evaluated the correlation between the ratios and provided it in the user-friendly matrix (Fig. 10).

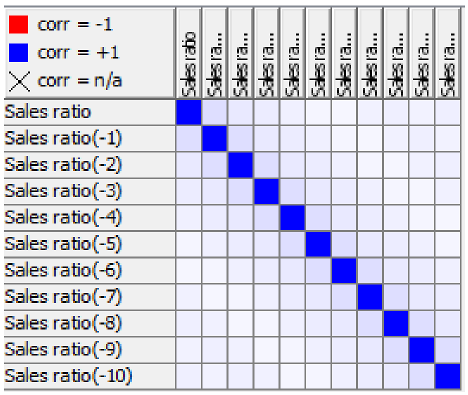

Fig. 10. Correlation matrix of sales ratios in training model

Source: prepared by author

The matrix above discloses the positive week correlation between the different sales ratios. For the purpose to improve the quality and to find out which columns and parameters are necessary for the model to be predictive the feature elimination and parameter optimization loops are added (Fig. 6 and Fig. 11).

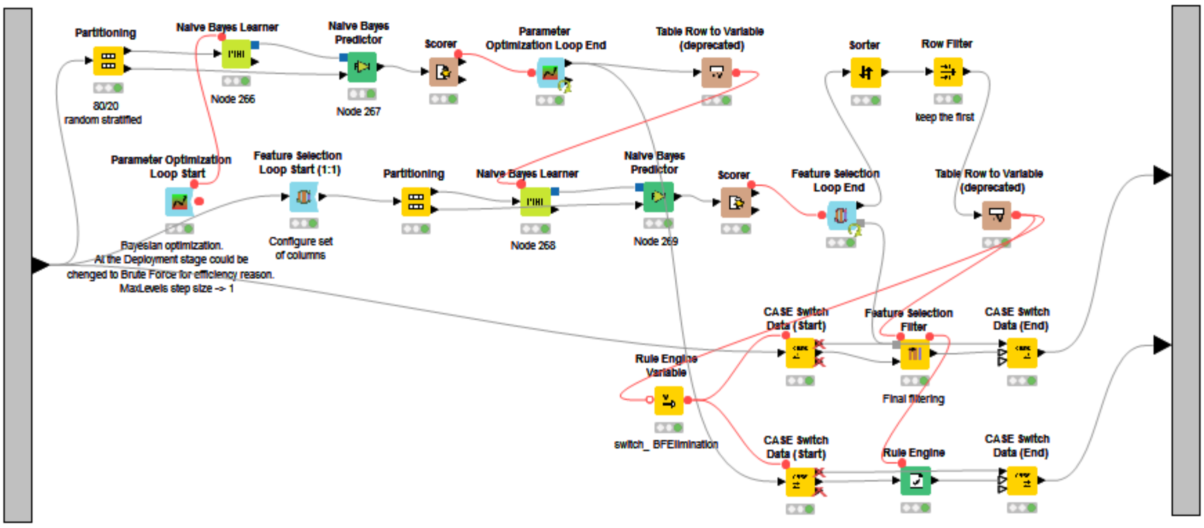

Fig. 11. Bayesian Parameter Optimization and Backforward Feature Elimination

Source: prepared by author

The goal of Parameter Optimization Loop is to find the best set of parameters for the prediction model. Like all the other loops, the Parameter Optimization Loop follows the classic KNIME loop mode. For parameter optimization we used Parameter Optimization Loop Start and Parameter Optimization Loop End nodes those surrounded by loop body train the prediction model. After mathematical methods review in theoretical part, method of Bayesian Additive Regression Trees was decided to be applied in our study as the method for Learner node to train the classification model and Predictor node to apply the model to the test dataset. The data is split into the training and test set by using the Partitioning node in proportion of 80/20. The Parameter Optimization Loop Start node loops through the list of parameter sets. In our study the parameters set for optimization – minimal standard deviation value (MinSDValue), that specifies a minimal spread of the data of observation, and threshold, that depends on the probability of the class through the process of classification. For each of the parameters selected to optimize the range of start and stop values are defined with exactly step size set. The strategy of our selected as follow:

Table 9. Process configuration for parameters optimization

Parameter	Start value	Stop value	Step size
MinSDValue	0,0001	0,01	0,0001
Threshold	0,1	0,5	0,1

Source: Prepared by author

The strategy of search applied for Parameter Optimization Loop Start Node – Bayesian Optimization (TPE). The strategy selected for model parameters optimization has two phases: the first phase of warm-up in which parameters combinations are randomly chosen and evaluated. The number of warm-ups may be manually set and our selection is 50 rounds. After these rounds, the actual Bayesian optimization begins in which parameter combinations are picked by analyzing the past scores and guessing new promising parameter combinations which are then evaluated. The algorithm used by Bayesian Optimization (TPE) strategy is based on an algorithm published by Bergstra and uses Tree-structured Parzen Estimation (TPE) in the second phase in order to select suitable parameter combinations. The Parameter Optimization Loop Start configuration window also requires additional configurations to set, in order to organize the further optimization procedures:

– Maximal number of iterations. Selected number – 50.

– Gamma. The value dividing already evaluated by TPE parameter combination into bad or good. In most of the studies gamma is chosen to be in range from 0,15 to 0,3. Our selection model training and testing phrase is 0,2, which means that 20% of best parameter combinations will rely to the good distribution and the rest to the bad one.

– Number of candidates per round – defines how many randomly candidates can be drown from parameter space by TPE when trying to find the next parameter combination by maximizing the expected improvements. Our number of candidates per round set – 100.

According to configurations set in each iteration a new set of parameters is fit to the learner node. The Parameter Optimization Loop End node collects iteration result and compares it with the previous ones until receive the best set of parameters. The goal on optimizing is to obtain the set of parameters for the Naive Bayes Learner algorithm that leads to the highest accuracy on the test set. The performance of model is measured by Scorer node, which compares values of actual with predicted columns and provides a number of classification accuracy metrics.

Table 10. Bayesian Additive Regression Trees method accuracy results

		Predicted data	
Real data		Bayesian Additive Regression Trees	
		No action	Add Products
	No action	16.102	477
	Add products	1.784	14.092
	Accuracy	97,19 %	

Source: Prepared by author

The results showed that the performance of parameter optimization model created is by 97,19% accurate. The error is 2,81% therefore it may be used in further modelling of replenishment. Best set of parameters combination calculated: MinSDValue (0,001) and Threshold (0,108).

An optimal set of parameters as flow variables are used in further feature elimination procedures of modelling. Feature elimination loop works in the way that lets to choose the column we want to include in the output table. The workflow shows all computed levels of the feature elimination together with the error rates. It is possible either select one level manually or specify an error threshold and then the level with the fewest features that has a prediction error below the threshold is automatically selected. In any case all columns from the input table that are not present in the selected level are filtered from the input table. In this way only the necessary columns remain in the model what ensure the more accurate results of the models (Fig. 11). The strategy of Feature Selection may be chosen from four strategies optional: Forward Feature Elimination, Backward Feature Elimination, Genetic Algorithm and Random Strategies. The strategies differ in the process of feature selection organization. Strategies of Forward and Backward Feature Elimination are iterative approaches, while Genetic Algorithm is a stochastic approach that bases optimization on the mechanics of biological evolution and genetics. The remaining random strategy is a simple approach, which selects feature combinations randomly. Our selection for practical study is Backward Feature Elimination strategy. The strategy of features selection starts with having all the features selected. The feature that has the least effect on the performance of the model upon its elimination is omitted with each iteration. Moreover in the Feature Selection Loop start the set of static columns may be configurated, those through the elimination procedures will not be removed from the model. In our case the columns selected as static – Store Number, Article Number and Flag (-1). In the further processing of feature elimination loop follows the classic KNIME loop mode. The Feature Selection Loop Start and Feature Selection Loop End work surrounded by loop body of Naive Bayes Learner for classification model to train and Naive Bayes Predictor to apply the model to the test dataset. The data is split into the training and test set using the same logic as with Parameter Optimization in proportion of 80/20. The process of feature elimination is finished with Rule Engine Variable and Case Switch Data nodes those are comparing whether the model after feature elimination stage is more accurate than it was after the parameter optimization. If results of feature elimination are more accurate then it was before this stage, we adapt them and pass improved accuracy indicator together with data needed to create more accurate model. The rule combined of defined flow variables, creates a new flow variable named by us as ”switch_BFElimination” with the resulted value of 1, showing that the model accuracy after feature elimination (97,66%) is higher than model accuracy only after the stage of parameter optimization (97,19%). This flow variable is connecting with two CASE Switch Data Start nodes. The first of the outcome is connecting with the Feature Selection Filter to obtain the results of the elimination (Fig. 11).

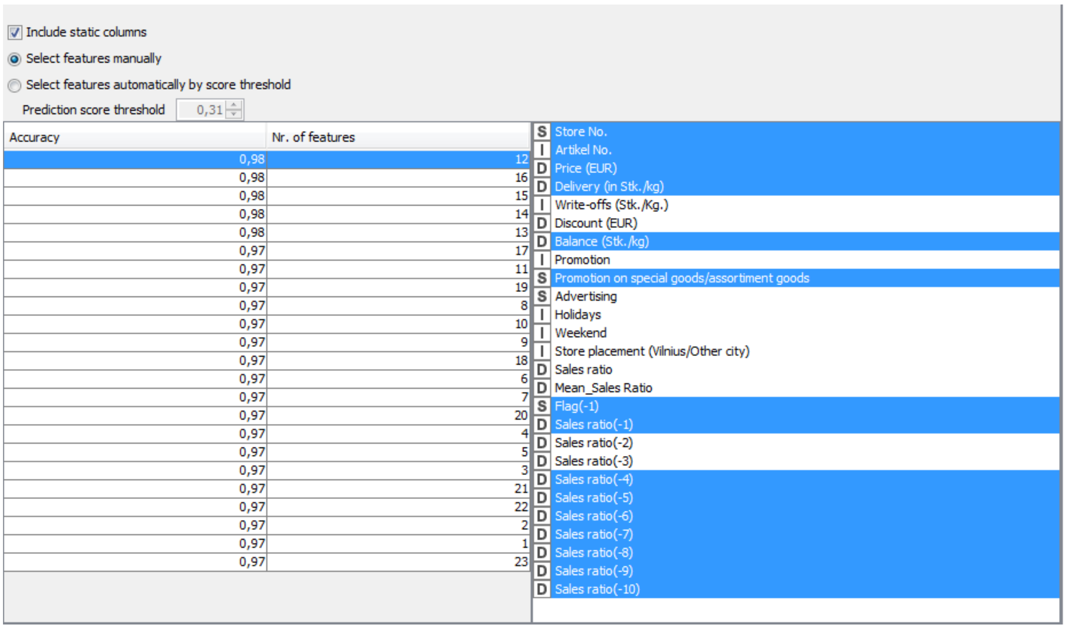

Fig. 12. Feature Selection Filter

Source: prepared by author

Configuration window of the Feature Selection Filter node following the Backward Elimination Loop is provided above with the results of the process provided (Fig. 12). A close to best accuracy value (97,66%) is obtained with 12 input features of Price (EUR), Delivery (Stk./kg.), Balance (Stk./kg.), Promotion on special goods/ assortment goods, and 8 of 10 sales ratios. The workflow removes the rest input features, some joined from an external file and some generated through the different stages of workflow. From the second CASE Switch Data Start node the Rule Engine node for the parameter optimization is created. The rule created compares the accuracies by setting the highest (97,66% > 97,19%). After this step with the help of CASE Switch Data End nodes the final outputs with the result of optimal parameters and selected features for modelling is presented. On the remaining input features a Naive Bayes is trained to predict the need for adding products for the future periods based on demand forecasted. Model Writer node is used to write model port object to a file, which may be read further in model deployment stage (part 6 of Fig. 6).

3.4. Replenishment Model Deployment

The definition of deployment in data science relates to the application of the model trained for prediction on an existing production environment to make practical business decisions based on new data. Building and training the model is generally not the end of the modelling. Even if the goal of the study would be the increase of the knowladge of the data, the model created would not be applicable to use on a real time situations in predicting the actions needed. We already created a machine learning model that can determinate if the specific product on a particular day in store needs a replenishment or no. However by training the data we used the historical data of sales per stores, while the goal of our is to determinate the need of replenishment in real-time, so that the business may prevent from the stockouts possible. In practice there are three ways to deploy the machine learning model based on the the situation having:

1. One-off – applicable for the case when there is no need to constantly learn machine learning model in order for it to be deployed. For such situations the model is only needed once or periodically. The model then, when it is neccessary, is simple trained ad-hoc and pushed to output until it deteriorates sufficiently to require any fixing.
2. Batch – the training that provides a posibility to have an up-to-date version of the model. It is a scalable method that takes a data subsample at a time, by removing the need for each upgrade to use the entire data set. This practise is good when the machine learning model is used on a consistent basis, but do not neccessary require the prediction in real-time.
3. Real-time – the practise when the accurate prediction is needed constantly in actual time. This is possible by using online machine learnings models, like Naive Bayes as in our practical case.

Data Scientists, regardless of what package they choose, are used to train machine learning models to solve business issues. To get the most value out of machine learning model created, it is essential to deploy it into everyday production so a business can start using it to make practical decisions. The goal of our work is to deploy a model on a real-time data thus the third scenario reviewed in order to build a beneficial tool for the business to apply and practically use in every day situations is needed. So, what we need here is two workflows –

the training one, and deployment. As we already have the first one prepared for the further use, deployment mode may be constructed with several important assumptions having. A model of deployment required the data prepared and integrated to it for generation in exactly the same way as it was created in the first workflow. In KNIME it can be proceeded either automatically by using the KNIME integrated deployment extensions or mannually by creating the new workflow from the stratch. In our study we select the second, manually option in order to better understand the model deploying, to do some changes requiring and to prevent from the mistakes possible in the process. The following Figure 13 shows the deployment forkflow modelled to alarm of replenishment for products needed.

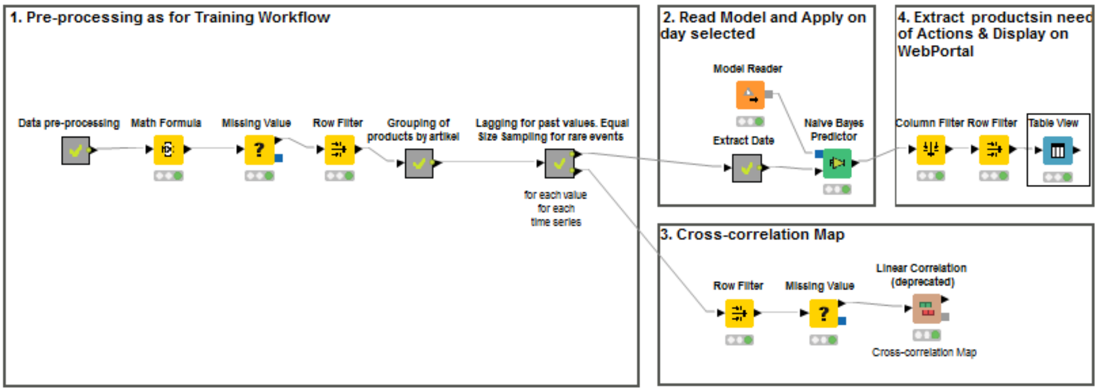

Fig. 13. Deployment model

Source: prepared by author

The modelling workflow accesses the data from Excel files. Six different Excel files with sales data of 2019.07-2019.12 are uploaded in data pre-processing metanode. In our particular case for exact day prediction we only need information of ten previous days. However we decided to add more information for wider forecasting day’s capabilities. The input data consists of all the exact columns as in the workflow of training, except for Flag column which has to be forecasted based on the model trained. The data is prepared by recomputing the type of each column, dealing with missing values and converting a few columns from string to numerical. Than the data are combined into one table by using the corresponding node of concatenate. As date in data composition is expressed as string value, for the further use column of date is converted to Date and Time. The separate columns for year, month and day are created for exact date needed to filter in further model deployment. As additional features with the help of Math Formula node sales ratio for each of the row is calculated. The feature performs a better view of sales of each row, as it is expressed as proportion of two important measures – sales and balance (prior day balance + exact day delivery). Moreover, two new features are added into modelling by grouping the articles into the groups. The first feature here is the iteration of each of the group while the second is the mean of sales ratio expressing each of iteration. After the grouping stage there are 744 different iterations encoding 744 different product groups with the meaning values of sales ratio from 0 to 1. For the new sales data past values to build the lagging for a past values and equal size sampling metanode with appropriate nodes inside was applied. For the new data the 10 past values of Sales Ratio and 1 past value for Flag are created. Comparing with the initial data training model, one change is applied in order to use it for the further modelling. Unlike in the training model, the date with the separate columns of year, month and day are not excluded from the data set in column filter node. The columns are left for the further application in filtering. For the further model deployment the exact date metanote were added (Fig. 14).

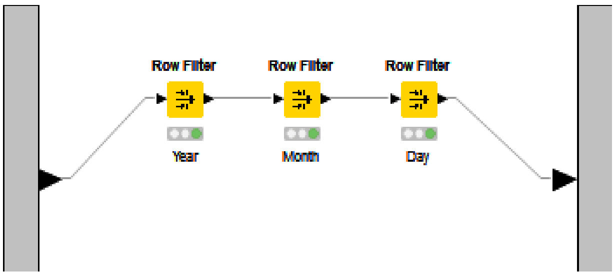

Fig. 14. Exact date configuration for deployment stage

Source: prepared by author

In the metanode above cretaed with the help of Row Filter nodes the exact year, month and day are set for the deployment results to filter. The configuration windows of nodes may be changed every time based on the exact situation. For our particular case we set the exact day (2019.07.11) to see what products in which stores require replenishment at the end of the day to satisfy the demand. We connected the outcome of this metanode to Naive Bayes Predictor node. This node has one more input – Model Reader, which reads KNIME model port objects that were written with the Model Writer node. Here we added out priviously trained model from the first training workflow. The Naive Bayes Predictor node predicts the class per row based on the learned model. By adding the trained model we want to deploy the priviously trained model to predict the need for replenishment for a new data. For the final results to disclose, only the rows with the pattern maching „Add products“ are filtered. Final results of the replenishment presented in user-friendly environment of Table View (Fig. 15).

Show 10 entries Search:

		Date	Store No.	Artikel No.	Price (EUR)	Balance (Stk./kg)	**Name**
		2019-07-11	1	12564	0.49	33	Yogurt
		2019-07-11	1	19397	2.47	25	Cheese
		2019-07-11	1	21274	0.21	32	Desrt
		2019-07-11	1	29838	0.4	12	Yogurt
		2019-07-11	1	116152	1.78	155	Sausages
		2019-07-11	1	116536	3.63	6	Salmon
		2019-07-11	1	5500116	1.81	65	Cheese
		2019-07-11	1	5500144	2.22	43	Butter
		2019-07-11	1	5500478	0.4	53	Curd
		2019-07-11	1	5500480	0.4	32	Curd

Showing 1 to 10 of 399 entries Previous 1 2 3 4 5 ... 40 Next

Fig. 15. Model results. Products with replenishment needed

Source: prepared by author

Results of the model provide the management with the information that 399 artickles in 10 different stores require replenishment for the demand of upcoming day to fullfilled. In the model created the days can be changed in order to get the results of different days. As in most of the retail companies orders for the upcoming periods are still entered manually by responsible employee, model created may assist as an accurate tool for products with action needed highlighted. With an application of such inventory optimization model company may inform the management of the inventory needed more precisely what can lead to better coordination of processes and reduction in costs.

The novelty of the work is also visible as we implement the model to solve the problem identified in the begining of work - no tool available in retail network for stock replenishment that may work based on up to date mathematical methods in demand prediction and existing algorithms of economic order quantity. Although the model created examines only one part of the problem - mathematical methods application in demand prediction by forecasting the need for replenishment, the further plan of is to extend results received by adding the economic order quantity algorithm analyzed in the theoretical part, which may assist by telling the optimal quantity of goods requiring to add. As any of the studies analyzed combine both of the methods on a real time data in retail market, results received may work as an effective innovation for the companies those want to improve the inventory management processes inside the organization.

CONCLUSIONS

1. Three different inventory systems often used in retail market were compared and practically analysed. The main difference between the systems identified is the principal of reordering and responsibility. If in Fixed Time Period Inventory system the key principal of reordering is constant time periods between orders, in Fixed Order Inventory system the constant size is quantity of order. However in both of methods responsibility to make an order is in the retailer side, while in Vendor Managed Inventory System, the vendor is decision maker, who set the order size and other related parameters for both.

2. Four often used IT applications for inventory analysis – Microsoft Power BI, IBM SPSS Modeler, KNIME and Apache Hadoop were proposed and compared. Each of the platform is configurated to help the users to construct accurate predictive models for individuals, groups and the systems of enterprise. The last two platforms works under the open-source principle where the original source code is made freely available and could be redistributed and implemented. There is no leading platform to use. Each application has some strong sides while at the same time areas for improvement indentified by their users: limited visualizations, crowded interfaces, issues with files size and security risks.

3. In the study Linear regression, Logistic regression, Probabilistic Neural Network, Bayesian Additive Regression Trees, Random Forest and Fuzzy Logic were suggested as the methods those could be used in modelling demand prediction of retail network. Practical application, advantages and disadvantages of each were evaluated and compared. Results of comparison showed all methods analysed have significant areas to improve. Recurring problems identified: possible deficiencies, high memory requirements, long testing time and possibly large computational cost.

4. Retail market of Lithuania in the last few years was growing from 3% to 6% annually. In Lithuania there are five main grocery retailers sharing more than 80% of the market. The leading retailer in Lithuania Maxima UAB has more than 30% of the market. Researches show that within the better economic situation customer's purchasing habits have changed – the sense of positive shopping experience is getting more and more important. Therefore for the retailers first, it is essential to ensure the right balance of the products in store and second, to put attention on marketing campaigns or the new extraordinary production offers to maintain the satisfaction of the customer. All these as affecting input features were included into the modelling of demand prediction.

5. Real time sales data of the period from 2019.01 to 2019.06 was collected and processed in order to use data for further modelling of demand prediction. Procedures performed to prepare and systemize the data: changes of data formatting, missing values cleaning, extraction of date & time fields, additional features creation, data grouping and lagging for the past values needed for time series prediction. The procedures performed with the help of nodes in KNIME analytic platform ensure both, the errors avoidance in the further modelling and higher accuracy as systemized and sorted data provides more precise results.

6. For model optimization parameters optimization and feature elimination procedures were applied. It was selected to optimize two parameters of Minimal Standard Deviation value and Threshold. The strategy of search applied for Parameter Optimization – Bayesian Optimization (TPE). The optimal values of parameters obtained with minimal error of 2,81% – 0,001 and 0,108. For Feature Elimination procedure the strategy of Backward Feature Elimination was applied. Highest accuracy of 97,66% reached with set of 12 input features of Price (EUR), Delivery (Stk./kg.), Balance (Stk./kg.), Promotion on special goods/ assortiment goods and 8 of 10 sales ratios. Defined set of parameters with the smallest error and selected features with the highest accuracy were used further for model training to predict the demand.

7. After the steps of model training, optimization and testing with the real time sales data from 2019.01 to 2019.06 demand prediction model trained to predict the replenishment needed was created. On deployment stage model was applied to sales data from 01.07.2019 to 10.07.2019 with the missing column of "Flag" which encode the replenishment actions of "Add products" or "No action" needed. The action needed was predicted by demand prediction model trained. In the case analysed, the exactly date of 11.07.2019 to predict the replenishment for products was selected. Results table shows that 399 products across 10 different stores require replenishment. Model created may assist as an accurate tool for company inventory optimization leading to better processes cooorditation, possible reduction in costs and increase in profit.

8. Author with the help of this practical case also underline that accuracy of the different methods applied in modelling can be evaluated and compared. As every case in the practice is different there is no best method perfectly suitable for all the situations. Only evaluation of model received can show which method provides the best quality on every specific study performing. Moreover, as model formed highlights only the need for replenishment without evaluating the optimal amount for particular goods required, for further studies, model improvement in combination with EOQ algorithm is introduced.

REFERENCES

Abusager, K., Baldwin, M., & Hsu, V. (2020). Using Power BI to Inform Clostridioides difficile Ordering Practices at an Acute Care Hospital in Central Florida. *American Journal of Infection Control*, *48*(8), S57–S58. https://doi.org/10.1016/j.ajic.2020.06.036

Achabal, D. D., McIntyre, S. H., Smith, S. A., & Kalyanam, K. (2000). A decision support system for vendor managed inventory. *Journal of Retailing*, *76*(4), 430–454. https://doi.org/10.1016/S0022-4359(00)00037-3

Agápito, A. D. O., Vianna, M. D. F. D., Moratori, P. B., Vianna, D. S., Meza, E. B. M., & Matias, I. D. O. (2019). Using multicriteria analysis and fuzzy logic for project portfolio management. *Brazilian Journal of Operations & Production Management*, *16*(2), 347–357. https://doi.org/10.14488/bjopm.2019.v16.n2.a14

Aghdaei, N., Kokogiannakis, G., Daly, D., & McCarthy, T. (2017). Linear regression models for prediction of annual heating and cooling demand in representative Australian residential dwellings. *Energy Procedia*, *121*, 79–86. https://doi.org/10.1016/j.egypro.2017.07.482

Ajidarma, P., & Irianto, D. (2019). Application of Bayesian Additive Regression Trees to Analyse the Growth of United States Electric Automobile Industry. *IOP Conference Series: Materials Science and Engineering*, *598*(1). https://doi.org/10.1088/1757-899X/598/1/012035

Akbari Kaasgari, M., Imani, D. M., & Mahmoodjanloo, M. (2017). Optimizing a vendor managed inventory (VMI) supply chain for perishable products by considering discount: Two calibrated meta-heuristic algorithms. *Computers and Industrial Engineering*, *103*, 227–241. https://doi.org/10.1016/j.cie.2016.11.013

Anghelache, C. (2015). Analysis of final consumption and gross investment influence on GDP – multiple linear regression model. *Theoretical and Applied Economics*, *22*(3), 137–142.

Arcelus, F. J., Pakkala, T. P. M., & Srinivasan, G. (2003). Special sales with guaranteed minimum duration but uncertain termination date. *Applied Mathematical Modelling*, *27*(9), 677–699. https://doi.org/10.1016/S0307-904X(03)00069-6

Arcelus, F. J., Shah, N. H., & Srinivasan, G. (2001). Retailer's response to special sales: Price discount vs. trade credit. *Omega*, *29*(5), 417–428. https://doi.org/10.1016/S0305-0483(01)00035-4

Arcelus, F. J., Shah, N. H., & Srinivasan, G. (2003). Retailer's pricing, credit and inventory policies for deteriorating items in response to temporary price/credit incentives. *International Journal of Production Economics*, *81–82*, 153–162. https://doi.org/10.1016/S0925-5273(02)00269-4

Barnes-Schuster, D., Bassok, Y., & Anupindi, R. (2006). Optimizing delivery lead time/inventory placement in a two-stage production/distribution system. *European Journal of Operational Research*, *174*(3), 1664–1684. https://doi.org/10.1016/j.ejor.2002.08.002

Berthold, M. R. (2003). Mixed fuzzy rule formation. *International Journal of Approximate Reasoning*, *32*(2–3), 67–84. https://doi.org/10.1016/S0888-613X(02)00077-4

Berthold, M. R., & Diamond, J. (1998). Constructive training of probabilistic neural networks. *Neurocomputing*, *19*(1–3), 167–183. https://doi.org/10.1016/S0925-2312(97)00063-5

Boulaksil, Y. (2016). Safety stock placement in supply chains with demand forecast updates. *Operations Research Perspectives*, *3*, 27–31. https://doi.org/10.1016/j.orp.2016.07.001

Burrascano, P. (1991). Learning vector quantization for the probabilistic neural network. *IEEE Transactions on Neural Networks*, *2*(4), 458–461. https://doi.org/10.1109/72.88165

Chen, Z. (2018). Optimization of production inventory with pricing and promotion effort for a single-vendor multi-buyer system of perishable products. *International Journal of Production Economics*, *203*(July), 333–349. https://doi.org/10.1016/j.ijpe.2018.06.002

Chipman, H. A., George, E. I., & McCulloch, R. E. (2012). BART: Bayesian additive regression trees. *Annals of Applied Statistics*, *6*(1), 266–298. https://doi.org/10.1214/09-AOAS285

Cobb, B. R., Johnson, A. W., Rumí, R., & Salmerón, A. (2015). Accurate lead time demand modelling and optimal inventory policies in continuous review systems. *International Journal of Production Economics*, *163*, 124–136. https://doi.org/10.1016/j.ijpe.2015.02.017

de Maio, A., & Laganà, D. (2020). The effectiveness of Vendor Managed Inventory in the last-mile delivery: An industrial application. *Procedia Manufacturing*, *42*, 462–466. https://doi.org/10.1016/j.promfg.2020.02.047

Feng, Y., & Wang, S. (2017). A forecast for bicycle rental demand based on random forests and multiple linear regression. *Proceedings - 16th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2017*, 101–105. https://doi.org/10.1109/ICIS.2017.7959977

George, E., Laud, P., Logan, B., McCulloch, R., & Sparapani, R. (2019). *Fully Nonparametric Bayesian Additive Regression Trees*. 89–110. https://doi.org/10.1108/s0731-90532019000040b006

Ghatasheh, N. (2014). Business Analytics using Random Forest Trees for Credit Risk Prediction: A Comparison Study. *International Journal of Advanced Science and Technology*, *72*(November), 19–30. https://doi.org/10.14257/ijast.2014.72.02

Graves, S. C. (1996). A multiechelon inventory model with fixed replenishment intervals. *Management Science*, *42*(1), 1–18. https://doi.org/10.1287/mnsc.42.1.1

Graves, S. C., & Willems, S. P. (2003). Erratum: Optimizing Strategic Safety Stock Placement in Supply Chains (Manufacturing & Service Operations Management (2000) 2:1 (68-83)). *Manufacturing and Service Operations Management*, *5*(2), 176–177. https://doi.org/10.1287/msom.5.2.176.16074

Grömping, U. (2016). Practical Guide to Logistic Regression. In *Journal of Statistical Software* (Vol. 71). https://doi.org/10.18637/jss.v071.b03

Gupta, B., Rawat, A., Jain, A., Arora, A., & Dhami, N. (2017). Analysis of Various Decision Tree Algorithms for Classification in Data Mining. *International Journal of Computer Applications*, *163*(8), 15–19. https://doi.org/10.5120/ijca2017913660

Güven, İ., & Şimşir, F. (2020). Demand forecasting with color parameter in retail apparel industry using artificial neural networks (ANN) and support vector machines (SVM) methods. *Computers and Industrial Engineering*, *147*(July). https://doi.org/10.1016/j.cie.2020.106678

Han, J., Lu, J., & Zhang, G. (2017). Tri-level decision-making for decentralized vendor-managed inventory. *Information Sciences*, *421*, 85–103. https://doi.org/10.1016/j.ins.2017.08.089

Hoque, M. A. (2013). A manufacturer-buyer integrated inventory model with stochastic lead times for delivering equal- And/or unequal-sized batches of a lot. *Computers and Operations Research*, *40*(11), 2740–2751. https://doi.org/10.1016/j.cor.2013.05.008

Huang, K. (2019). *Inventory Optimization of Fresh Foods Based on Shang Mao Cheng Supermarket*. 23–27.

Ignaciuk, P., & Bartoszewicz, A. (2009). Robust sliding-mode supply policy for periodic-review inventory systems with time-varying lead-time delay. In *IFAC Proceedings Volumes (IFAC-PapersOnline)* (Vol. 14). https://doi.org/10.3182/20090819-3-pl-3002.00121

Ihlayyel, H. A. K., Sharef, N. M., Nazri, M. Z. A., & Bakar, A. A. (2018). An enhanced feature representation based on linear regression model for stock market prediction. *Intelligent Data Analysis*, *22*(1), 45–76. https://doi.org/10.3233/IDA-163316

Inderfurth, K., & Vogelgesang, S. (2013). Concepts for safety stock determination under stochastic demand and different types of random production yield. *European Journal of Operational Research*, *224*(2), 293–301. https://doi.org/10.1016/j.ejor.2012.07.040

Investment, F., & Survey, T. M. (n.d.). *RE*.

Joubert, M., Verster, T., & Raubenheimer, H. (2019). Making use of survival analysis to indirectly model loss given default. *ORiON*, *34*(2), 107–132. https://doi.org/10.5784/34-2-588

Jouve, O., Martin, E., & Guerin, M. C. (2012). Case Study. Sentiment-Based Text Analytics to Better Predict Customer Satisfaction and Net Promoter® Score Using IBM®SPSS® Modeler. In *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*. https://doi.org/10.1016/B978-0-12-386979-1.00021-9

Jung, J. Y., Blau, G., Pekny, J. F., Reklaitis, G. V., & Eversdyk, D. (2008). Integrated safety stock management for multi-stage supply chains under production capacity constraints. *Computers and Chemical Engineering*, *32*(11), 2570–2581. https://doi.org/10.1016/j.compchemeng.2008.04.003

Kapelner, A., & Bleich, J. (2016). bartMachine: Machine learning with bayesian additive regression trees. *Journal of Statistical Software*, *70*. https://doi.org/10.18637/jss.v070.i04

Korponai, J., Tóth, Á. B., & Illés, B. (2017). Effect of the Safety Stock on the Probability of Occurrence of the Stock Shortage. *Procedia Engineering*, *182*, 335–341. https://doi.org/10.1016/j.proeng.2017.03.106

Kostić, K. (2009). Inventory control as a discrete system control for the fixed-order quantity system. *Applied Mathematical Modelling*, *33*(11), 4201–4214. https://doi.org/10.1016/j.apm.2009.03.004

Kusy, M. (2015). Dimensionality reduction for probabilistic neural network in medical data classification problems. *International Journal of Electronics and Telecommunications*, *61*(3), 289–300. https://doi.org/10.1515/eletel-2015-0038

Lesnaia, E. (2004). Optimizing Safety Stock Placement in General Network Supply Chains. *Massachusetts Institute of Technology*.

Li, Y., & Ou, J. (2020). Optimal ordering policy for complementary components with partial backordering and emergency replenishment under spectral risk measure. *European Journal of Operational Research*, *284*(2), 538–549. https://doi.org/10.1016/j.ejor.2020.01.006

Lietuvos bankas. (2016). *Lietuvos ekonomikos apžvalga 2016*. 11–22.

Lithuanian Official Statistics Portal. Retrieved from https://osp.stat.gov.lt/en

Linero, A. R., & Yang, Y. (2018). Bayesian regression tree ensembles that adapt to smoothness and sparsity. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *80*(5), 1087–1110. https://doi.org/10.1111/rssb.12293

Logan, B. R., Sparapani, R., McCulloch, R. E., & Laud, P. W. (2019). Decision making and uncertainty quantification for individualized treatments using Bayesian Additive Regression Trees. *Statistical Methods in Medical Research*, *28*(4), 1079–1093. https://doi.org/10.1177/0962280217746191

Lu, M. (2014). *Discovering Microsoft Self-service BI solution : Power BI*. (May).

Maiti, A. K., Maiti, M. K., & Maiti, M. (2009). Inventory model with stochastic lead-time and price dependent demand incorporating advance payment. *Applied Mathematical Modelling*, *33*(5), 2433–2443. https://doi.org/10.1016/j.apm.2008.07.024

Malik, A. I., & Sarkar, B. (2018). Optimizing a multi-product continuous-review inventory model with uncertain demand, quality improvement, setup cost reduction, and variation control in lead time. *IEEE Access*, *6*, 36176–36187. https://doi.org/10.1109/ACCESS.2018.2849694

Microsoft data centre servers run on fuel cell backup power for 48 h. (2020). *Fuel Cells Bulletin*, *2020*(8), 7. https://doi.org/10.1016/s1464-2859(20)30342-4

Minner, S. (2007). A note on how to compute economic order quantities without derivatives by cost comparisons. *International Journal of Production Economics*, *105*(1), 293–296. https://doi.org/10.1016/j.ijpe.2006.04.012

Mishra, V. K. (2013). Deteriorating inventory model with quadratically time varying demand and partial backlogging. *WPOM - Working Papers on Operations Management*, *4*(2). https://doi.org/10.4995/wpom.v4i2.1170

Nechval, N. A., & Nechval, K. N. (1999). Applications of invariance to estimation of safety stock levels in inventory model. *Computers and Industrial Engineering*, *37*(1), 247–250. https://doi.org/10.1016/S0360-8352(99)00066-2

Nikolla, E. (n.d.). *BIT s tudents ' performance analysis with KNIME Analytics Platform*.

Onggo, B. S., Panadero, J., Corlu, C. G., & Juan, A. A. (2019). Agri-food supply chains with stochastic demands: A multi-period inventory routing problem with perishable products. *Simulation Modelling Practice and Theory*, *97*(July), 101970. https://doi.org/10.1016/j.simpat.2019.101970

Pan, J. C. H., & Hsiao, Y. C. (2002). Inventory models with back-order discounts and variable lead time. *International Journal of Systems Science*, *32*(7), 925–929. https://doi.org/10.1080/00207720120201

Parry, M. E., Cao, Q., & Song, M. (2011). Forecasting new product adoption with probabilistic neural networks. *Journal of Product Innovation Management*, *28*(SUPPL. 1), 78–88. https://doi.org/10.1111/j.1540-5885.2011.00862.x

Penpece, D., & Elma, O. E. (2014). Predicting Sales Revenue by Using Artificial Neural Network in Grocery Retailing Industry: A Case Study in Turkey. *International Journal of Trade, Economics and Finance*, *5*(5), 435–440. https://doi.org/10.7763/ijtef.2014.v5.411

Popluga, D., Pilvere, I., & Nipers, A. (2014). Main Development Trends of Grocery Retail Industry: Case Studies of Latvia, Lithuania and Estonia. *Economic Science for Rural Development: Marketing and Sustainable Consumption - Rural Development and Entrepreneurship - Home Economics*, (35), 129–136. Retrieved from https://apps.webofknowledge.com/full_record.do?product=WOS&search_mode=GeneralSearch&qid=162&SID=U1pmBw4lPObrKTl9Kkh&excludeEventConfig=ExcludeIfFromFullRecPage&page=1&doc=22

Prasad, K., & Mukherjee, B. (2016). Optimal inventory model under stock and time dependent demand for time varying deterioration rate with shortages. *Annals of Operations Research*, *243*(1–2), 323–334. https://doi.org/10.1007/s10479-014-1759-3

Pratola, M. T., Chipman, H. A., Gattiker, J. R., Higdon, D. M., McCulloch, R., & Rust, W. N. (2014). Parallel Bayesian Additive Regression Trees. *Journal of Computational and Graphical Statistics*, *23*(3), 830–852. https://doi.org/10.1080/10618600.2013.841584

Rahdar, M., Wang, L., & Hu, G. (2018). A tri-level optimization model for inventory control with uncertain demand and lead time. *International Journal of Production Economics*, *195*(October 2017), 96–105. https://doi.org/10.1016/j.ijpe.2017.10.011

Rezaei, J. (2014). Economic order quantity for growing items. *International Journal of Production Economics*, *155*, 109–113. https://doi.org/10.1016/j.ijpe.2013.11.026

Ridwan, M. (2019). *Prediction Of Lamp Price Using Adaptive Neuro Fuzzy Inference System*. (January 2018). https://doi.org/10.4108/eai.24-10-2018.2280522

Sadeghi, J., Sadeghi, S., & Niaki, S. T. A. (2014). A hybrid vendor managed inventory and redundancy allocation optimization problem in supply chain management: An NSGA-II with tuned parameters. *Computers and Operations Research*, *41*(1), 53–64. https://doi.org/10.1016/j.cor.2013.07.024

Sajadieh, M. S., Jokar, M. R. A., & Modarres, M. (2009). Developing a coordinated vendor-buyer model in two-stage supply chains with stochastic lead-times. *Computers and Operations Research*, *36*(8), 2484–2489. https://doi.org/10.1016/j.cor.2008.10.001

San-José, L. A., Sicilia, J., González-De-la-Rosa, M., & Febles-Acosta, J. (2019). Analysis of an inventory system with discrete scheduling period, time-dependent demand and backlogged shortages. *Computers and Operations Research*, *109*, 200–208. https://doi.org/10.1016/j.cor.2019.05.003

Sarker, B. R., & Al Kindi, M. (2006). Optimal ordering policies in response to a discount offer. *International Journal of Production Economics*, *100*(2), 195–211. https://doi.org/10.1016/j.ijpe.2004.10.015

Shi, N., Zhou, S., Wang, F., Xu, S., & Xiong, S. (2014). Horizontal cooperation and information sharing between suppliers in the manufacturer-supplier triad. *International Journal of Production Research*, *52*(15), 4526–4547. https://doi.org/10.1080/00207543.2013.869630

Sitompul, C., Aghezzaf, E. H., Chen, H., & Dullaert, W. (2006). A preliminary study on safety stock placement in capacitated supply chains. *IFAC Proceedings Volumes (IFAC-PapersOnline)*, *12*(PART 1). https://doi.org/10.3182/20060517-3-FR-2903.00318

Sun, Q., Wu, C., & Li, Y. L. (2017). A new probabilistic neural network model based on backpropagation algorithm. *Journal of Intelligent and Fuzzy Systems*, *32*(1), 215–227. https://doi.org/10.3233/JIFS-151415

Syahputra, R. (2016). Application of neuro-fuzzy method for prediction of vehicle fuel consumption. *Journal of Theoretical and Applied Information Technology*, *86*(1), 138–150.

Taleizadeh, A. A., Noori-Daryan, M., & Cárdenas-Barrón, L. E. (2015). Joint optimization of price, replenishment frequency, replenishment cycle and production rate in vendor managed inventory system with deteriorating items. *International Journal of Production Economics*, *159*, 285–295. https://doi.org/10.1016/j.ijpe.2014.09.009

Taleizadeh, A. A., & Pentico, D. W. (2013). An economic order quantity model with a known price increase and partial backordering. *European Journal of Operational Research*, *228*(3), 516–525. https://doi.org/10.1016/j.ejor.2013.02.014

Taleizadeh, A. A., Pentico, D. W., Aryanezhad, M., & Ghoreyshi, S. M. (2012). An economic order quantity model with partial backordering and a special sale price. *European Journal of Operational Research*, *221*(3), 571–583. https://doi.org/10.1016/j.ejor.2012.03.032

Tan, Y. V., & Roy, J. (2019). Bayesian additive regression trees and the General BART model. *Statistics in Medicine*, *38*(25), 5048–5069. https://doi.org/10.1002/sim.8347

Tarhini, H., Karam, M., & Jaber, M. Y. (2020). An integrated single-vendor multi-buyer production inventory model with transshipments between buyers. *International Journal of Production Economics, 225*(December 2019), 107568. https://doi.org/10.1016/j.ijpe.2019.107568

Teng, J. T. (2009). A simple method to compute economic order quantities. *European Journal of Operational Research*, *198*(1), 351–353. https://doi.org/10.1016/j.ejor.2008.05.019

Tersine, R. J. (1996). Economic replenishment strategies for announced price increases. *European Journal of Operational Research*, *92*(2), 266–280. https://doi.org/10.1016/0377-2217(95)00098-4

Tyan, J., & Wee, H. M. (2003). Vendor managed inventory: A survey of the Taiwanese grocery industry. *Journal of Purchasing and Supply Management*, *9*(1), 11–18. https://doi.org/10.1016/S0969-7012(02)00032-1

Van Donselaar, K. H., & Broekmeulen, R. A. C. M. (2013). Determination of safety stocks in a lost sales inventory system with periodic review, positive lead-time, lot-sizing and a target fill rate. *International Journal of Production Economics*, *143*(2), 440–448. https://doi.org/10.1016/j.ijpe.2011.05.020

Yadav, H. B., Kumar, S., Kumar, Y., & Yadav, D. K. (2018). A fuzzy logic based approach for decision making. *Journal of Intelligent and Fuzzy Systems*, *35*(2), 1531–1539. https://doi.org/10.3233/JIFS-169693

Yin, Yu-chun, Lee, C., & Wong, Y. (2012). *Demand Prediction of Bicycle Sharing Systems*. (2), 1–5.

Yin, Yuyu, Zhang, R., Gao, H., & Xi, M. (2019). New Retail Business Analysis and Modelling: A Taobao Case Study. *IEEE Transactions on Computational Social Systems*, *6*(5), 1126–1137. https://doi.org/10.1109/TCSS.2019.2933486

Yu, Y., Chu, F., & Chen, H. (2009). A Stackelberg game and its improvement in a VMI system with a manufacturing vendor. *European Journal of Operational Research*, *192*(3), 929–948. https://doi.org/10.1016/j.ejor.2007.10.016

Zhang, R. Q., Kaku, I., & Xiao, Y. Y. (2011). Deterministic EOQ with partial backordering and correlated demand caused by cross-selling. *European Journal of Operational Research*, *210*(3), 537–551. https://doi.org/10.1016/j.ejor.2010.10.001

APPENDICES

APPENDIX 1. Scientific article

11th International Scientific Conference
"Business and Management 2020"
May 7–8, 2020, Vilnius, LITHUANIA
Section: Business Technologies and Sustainable Entrepreneurship
http://www.bm.vgtu.lt

ISSN 2029-4441 / eISSN 2029-929X
ISBN 978-609-476-231-4
eISBN 978-609-476-230-7
https://doi.org/10.3846/bm.2020.604

STUDY OF DIFFERENT DATA SCIENCE METHODS FOR DEMAND PREDICTION AND REPLENISHMENT FORECASTING AT RETAIL NETWORK

Aleksei Iurasov[1], Giedre Stanelyte[2]

[1, 2]Department of Business Technologies and Entrepreneurship, Faculty of Business Management, Vilnius Gediminas Technical University, Saulėtekio av. 11, LT-10223, Vilnius, Lithuania
E-mail: [1] aleksei.iurasov@vgtu.lt (corresponding author); [2]giedre.stanelyte@stud.vgtu.lt

Received 11 March 2020; accepted 05 May 2020

Abstract. The demand prediction becoming an essential tool to remain or even lead in the competition among the retail businesses. A well-done demand prediction model could help retailer to track the level of inventory, orders and sales in the most effective way in which the best results could be achieved. However, there are many different methods and opinions of how to create a demand prediction model. In this paper, we will analyse the most commonly used methods of Linear regression, Logistic Regression, Probabilistic Neural Network, Bayesian Additive Regression Trees, Random Forest and Fuzzy Logic with their specifications and limitations found in studies of authors. After review performed all methods will be compared according to characteristics selected. Moreover, in order to get more practical results the accuracy of Logistic Regression and Random Forest methods will be compared based on data of milk sales collected from retail network. For constructing of decision support system for retail network, we need to go beyond demand prediction one-step to replenishment forecasting. It was concluded that there is no best method to forecast replenishment and results can differ based on the data and conditions analysing. In every situation authors seeking to select the method with the highest accuracy and the lowest number of errors possible. Limitations of research: limited number of goods and stores included in the modelling.

Keywords: demand prediction, replenishment forecasting, retail network, logistic regression, random forest.

JEL Classification: L81.

1. Introduction

Retail market today is one of the fastest growing markets in the world. This rapid growth of consumption and information technologies provides a number of opportunities for retail companies. A quick reaction and ability to work efficiently in changing business can deliver great results. Unfortunately, not all the organizations are working in the most effective way and modelling of demand prediction and replenishment here could be suggested as solution. A well-done demand prediction and replenishment model could help retailer not even to work more efficiently but also to increase company profit by saving the cost, increasing revenue and customer satisfaction. However even if decision to create a model would be accepted, very often retailers face the problem of lack of the knowledge about the methods those could be applied in modelling demand prediction in retail network. Moreover, as there are many different methods of demand prediction modelling, whether the best method for the most accurate prediction exists.

The aim of this study is to compare the application of different methods and by using the real time, data to evaluate the accuracy of two most commonly mentioned methods in practise. Tasks set for achieving the goal:

- To propose different mathematical methods for prediction and study its application in practice.
- To compare the methods by emphasizing its advantages and disadvantages.
- By forming demand prediction model with help of two different methods evaluate the accuracy of methods used.

Research object is enterprise demand prediction and replenishment modelling. Modelling prepared by using logistic regression and random forest methods. For modelling KNIME Analytics Platform was used. Due to the wide range of the products and business transactions in the analysing market model was formed based on sales of limited number of goods in defined number of stores.

2. Mathematical methods for demand prediction modelling

Demand prediction is the combination of two words. The first one – demand, and second – prediction. Before combining those words into phrase, and further reviewing different methods applicable, it is important to understand economical meaning and value of this phrase. Word demand – could be define as requirement of products and services, while prediction – in general, means making estimation in present for future events, at this case – future demand of products. Demand plays a vital role in the decision making of a business. In competitive market conditions, there is a need to take correct decision and make planning for future events related to business like a sale, production or inventory optimization. The better analysis of different factors are performed – the better decisions based on demand prediction results could be made. As the process by itself is difficult, there are many different methods invented and studies made based on the applications of those methods. In this part – application of the methods of linear regression, Logistic regression, Probabilistic Neural Network, Bayesian Additive Regression Trees, Random Forest and Fuzzy Logic will be reviewed and discussed.

The first method – linear regression is the basic and often used type of predictive analysis. Linear regression uses only one independent variable as a predictor, which has an effect to dependent variable (outcome). The main idea of regression analysis is to determinate the strength of predictor, to forecast an effect, and to use the model formed in forecasting the further results. However, the method when there is only one predictor is not clear enough to predict possible subsequent development of the analyzing process, therefore the authors commonly choose a *Multiple Linear regression.* This means that linear regression can be extended and instead of one independent variable, we have a many different variables (Anghelache, 2015). As the method is treated as clearly understandable, it is widely used in various studies and researches to predict outcomes selected by different authors. Author (Anghelache, 2015) used multiple linear regression to analyze the final consumption and gross investment influence to Romania GDP. Based on the data collected authors formed an equation and based on statistical tests evaluate the accuracy of the model. Authors conclude that higher number of factors in regression model allows the researcher to draw results that are more conclusive in macroeconomic analysis. Authors (Aghdaei et al., 2017) also used a regression in building of energy simulation model. The aim of the study was to predict the space heating and cooling requirements in different cities of Australia. Findings of the research show that the linear regression with simple independent variables can predict the requirements for space heating and cooling of the residential buildings in the specific climates within acceptable errors. It can even be applied in the studies where the relation between the financial news as an independence variable and the stock price of financial market is evaluating. In this case the author of the work named regression as machine learning-based approach (Ihlayyel et al., 2018). However, as method is quite simple, for better results it has to be used in combination with some rules, methods or algorithms. In the researches mentioned above to reduce the modelling cost of the parametric analysis authors (Aghdaei et al., 2017) use methods of Taguchi and Analysis of variance (ANOVA). Other author Ihlayyel mentioned above applied Enhanced ELR-BoW (Bag of Words) algorithm. Only after simplification of the data regression models were formed. The research of the authors (Cankurt & Subasi, 2015) also tried to compare the linear regression with neural network and agree with the idea above. Formed forecast shows that neural network present higher accuracy than the linear regression when there is no combination with linear regression and other methods.

When it comes to regression it is always important to mention – logistic regression. Unlike traditional linear regression – logistic regression is appropriate for modelling a binary variable. In wider perspective it means that results can procedure two outcomes (1 or 0), those could be considered as "positive" and "negative". Such results are useful in the practice however can't be received by using simple linear regression. This is mostly due to two main reasons. Firstly, a simple linear regression can only predict values outside the acceptable range. Secondly, as the dichotomous experiments can only have one of two possible outcomes for every experiment, the residuals will not be normally allocated about the predicted line. Performing analytics with logistic regression includes three main goals: prediction that the outcome or response variable equals to 1, categorization of outcomes and predictions and finally, access to the odds or risk associated with model predictors (Grömping, 2016). Logistic regression is considered as very important statistical procedure in predictive analytics in areas of health-care, medical analysis, social statistics and economy. The authors Joubert, Verster, and

Raubenheimer (2019) included logistic regression as commonly used method to predict probability components and loss severity in study of Loss Given Default (LGD) evaluation in banks. In authors study probability components were modelled by making use of logistic regression binary outcomes (write-offs or not write-offs). Moreover in study logistic regression was used in combination with method of survival analysis, what let to increase the model's predictive power and accuracy of results.

As it is mention above, one more method often included into studies is – Probabilistic Neural Network (PNN). PNN is the method of artificial intelligence that allow to form a complex nonlinear relationship between response variables and explanatory ones. The network was introduced in late 20ths – in 1990 by Specht. The main characters after presentation of network were – easy to use and possibility to interpret the network's structure in the form of a probability density function which is simple to understand. For those reasons the method is used in various sectors to analyze. Authors Penpece and Elma (2014) used neural networks for the purpose to forecast Sales Revenue in Grocery Retailing Industry. Based on results received authors stated that neural network method is more organic and predicting results better than the other methods. Revenue forecasts calculated based on neural networks were very close to actual data of sales revenue. However, the model also faced some problems of estimation of probability density function and high space complexity of PNN pattern layer. Moreover, model by itself has only one parameter of training and the smoothing parameter (σ), which must be optimized in order to make the network achieve the highest prediction ability. Based on different scientists' network is composed of 3 either 4 layers: an input layer, a pattern layer, a summation layer, and an output layer. Some scientists (Sun et al., 2017) does not count last layer and named third layer as Classes. The neurons in the input layer are simply the features of input vectors. The pattern layer consists of as many neurons as training examples. In the summation layer, the number of neurons is equal to the cardinality of classes in the data set. Finally, the output layer consists of a single neuron that provides the classification result. As the structure of the network is considered as a complex and probabilistic neural network is a frequently exploited model in the field of data mining in different researches of scientists, certain PNN reduction techniques have been established. These techniques include dynamic decay adjustment algorithm (DDA) (Berthold & Diamond, 1998) backpropagation mechanism (Sun et al., 2017), dimensionality reduction (Kusy, 2015) and also some other, presented earlier (in 1991–1994) – learning vector quantization (Burrascano, 1991) or maximum-likelihood algorithm techniques. All the techniques have the own specific parameters and the influence on the network. Further mainly parameters and the practical value of techniques are reviewed:

- Dynamic adjustment algorithm (DDA). Operation of algorithm required two phases – training and classification. New neurons are added if necessary. Less than five epochs are needed to complete training. The algorithm can be proven to terminate when a finite number of training examples is used. And finally, only two thresholds are required to be adjusted manually (Berthold & Diamond, 1998).
- Dimensionality reduction. For the reduction creation model by author (Kusy, 2015) two main steps – feature selection (methods of single decision tree (SDT) and random forest (RF)) and feature extraction (method of principal component analysis (PCA)) need to be follow. The main idea of the SDT is treated method as a predictive model – it maps the input data into desired targets. If the desired targets take the form of groups to which the data belong, SDT is treated as classification tree. After this process RF utilizes the collection of independent decision trees formed by SDT. Within the training process, the trees grow in parallel, not interacting until all of them have been grow. Once the training is completed, the need to move to the next phase appears and predictions of single trees are combined to make the overall prediction of RF. For the further step – principal component analysis is used. PCA is one of most popular feature extraction method to use. The methods combine the statistical technique which converts a set of input features into a set of new values by means of linear transformation. The results are names principal components and are linearly uncorrelated. With the help of method patterns of similarities and differences in data can be identified. Once these patterns are determined, the data can be compressed by decreasing the number of dimensions without significance loss of information. According to

the author Kusy (2015), who applied this variation of method for medical data classification tasks, the results showed an increase of prediction ability and decrease in computational time needed to complete the task in every single case.

- Backpropagation mechanism (BP). The authors (Parry et al., 2011) define the method as feed-forward neural network and included it into the test of prediction accuracy in the first-time adoption of DVD players. The authors forecasted performance of the logit model and the three neural network models. At the end it was found that the PNN algorithm significantly outperforms the logit model and the two remaining neural network algorithms. BP was one of those methods. However in 2017 scientists (Sun et al., 2017) present the idea of new PNN model in combination with backpropagation algorithm. New BP-PNN model has two phases – learning phase, where the idea is to receive the initialized value of the variable weights and training phase – where the error function is propagated back. Based on the analyses performed by the authors comparing with PNN, BP-PNN has less components in the pattern layer, which helps to reduce the space complexity of the model. Moreover, comparing with PNN, BP-PNN is designed with the much clearer structure and ability to identify the importance of indicators. Nevertheless, there are still some limitation of model emphasized by the authors – the model required further studies as the number of parameters trained is higher than in other models and thus requires a long time for calculations in the model.

Review emphasized that the method of PNN PNN needs to be combine with one of the algorithms reviewed for the purpose to receive the better results.

One more popular method – Bayesian Additive Regression Trees (BART) by authors Pratola et al. (2014) were described as nonparametrically method. Logan, Sparapani, McCulloch, and Laud (2019) define BART method as fully nonparametric and flexible model of prediction which can deal with complex functional forms as well as interactions among wide range of variables The authors Ajidarma and Irianto (2019) in their study agree with the mentioned definition by adding the idea that BART regression method harnesses dimensionally adaptive random basis elements. The method consisting from a prior and likelihood and is a function of an ensemble of trees. As method can combine and compare a number of different factors it is widely used in the researches of different authors. Method can use different algorithms with whose help more precise search of the model space and variation across the algorithm draws can be organized. However there are some problems as publicly available version of algorithm in R package can process only the limited number of observations (Pratola et al., 2014). Moreover the authors Linero and Yang (2018) performed a study for one more problem of decision threes to analyze. According to the authors in those methods – a high possibility of deficiency in ensembles is possible and could be caused by lack of smoothness and vulnerability to the curse of dimensionality. The idea of soft decision trees was suggested and implemented on the BART method. The authors demonstrate that their methods can have meaningful improvement over existing methods. Yet as there are still a lot of limitation, further studies of idea is performing. Ajidarma and Irianto (2019) have included the BART in the analysis and prediction the growth of electric automobile industry in different states of United States. BART method was used to analyze the relationship between several factors chosen and the sales of electric vehicles. With help of the algorithms (on this case Markov chain Monte Carlo (MCMC) algorithm was chosen) four BART models were generated and fitted into the data. The models identified the top predictors those correlation with the sales of electric vehicles were confirmed in the further steps. Authors of the work concluded that to use the method was beneficial – as the method enables a full assessment of prediction uncertainty while remaining highly competitive in terms of prediction accuracy.

Moving forward to Random forest method – it is one more decision trees method which consist of a chosen number of decision trees, which are used for classification and regression analysis (Feng & Wang, 2017). Authors Gupta, Rawat, Jain, Arora, and Dhami (2017) in their study of decision methods described this random forest method as tool that form the ability of multiple varied analyses, organization strategies, supply and demand prediction modelling, perceptive variables and importance ranking on the record-by-record basis for deep data understanding. Instead of tool authors Yin, Lee, and Wong (2012) define random forest method as the algorithm with an ensemble random method. Author Ghatasheh (2014) agree with the definitions mentioned and emphasized some more positive attributes, such as an immunity to overfit, good estimation of internal errors, and high accuracy comparing with other learning algorithms. Because of these reasons' method is included in various calculations and studies performed. In 2017 authors Feng & Wang made a study, where the demand of the bicycle rental was evaluating. Two methods – multiple linear regression analysis and random forest were included into calculation. However, research shows that accuracy of linear regression model to forecast is too low even though the normal distribution of factors and good relationship between them was identified. Authors identified the high possibility of error due to specific characteristics of factors. The method was changed to random forest method and this improve the accuracy of result to 82%.

The last method reviewed in this article – Fuzzy method. The authors Agápito et al. (2019) characterize Fuzzy method as logic which with a help of specific set of rules can make an associations between linguistic and numeric data of a database. Rules can be provided by two groups – artificial intelligence algorithms, which are the target of researches in nowadays and by group of experts. The author Syahputra (2016) in her study agree with the definition about the logic, but additionally emphasized that this logical function recognizing only two parameters, either „Yes“ or „No“ („1“ or „0“). The logic mostly used to create an expert systems and knowledge – based control settlements. However there are some limitations, those are important to know before analysing the method in more detail. Firstly, the method is case-dependent – every time the changing scenario can have a different influencing factors, where each of them need to be evaluated as important. Secondly, contribution of domain experts is of significant importance in process of forming control settlements (Yadav et al., 2018). Moreover there also could be the problem of too many rules. The problem usually arise among the rule-based models, when rules are created for every single factors and the outliers, those were not identified in the beginning of modelling are detected only in the process (Berthold, 2003). Depside it, the fuzzy logic method is applying in studies, yet usually in combination with other methods. The author Syahputra (2016) performed a study to predict a vehicle fuel consumption by using the combination of artificial neural networks and fuzzy logic (ANFIS). Prediction was made for different models of cars based on two criteria – weight and age. Results show that with an increase of weight of the motor vehicle, an amount of fuel needed to travel the same distance is increasing. Moreover car age also affects fuel efficiency. For the younger car, the higher efficiency of fuel consumption is calculated. The same combination of methods were also used by author Ridwan (2019) who used adaptive network-based fuzzy inference system (ANFIS) in the study to predict the price of good – lamp.

As all methods reviewed are different, for better understanding it is important to summarize what are the advantages and disadvantages of every method and what are the areas the method could be applied. The comparison of methods shows that there are no one perfectly suitable method to use (see Table 1). All the methods have their advantages and also areas to improve. For further calculations two methods: Logistic Regression (as the most popular for easy interpretation and implementation) and Random Forest (by studies emphasized as one of the most accurate learning algorithm) were selected for demand modelling and replenishment forecasting.

3. Methodology of Logistic regression and Random forest

Data processing for the practical part of this study is conducted by using KNIME Analytics Platform. KNIME (Konstanz Information Miner) is an open-source Big Data Analytics Platform, which used for data analytics and reporting. These processes organized in KNIME in form of workflows. The main principals of workflow are – visualization, modularity and easy extensibility (Berthold et al., 2009). The workflow consist of many nodes, those are processing the data and transporting results via connections between the nodes. The work in this workflow is organized based on the structure of 4 main stages: a) data collection; b) pre-processing c) model development and training and d) prediction and review of the results (scores) (Ranji et al., 2019). Included methods – logistic regression and random forest.

Logistic regression, as mentioned above models the probabilities for classification problems with two possible outcomes (1 & 0, e.g. "No action"&"Replenish"). The method is widely used in various fields of the practise. Logistic regression curve is constructed using the natural logarithm of the "odds" of the target variable. The formula of logistic regression:

$$P = \frac{1}{1+e^{-(a+bX)}}, \quad (1)$$

where: P – probability of 1; e – the base of natural logarithm; a, b – parameters of the model.

Table 1. Advantages, disadvantages and application of mathematical methods proposed

Characteristics	Linear Regression	Logistic regression	PNN	BART	Random forest	Fuzzy logic
Basic principles of method	Determination of relationship between independent and dependent variables.	Prediction and determination of relationship between variables when the dependent variable is binary.	Classification of patterns based on learning from examples.	Creation of sum-of-trees model and regularization prior on the parameters of that model.	Combination of tree predictors where each tree depends on the values of a random vector sampled independently with the same distribution for all trees in the forest.	Form of many-valued logic in which the truth values of variables may be any real number between 0 and 1 both inclusive.
Authors	Anghelache; Aghdaei; Kokogiannakis, Daly & McCarthy; Ihlayyel, Sharef, Nazri & Bakar, Aghdaei et al.; Cankurt, Subasi.	Grömping; Joubert, Verster, & Raubenheimer.	Penpece & Elma; Sun,Wu&Li; Berthold & Diamond; Kusy;Sun; Burrascano;Parry, Cao& Song.	Pratola; Logan, Sparapani, McCulloch, & Laud; Ajidarma & Irianto; Pratola; Linero & Yang.	Feng & Wang; Gupta, Rawat, Jain, Arora, & Dhami; Yin, Lee, & Wong; Ghatasheh.	Agápito; Syahputra; Yadav, Kumar, Kumar, & Yadav; Berthold.
Advantages	– Simple estimation procedure; – Easy to understand interpretation on a modular level (i.e. the weights).	– Gives not only a measure of how relevant a predictor is, but also its direction of association; – Is easy to implement, efficient to train.	– Less time consumed to train virtually; – Relatively sensitive to outliers; – Can calculate probability scores.	– Provides a flexible approach to fitting a variety of regression models and algorithms while avoiding strong parametric assumptions.; – Able to represent interactions; – Can handle missing values.	– One of the most accurate learning algorithms; – Runs efficiently on large database; – Provide a reliable feature estimate; – Offer estimates of the test error and missing data.	– Using simple mathematics for non linear, integrated and complex systems; – Is based on linguistic model; – Has rapid operations; –Can handle trouble with inaccurate data.
Disadvantages	– Can oversimplifies the problems, and make a linear relationship where there is no such; –Is sensitive to outliers; –Linearity between variables. Hard to achieve in real world; – The data must be independent.	– Dependent variable is restricted; – If number of observations < number of features, lead to possible overfit.	– Required a lot of memory for data; – Usually long testing time; – Possible large computational cost.	– In publicly available version of algorithm in R package can process only the limited number of observations (Pratola et al., 2014); – A high possibility of deficiency due to lack of smoothness and vulnerability to the curse of dimensionality. (Linero & Yang, 2018).	– An ensemble model is inherently less interpretable than an individual decision tree; – Training a high number of trees could lead to large computational cost and can use a lot of memory.	– Case-dependent; – Contribution of domain experts is of significant importance in process of forming control settlements (Yadav et al., 2018) – There also could be the problem of too many rules; – Based on the size, high memory and cost required.
Business processes were methods are applied	Inventory, sales prediction, evaluation of marketing effectiveness, promotions.	Supply chain management, inventory, sales forecasting.	Selection of retail stores location, prediction of sale revenue, promotions.	Supply chain management, demand and orders prediction modelling, correlated outcomes.	Prediction of retail demand, optimal retail location, pricing.	Market trend analysis, retail service quality evaluation, consumption and demand predictions.

In the logistic regression the parameter *a* moves the curve left and right and parameter *b* – defined the steepness of the curve. To obtain model coefficients, those best relate predictors to target logistic regression uses maximum likelihood estimation method.

Random forest method – adding to the definitions above, this method also could be defined as an algorithm consisting of a collection of tree structured classifiers that created for independently and identically distributed random vectors. Each tree casts a unit vote for the most popular class for each input when the response is binary. For regression problems, the random forest forecast is the average prediction from the regression trees (Weng et al., 2018). Main steps of the method are showed in the Figure 1:

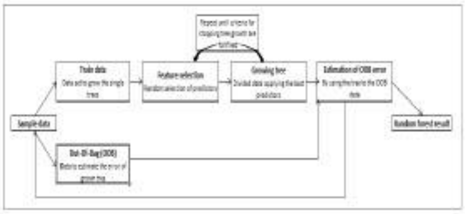

Figure 1. Random forest algorithm (source: Boulesteix, Janitza, and Kruppa (2012))

Firstly, random samples from a given dataset are selected. Then when the data is selected, the best predictors is found and applied on the data analyzing. This step has to be repeated until criteria for stopping tree growth are fulfilled. Moreover, it is important to estimate and include into calculations the Out of Bag (OOB) error. Finally, when all the requirements are fulfilled, final prediction result is formed.

4. Demand prediction and replenishment modelling on behalf of Logistic regression and Random forest methods

In the practical part the KNIME workflow was created, the logistic regression and random forest methods based demand prediction and replenishment models were generated and assessed (see Figure 2).

As it is mentioned in the methodology part, there are some main steps to do to prepare the models. In the first step data is collected. Data selected to use in this practical case:

- Sales of one product – milk (dependent variable). Sales data collected from 10 different shops (product number, product name, quantity sold, etc.).
- Time information: year; day, day of month, day of week.
- Other additional information, as an independent variables – product delivery information (frequency, delivery time), number of customers per day, average basket information, discounts information, holidays, weather information.

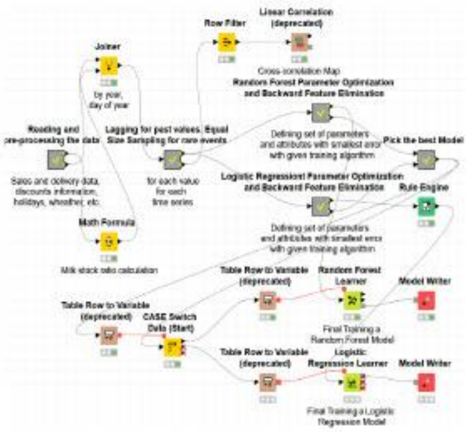

Figure 2. Demand prediction and replenishment models (top level of KNIME workflow)

After uploading the data are prepared for the further use: new parameters, e.g. Milk ratio (Milk stock ratio) as an additional measure is calculated by dividing milk stocks (in every store in every time frame – day) by optimal order quantity. We did this because retail companies don't use demand prediction itself but as part of decision support to replenish goods. Therefore the ratio can better identify the need for replenishment than just the quantity of items currently available and quantity of items sold.

Actually, we need to forecast whether an action is needed: "No action" or "Replenish". The model evaluated the correlation between the last 10 ratios and provided it in the user-friendly matrix (see Figure 3).

For the purpose to improve the prediction accuracy and to find out which columns are necessary for the model the parameter optimization and backward feature elimination loops was added (see Figure 4). Feature elimination loop work in the way that lets to include in the output table the columns that is best for model construction. We specify an error threshold and then the level with the fewest features that has a prediction error below the threshold is automatically selected. In any case all columns from the input table that are not present in the selected level are filtered from the input table. In this way only the necessary columns remain in the model what ensure the more accurate results. From the economic side – feature elimination loop helps the company to identify only the real attributes with significant impact to result. It define set of attributes with smallest error possible. Need of this step appear as in the beginning of modelling more than one different attributes are adding.

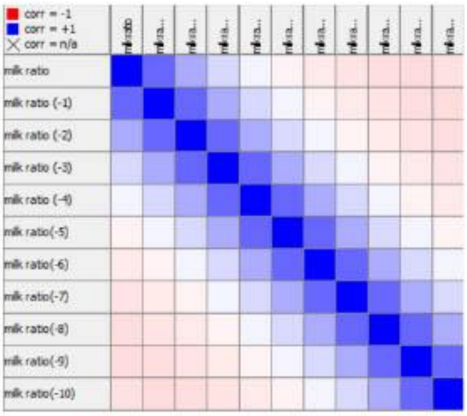

Figure 3. Milk ratios correlation matrix

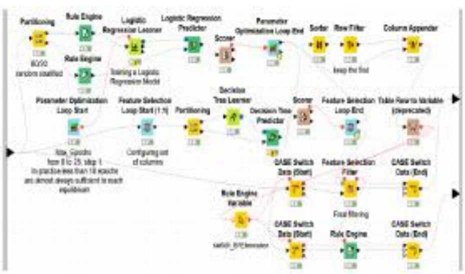

Figure 4. Logistic Regression Parameter Optimization and Backward Feature Elimination (second level of KNIME workflow)

However without separate procedures it would be hard to evaluate possible impact to result. And here feature elimination loop works as combined filter, where the attributes identified as high in level of error are eliminating from the model. In this specific case discounts information, holidays, weather and some milk stock ratios of previous periods (–3,…,–10) were identified as attributes those do not have a significant impact to decision of milk replenishment. All the others attributes were left as significant and meaningful in final calculations. After this step the final calculations are proceed based on eliminations made and results of the models are provided.

The results are showed that in this specific case of Milk product replenishment modelling logistic regression model is by 11% less accurate than the model prepared on behalf of Random Forest method (see Table 2).

Table 2. Logistic regression and Random forest methods accuracy results in prediction models formed

		Predicted data			
	Forecasting methods	Logistic regression		Random forest	
	Actions	No action	Replenishment	No action	Replenishment
Factual data	No action	2491	815	2916	390
	Replenishment	665	3289	307	3647
Accuracy		79.61%		90.40%	
Error		20.39%		9.60%	
Cohen's kappa (k)		0.587478		0.806048	

However the small difference and ideas marked before in the article only confirmed that every case is different and there is no one best method for prediction. Only the precise evaluation of data can shows which method provide the best quality on every specific study performing.

5. Conclusions

In the study Linear regression, Logistic regression, Probabilistic Neural Network, Bayesian Additive Regression Trees, Random Forest and Fuzzy Logic were suggested as the methods those could be used in modelling demand prediction of retail network. Practical application, advantages and disadvantages of each were evaluated and compared. Results of comparison showed all methods analysed have areas to improve. Recurring problems identified: possible deficiencies, high memory requirements, long testing time and possibly large computational cost.

In practical study the model of milk demand prediction and replenishment on behalf of two methods proposed in theoretical part was formed.

Results showed that model formed on behalf of logistic regression method was by 11% less accurate than the model created with help of random forest method.

Authors with the help of this practical case also underline that accuracy of the different methods applied in modelling can be evaluated and compared. As every case in the practice is different there is no best method perfectly suitable for all the situations. The authors Sarkar and Mahapatra (2017) confirmed the statement by adding that in real-life situation, it is very difficult to know all the information about the demand. Moreover the information can differ base on the situation, therefore in every case factors affecting the model may vary. Only evaluation of model received can show which method provide the best quality on every specific study performing.

From economic perspective – practical case shows that modeling could help the business to understand better what are the attributes they need to include into considerations before making significant decisions. Discounts information, holidays, weather and some milk stock ratios were not attributes having significant impact to decision of milk replenishment. Their identification helps to adjust the model and by doing this to receive a more precise results.

As practical study evaluated only two methods proposed, future researches could focus on evaluation of more complex models with higher number of methods to include.

References

Agápito, A. D. O., Vianna, M. D. F. D., Moratori, P. B., Vianna, D. S., Meza, E. B. M., & Matias, I. D. O. (2019). Using multicriteria analysis and fuzzy logic for project portfolio management. *Brazilian Journal of Operations & Production Management, 16*(2), 347–357. https://doi.org/10.14488/bjopm.2019.v16.n2.a14

Aghdaei, N., Kokogiannakis, G., Daly, D., & McCarthy, T. (2017). Linear regression models for prediction of annual heating and cooling demand in representative Australian residential dwellings. *Energy Procedia, 121*, 79–86. https://doi.org/10.1016/j.egypro.2017.07.482

Ajidarma, P., & Irianto, D. (2019). Application of bayesian additive regression trees to analyze the growth of United States electric automobile industry. *IOP Conference Series: Materials Science and Engineering, 598*(1). https://doi.org/10.1088/1757-899X/598/1/012035

Anghelache, C. (2015). Analysis of final consumption and gross investment influence on GDP – multiple linear regression model. *Theoretical and Applied Economics, 22*(3), 137–142.

Berthold, M. R. (2003). Mixed fuzzy rule formation. *International Journal of Approximate Reasoning, 32*(2–3), 67–84. https://doi.org/10.1016/S0888-613X(02)00077-4

Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Thiel, K., & Wiswedel, B. (2009). KNIME – the Konstanz information miner. *SIGKDD Explorations, 11(1)*, 26–31. https://doi.org/10.1145/1656274.1656280

Berthold, M. R., & Diamond, J. (1998). Constructive training of probabilistic neural networks. *Neurocomputing, 19*(1–3), 167–183. https://doi.org/10.1016/S0925-2312(97)00063-5

Boulesteix, A., Janitza, S., Kruppa, J., & König, I. R. (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *WIREs, 2*(6), 493–507. https://doi.org/10.1002/widm.1072

Burrascano, P. (1991). Learning vector quantization for the probabilistic neural network. *IEEE Transactions on Neural Networks, 2*(4), 458–461. https://doi.org/10.1109/72.88165

Cankurt, S., & Subasi, A. (2015). Comparison of linear regression and neural network models forecasting tourist arrivals to Turkey. *Eurasian Journal of Science & Engineering, 1*(1), 21–26.

Feng, Y., & Wang, S. (2017). A forecast for bicycle rental demand based on random forests and multiple linear regression. *Proceedings – 16th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2017*, 101–105. https://doi.org/10.1109/ICIS.2017.7959977

Ghatasheh, N. (2014). Business analytics using random forest trees for credit risk prediction: A comparison study. *International Journal of Advanced Science and Technology, 72*, 19–30. https://doi.org/10.14257/ijast.2014.72.02

Grömping, U. (2016). Practical guide to logistic regression. *Journal of Statistical Software, 71*. https://doi.org/10.18637/jss.v071.b03

Gupta, B., Rawat, A., Jain, A., Arora, A., & Dhami, N. (2017). Analysis of various decision tree algorithms for classification in data mining. *International Journal of Computer Applications, 163*(8), 15–19. https://doi.org/10.5120/ijca2017913660

Ihlayyel, H. A. K., Sharef, N. M., Nazri, M. Z. A., & Bakar, A. A. (2018). An enhanced feature representation based on linear regression model for stock market prediction. *Intelligent Data Analysis, 22*(1), 45–76. https://doi.org/10.3233/IDA-163316

Joubert, M., Verster, T., & Raubenheimer, H. (2019). Making use of survival analysis to indirectly model loss given default. *ORiON, 34*(2), 107–132. https://doi.org/10.5784/34-2-588

Kusy, M. (2015). Dimensionality reduction for probabilistic neural network in medical data classification problems. *International Journal of Electronics and Telecommunications*, *61*(3), 289–300. https://doi.org/10.1515/eletel-2015-0038

Linero, A. R., & Yang, Y. (2018). Bayesian regression tree ensembles that adapt to smoothness and sparsity. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *80*(5), 1087–1110. https://doi.org/10.1111/rssb.12293

Logan, B. R., Sparapani, R., McCulloch, R. E., & Laud, P. W. (2019). Decision making and uncertainty quantification for individualized treatments using Bayesian Additive Regression Trees. *Statistical Methods in Medical Research*, *28*(4), 1079–1093. https://doi.org/10.1177/0962280217746191

Parry, M. E., Cao, Q., & Song, M. (2011). Forecasting new product adoption with probabilistic neural networks. *Journal of Product Innovation Management*, *28*(Suppl 1), 78–88. https://doi.org/10.1111/j.1540-5885.2011.00862.x

Penpece, D., & Elma, O. E. (2014). Predicting sales revenue by using artificial neural network in grocery retailing industry: A case study in Turkey. *International Journal of Trade, Economics and Finance*, *5*(5), 435–440. https://doi.org/10.7763/ijtef.2014.v5.411

Pratola, M. T., Chipman, H. A., Gattiker, J. R., Higdon, D. M., McCulloch, R., & Rust, W. N. (2014). Parallel bayesian additive regression trees. *Journal of Computational and Graphical Statistics*, *23*(3), 830–852. https://doi.org/10.1080/10618600.2013.841584

Ranji, R., Thanavanich, C., Sukumaran, S. D., Kittiwachana, S., Zain, S., Sun, L. C., & Lee, V. S. (2019). An automated workflow by using KNIME analytical platform: A case study for modelling and predicting HIV-1 protease inhibitors. *Progress in Drug Discovery & Biomedical Science*, *2*(1), 4–8. https://doi.org/10.36877/pddbs.a0000035

Ridwan, M. (2018). Prediction of lamp price using adaptive neuro fuzzy inference system. *ICCSET 2018* (pp. 742–751), 25–26 October 2018. Kudus, Indonesia. https://doi.org/10.4108/eai.24-10-2018.2280522

Sarkar, B., & Mahapatra, A. S. (2017). Periodic review fuzzy inventory model with variable lead time and fuzzy demand. *International Transactions in Operational Research*, *24*(5), 1197–1227. https://doi.org/10.1111/itor.12177

Sun, Q., Wu, C., & Li, Y. L. (2017). A new probabilistic neural network model based on backpropagation algorithm. *Journal of Intelligent and Fuzzy Systems*, *32*(1), 215–227. https://doi.org/10.3233/JIFS-151415

Syahputra, R. (2016). Application of neuro-fuzzy method for prediction of vehicle fuel consumption. *Journal of Theoretical and Applied Information Technology*, *86*(1), 138–150.

Weng, B., Lu, L., Weng, B., Lu, L., Wang, X., Megahed, F. M., & Martinez, W. (2018). Predicting short-term stock prices using ensemble methods and online data sources predicting short-term stock prices using ensemble methods and online data sources. *Expert Systems with Applications*, *112*, 258–273. https://doi.org/10.1016/j.eswa.2018.06.016

Yadav, H. B., Kumar, S., Kumar, Y., & Yadav, D. K. (2018). A fuzzy logic based approach for decision making. *Journal of Intelligent and Fuzzy Systems*, *35*(2), 1531–1539. https://doi.org/10.3233/JIFS-169693

Yin, Y., Lee, C., & Wong, Y. (2012). Demand prediction of bicycle sharing systems. (2), 1–5. http://cs229.stanford.edu/proj2014/Yu-chun%20Yin,%20Chi-Shuen%20Lee,%20Yu-Po%20Wong,%20Demand%20Prediction%20of%20Bicycle%20Sharing%20Systems.pdf