Lappeenranta-Lahti University of Technology LUT
School of Engineering Science
Industrial Engineering and Management
Data Analytics in Decision Making

Emmi Huovila

USE OF MACHINE LEARNING IN SUPPLY CHAIN MANAGEMENT - CASE STUDY WITH DATAROBOT

Master's Thesis

Examiners: Professor Janne Huiskonen
Professor Ajantha Dahanayake

ABSTRACT

Lappeenranta-Lahti University of Technology LUT
School of Engineering Science
Industrial Engineering and Management
Data Analytics in Decision Making

Emmi Huovila

Use of Machine Learning in Supply Chain Management - Case Study with DataRobot

Master's Thesis

2021

67 pages, 16 figures, 6 tables, 1 appendix.

Examiners: Professor Janne Huiskonen
Professor Ajantha Dahanayake

Keywords: machine learning, supply chain management, DataRobot

In this Master's thesis, machine learning in supply chain management was studied. The goal of this thesis was to find out in which problems machine learning is suitable in the field of supply chain management and what are its benefits. The thesis was divided into two sections, literature review and empirical part. In the literature review, case examples were looked at from literature. Case examples were found for many different needs on different fields, such as demand prediction, inventory management, supplier selection and evaluation, and location and logistics problems. The benefits of machine learning were financial, and it gave aid in decision making and in improvement of customer satisfaction. In the empirical part, DataRobot platform's functions were studied by creating a case study of inventory management. The challenge in the case study was significant imbalance between classes. In this case study, the results were tolerable, but DataRobot offers for companies the possibility to try machine learning without great resources.

TIIVISTELMÄ

Lappeenrannan-Lahden teknillinen yliopisto LUT
School of Engineering Science
Tuotantotalous
Data-analytiikka päätöksenteossa

Emmi Huovila

Koneoppimisen käyttö toimitusketjun johtamisessa - case-esimerkki DataRobotilla

Diplomityö

2021

67 sivua, 16 kuvaa, 6 taulukkoa, 1 liite.

Tarkastajat: Professori Janne Huiskonen
Professori Ajantha Dahanayake

Hakusanat: koneoppiminen, toimitusketjun johtaminen, DataRobot

Tässä diplomityössä tutustuttiin koneoppimiseen toimitusketjun johtamisen alueella. Tavoitteena oli selvittää, millaisiin ongelmiin koneoppiminen sopii toimitusketjun johtamisen alueella ja kuinka siitä voidaan hyötyä. Tämän selvittämiseksi työ on jaettu kahteen osaan, kirjallisuuskatsaukseen ja kokeelliseen osuuteen. Kirjallisuuskatsauksessa etsittiin case-esimerkkejä. Case-esimerkkejä löydettiin erilaisiin tarpeisiin useilla osa-alueilla, kuten kysynnän ennustaminen, varaston hallinta, toimittajan valinta ja arviointi sekä sijainti- ja logistiikkaongelmat. Koneoppimisesta hyödyttiin taloudellisesti sekä saatiin apua esimerkiksi päätöksen tekoon ja asiakastyytyväisyyden parantamiseen. Kokeellisessa osuudessa DataRobot-ohjelman toimintoihin tutustuttiin luomalla oma case-esimerkki varaston hallintaan liittyen. DataRobotilla tehdyssä case-esimerkissä haasteena oli merkittävä epätasapaino luokkien välillä. Tässä tapauksessa saatiin välttäviä tuloksia, mutta DataRobot tarjoaa yrityksille mahdollisuuden kokeilla koneoppimista ilman suuria resursseja.

PREFACE

I want to thank everyone who helped me with this thesis. As well, I want to thank those who have been beside me during my studies. The time spent in university has been great but it is also nice to conclude my studies and move forward to new challenges.

Lappeenranta, March 25, 2021

Emmi Huovila

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ANN	Artificial Neural Network
API	Application Programming Interface
CTL	Center for Transportation and Logistics
EGBTC	eXtreme Gradient Boosted Trees Classifier with Early Stopping
GLM	GLM Blender Algorithm
GPS	Global Positioning System
MCC	Matthews Correlation Coefficient
ML	Machine Learning
RFID	Radio Frequency Identification
ROC	Receiver Operating Characteristic
SCM	Supply Chain Management
SCOM	Supply Chain Ordering Management
SCOR	Supply Chain Operations Reference
SVM	Support Vector Machine
TPR	True Positive Rate

1 INTRODUCTION

1.1 Background

Increased amount of data leads to the development in the area of the supply chain management (SCM) these days. To be able to succeed, companies have to understand what the key forces of the marketplace are and how to respond to the chances fast enough (Ittmann 2015). It is estimated in MIT's Center for Transportation and Logistics (CTL) by Sergio Caballero, research scientist, that half of the 80 companies working on projects with CTL do not use artificial intelligence (AI) at all, when 25% have done AI pilots, and 25% are already using AI (Forger 2020). One of the methods of AI is machine learning (ML), which has increased in the area of SCM starting from 1999 based on the amount of the released articles (Ni, Xiao and Lim 2019). The peaks in the articles released have been in 2008-2009, due to financial crisis, and 2016-2017 when AI have become more popular (Ni et al. 2019).

This thesis will be focused on ML as it is less researched than data analytics generally. In the literature review of supply chain risk management, only small proportion, 5%, were machine learning and big data applications. Most of those ML articles were written between 2015-2018, so ML have started to become more popular on recent years. (Baryannis, Validi, Dani and Antoniou 2018) Literature reviews of ML in SCM have been done few, to be mentioned Nguyen et al. (2018), Ni et al. (2019), Wang, Gunasekaran, Ngai and Papadopoulos (2016) and Wenzel, Smit and Sardesai (2019). Ni et al. (2019) found out that using ML in SCM is still in a developmental stage. Also, in the articles that Ni et al. (2019) found in their research, half of the cases were made with simulated data, so there should be more practical case studies.

1.2 Objectives and delimitations

This study will focus more on practical side of the ML applications in SCM area, as the literature reviews has been done. More practical research was needed. To achieve more knowledge about to what specific situations ML is appropriate, this study will concentrate on case studies. It will be examined what benefits ML brings to those problems. These problems are defined as following questions:

Research question 1. What SCM problems ML is suitable for?
Research question 2. What benefits are achieved with ML in SCM?

Restrictive factor in the thesis was amount of work. Case articles was handled one by one, and information was collected separately, so it restricted the amount of collected use cases. Getting familiarized with a new software takes time, so it was not possible to consider and compare other software. Also, only one case study is presented because of same reason.

1.3 Research approach

The thesis consists of two sections: literature review and empirical study. In literature review, use cases are looked from literature. Found use cases are summarized and results are considered. This will provide overview of how ML is already used. In empirical part, case study is done with DataRobot software. Example dataset is available on website. The purpose of this section is to find out how ML can be used in practice.

1.4 Structure of the thesis

The structure of the thesis is following: in Chapter 2 , theory of analytics, data and ML is studied. Chapter 3 concentrates on literature review of ML use cases in SCM. The case study is in Chapter 4. In Chapter 5 is discussion. Finally, conclusions are in Chapter 6.

2 DATA ANALYTICS IN SUPPLY CHAIN MANAGEMENT

Data mining "is a process of discovering interesting patterns and knowledge from large amounts of data". Knowledge process consists of data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge presentation, where data mining is "an essential process where intelligent methods are applied to extract data patterns". Data mining is one step in the knowledge process, but usually the term data mining is used to describe the entire process, so terms data mining and knowledge discovery from data are used as synonyms. Data mining can use many techniques and one of them is ML, which is studied in this thesis. (Han, Pei and Kamber 2011, p. 1-26)

2.1 Types of data analytics in supply chain management

Traditionally, data analytics is divided into three categories: descriptive, predictive and prescriptive analytics. Descriptive analytics describes what have happened in the past and why by explaining it with statistics and reports (Olson and Wu 2020, p. 5). The goal of descriptive analytics is to find current problems and root causes for later analysis (Zhao 2019). In the field of SCM, descriptive analytics is used to transform the data collected for example with global positioning system (GPS) and radio frequency identification (RFID) to understandable and value providing for the company or organization (Souza 2014). Visualizations and reports provided can be used to improve SCM processes, such as delivery schedules, replenishment orders, emergency orders and transportation modes (Souza 2014). Souza (2014) have listed that descriptive techniques in SCM are supply chain mapping and supply chain visualization.

Predictive analytics concentrates on what will happen in the future and why. Predictive analytics forecasts the future, but also does classification (Olson and Wu 2020, p. 5). Mathematical algorithms and programming are used to provide information from the past data to find the patterns that descriptive analytics cannot find (Wang et al. 2016). Following predictive analytics techniques are used in SCM (Souza 2014):

- Time series methods including
  - moving average,
  - exponential smoothing and

  - autoregressive models,
- Linear, non-linear and logistic regression,
- Data-mining techniques including
  - cluster analysis and
  - market basket analysis.

Obvious example of predictive analytics in SCM is demand forecasting.

Prescriptive analytics provides information for the question of what should be happening (Souza 2014). The purpose of prescriptive analytics is to provide suggestion of the best decision to the situation by comparing different decisions and their outcomes (Wang et al. 2016). Prescriptive analytics combines descriptive and predictive analytics with mathematical models (Souza 2014), such as multi-criteria decision making, optimization and simulation (Wang et al. 2016). In SCM, prescriptive analytics "can optimize the supply chains to balance the trade-offs between cost efficiency and customer service requirement" (Zhao 2019). Appropriate prescriptive techniques for SCM are analytic hierarchy process game theory, mixed-integer linear programming, non-linear programming, network flow algorithms and stochastic dynamic programming (Souza 2014).

ML belongs to predictive analytics, so this thesis concentrates on predictive analytics. Chakroun, Bouchti and Abbar (2018) researched articles of using big data analytics in warehouse management from years 2010-2018. Of 64 researched articles, only 3 articles included predictive analytics. Rest 93% was descriptive analytics. This confirms that more research on predictive analytics in SCM is needed.

Wang, Gunasekaran, Ngai and Papadopoulos (2016) have defined supply chain analytics (SCA) as big data business analytics within logistics and SCM. SCA is "tools and techniques that harness data from a wide range of internal and external sources to produce breakthrough insights that can help supply chains reduce costs and risk whilst improving operational agility and service quality" (Deloitte and MHI 2014). SCA is used to help for example in decision making and optimization in SCM.

2.2 Data in supply chain management

One model for supply chain strategic decision making is the supply chain operations reference (SCOR), which consists of four processes: plan, source, make and delivery (Huan, Sheoran and Wang 2004). Based on SCOR, Biswas and Sen (2016b) has developed data driven structure for supply chain (Fig. 1). Supply chain goes from left to right and demand conversely. Suppliers are under source process and provides supplier data and manufacturer is responsible for make process and provides manufacturing data. Deliver process is between warehouses, distributors/retailers and customers. Delivery, sales and customer data are received from deliver process. Biswas and Sen (2016a) has listed the information which can be collected from SCM. Those are presented in the Fig. 2.

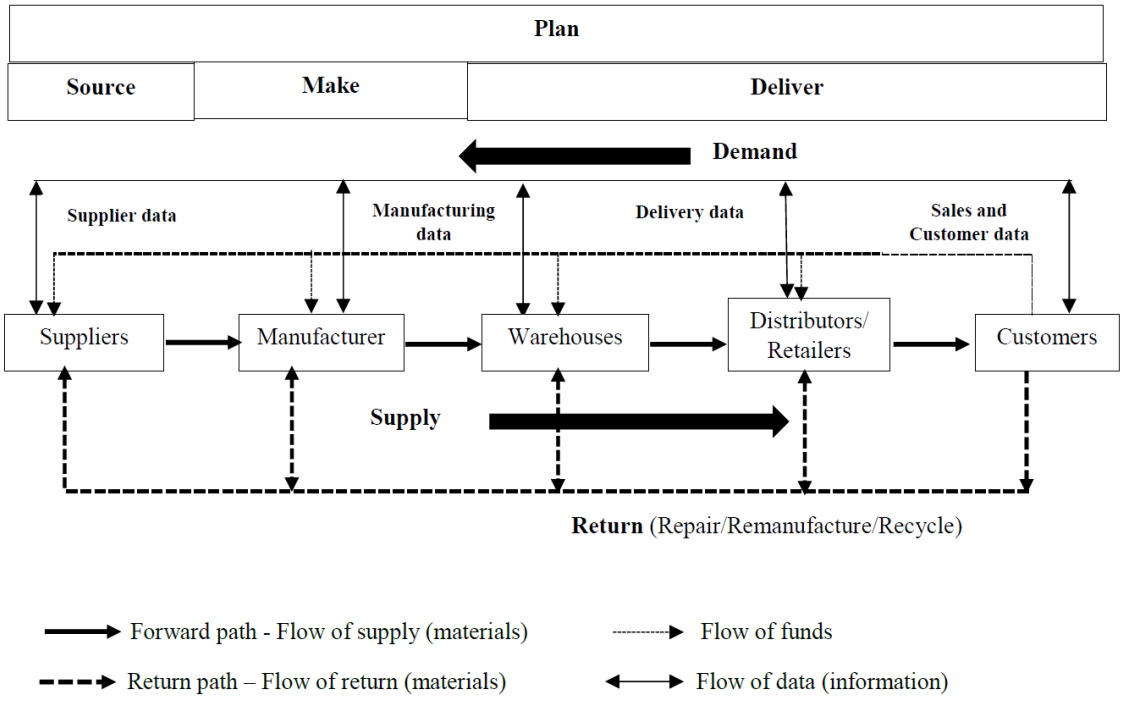


Figure 1. Data in supply chain (Biswas and Sen 2016b)

Big data is a buzzword these days. The definition of big data has sharpened in the 2010s (Power 2014). First was known 3 Vs of big data, which included Volume, Velocity and Variety. In 2013, Demchenko, Grosso, de Laat and Membrey (2013) extended the definition to 5 Vs of big data, where Value and Veracity were added. Volume describes quantity of the data with features size, scale, amount and dimension. The data is collected from records and transactions and the size of the data is expressed in terabytes. Google, Facebook and Twitter are taken as examples of producing vast amount of data. Velocity is

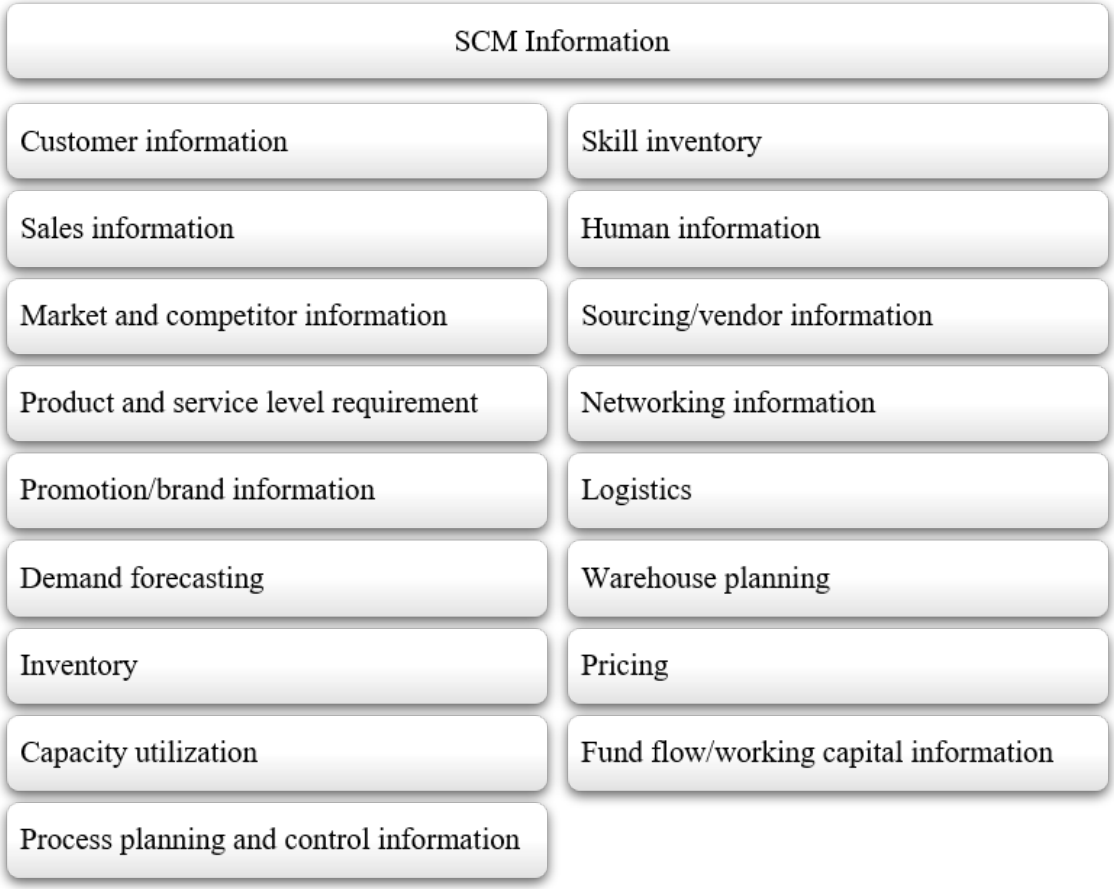


Figure 2. Information which can be collected from SCM (Biswas and Sen 2016a)

related to the speed of data, meaning that data is real-time as it is taken from sensors and events. Velocity needs be taken into account when processing the data, so that the data is processed also in real-time, or the data is taken in batches or visualized as streams. Variety means the complexity of the data. Data from different sources is different and also is structured differently: structured, unstructured, semi-structured and mixed. In addition to numbers, data can be different types, for example text, picture, sound and video (Ittmann 2015). Value is additional benefit that data provides. Value effects which data is needed and how it is collected and stored. Last characteristic is Veracity, which means consistency and trustworthiness of the data. Purpose of Veracity is that for the whole life cycle of the data, the data is trusted and timeliness. To meet this purpose, data have to be secured from collecting to stocking and using.

Biswas and Sen (2016b) has created the table (Table 1) how big data is used in SCM. The table is categorized with big data characteristics and SCM is divided into supplier, manufacturing, delivery, and sales and customer. In Volume section, many different data sources of SCM are listed. Data can be collected almost every action in whole supply

chain. Velocity adds time information to the data. In SCM real time, hourly, daily or monthly data is used in different cases. For example, logistics data is needed in real time, but customer satisfaction is analyzed in longer period. Variety presents different sources of data, such as sensors, E-mails, RFID and physical documents. Value describes different ways where data can be used to gain value. Those are for example planning of different situations, such as production, inventory and store. Veracity is same for all parts of SCM. Veracity is achieved so that data is from different sources and formats, and lack of reliability and noise in data is considered. The last in the table is added analytics, which describes in which analytics big data can be used. Common analytics are sentiment analytics, optimization and forecasting.

Table 1. Various Characteristics of big data in context of supply chain (Biswas and Sen 2016b)

Type of data	Supplier	Manufacturing	Delivery	Sales and Customer
Volume	More detail around design data for products, type of products, process, order, inventory, lot size, delivery, lead time, shipment and routing, pricing, tax, payment, return/ dispose.	Product design, customer requirement (e.g. specification, choice, demand, order, time of delivery, feedback), process metrics (e.g. throughput time, cycle time, % rejection, capability, reliability, maintenance), production planning and scheduling, inventory (e.g. lot size, order, WIP, scrap/disposal, finished goods, raw material), material storage, shipment and routing, vendor data (e.g. vendor list, purchase data, vendor evaluation, lead time), people data (e.g. skill inventory, training data, deployment details), finance data (e.g. wage, conversion cost)	Demand data (e.g. order, variety, forecasting), lead time, delivery schedule, location data, inventory (e.g. stock level, aging data), shipping and routing (e.g. mode of transport, load, network and path), finance data (e.g. pricing, exchange rate, tax, payment), miscellaneous (e.g. weather, social, economic, regional data), customer data (e.g. choice, feedback), manufacturing data (e.g. inventory status, production plan and schedule, product details), sales data (e.g. promotion, POS data), return/dispose	Point of Sales (POS) data, order status, demand data, customer data (e.g. product, quantity, delivery, lead time, sentiments, feedback, new product, profile, choice, purchase pattern), promotion, finance data (e.g. payment, pricing, discount, exchange), shipment and routing, return/dispose
Velocity	Hourly, daily, weekly, monthly, yearly	Hourly, daily, weekly, monthly, yearly	Real time, hourly, daily, weekly, monthly, yearly	Real time, hourly, daily, weekly, monthly, yearly
Variety	Various database, web, audio (verbal/telephonic), E-mail, physical document, sensor data, RFID data	Physical document, sensor data, RFID data, camera, various chips, web data, E-mail	Physical document, sensor data, RFID data, E-mail, various database, web, audio (verbal/telephonic)	Physical document, sensor data, RFID data, E-mail, various database, web, audio (verbal/telephonic)
Value	New product development, production planning and scheduling, Optimal lot size and inventory planning, shipping and routing, disposal/ recycle	Optimal lot size and inventory planning, product decision, process selection, execution and control, production planning and scheduling, supplier selection, optimizing delivery lead time, routing decision, remanufacture/recycle/disposal	Transportation and network planning, store planning, inventory planning, customer analysis	Predictive demand modelling, customer analysis, network planning, market basket planning, assortment planning, branding and promotion
Veracity	Multiple data sources, different formats, lack of reliability in some data sources, presence of noise in the network communication.			
Analytics	Association rule mining, optimization, network planning, logistics and supply chain planning, Sentiment analytics, stock planning	Optimization, operations research, assignment and schedule planning, new product development, inventory planning, distribution and warehouse planning, Sentiment analytics, forecasting, predictive demand modelling	Logistics and distribution planning, network planning, retailer selection, Sentiment analytics, forecasting	Sentiment analytics, market basket analysis, forecasting, product shelf layout planning

There are also problems in use of data. Increased amount of data includes much value, but also redundant information. One common problem is imbalanced data, where classification data has huge differences between class sizes. Minority class is difficult to predict. Imbalanced data is frequently occurring problem in real life situations. Thera are different methods which can be used to help with imbalanced data. Those methods are divided into three categories: “sample methods based on the modification of the training data, kernel methods based on the modification of the kernel and optimization methods based on the modification of the problem formulation”. (Li, Jiang, Yang and Wu 2018)

Hazen, Boone, Ezell and Jones-Farmer (2014) have defined four dimensions of data quality: accuracy, timeliness, consistency and completeness. Accuracy answers to the question if the data are free of errors. Accuracy can be tested by checking if the data matches with the known values. Example in supply chain is a customer shipping address in a customer management system, which can be verified with the most recent customer order and its address. Timeliness measures if the data are up-to-date. Timeliness can be divided to currency, which describes when the data is last updated, and volatility is the frequency of the data updates. In SCM inventory levels are monitored in real-time with inventory management. Consistency refers to the format and structure of data. For example, in SCM when entering information to systems, each part of the information is formatted in the same way, such as delivery date formatted as DD/MM/YY or name as "Surname, Forename". Completeness tells if there are necessary values missing in data. This not only refers to single values missing, but also all information missing of some subject. For example, customer shipping address is not complete if it includes street address, but not city. (Hazen et al. 2014)

3 MACHINE LEARNING IN SUPPLY CHAIN MANAGEMENT

In this chapter, the goal is to find out how ML is used in the field of SCM. To find which kind of problems it is used and what benefits it provides, a literature review was conducted. Existing case examples was looked from the literature.

3.1 Machine learning

Artificial intelligence (AI) has started to effect on everything during recent decades and the biggest steps have been in the 21st century (Li et al. 2018). ML is one of the methods of AI. The term machine learning was first used in the end of 1950s. Joshi (2020, p. 4) describes ML as "a computer program that can learn to produce a behavior that is not explicitly programmed by the author of the program". ML can be described as a combination of math, finance and computer science (Li et al. 2018). ML can use different algorithms to solve problems or ML finds patterns from data which helps to make prediction. ML is part of AI, because ML is intelligent as it can learn (Alpaydin 2014, p. 1-4). ML can provide models that are descriptive or predictive (Alpaydin 2014, p. 1-4).

ML methods are categorized to three categories based on how the learning is happening. Categories are supervised, unsupervised and reinforcement learning algorithms. In supervised learning, the historical data includes both input and output data. When creating a model, ML receives a feedback of prediction from labelled output data. Example of supervised learning is classification. (Joshi 2020, p. 10-11)

In unsupervised learning, data does not include output information. Unsupervised learning can create clusters from the data based on similarities. New data row will be added to the cluster which is closest to the data point. (Joshi 2020, p. 11; Rebala, Ravi and Churiwala 2019, p. 21) Unsupervised learning need more decision making from humans as the clusters must be analyzed and named by human.

The third category, reinforcement learning has components from both supervised and unsupervised learning. In reinforcement learning, data does not include label data to give feedback for machine about success, but it is getting feedback from environment. Example of reinforcement is game playing, where feedback is got from opponent's choice and situation is changing every time. (Joshi 2020, p. 11; Rebala et al. 2019, p. 21)

Ni, Xiao and Lim (2019) have found four reasons to use ML in SCM decision making situations:

- ML can find non-linear relationships,
- ML can use also unstructured data in modeling,
- ML techniques can find key factors, which effect on SCM performance and
- ML can help with visual pattern recognition in physical maintenance.

To success in using ML, the company has to adopt features that effect on succeeding. Understanding that AI is not developed to solve every problem is essential (Forger 2020). Ten characteristics of the companies that success are "agile, responsive, customer focused, technology savvy, data driven, collaborative, comfortable with experimentation, constantly challenging the status quo, adaptable to change, and ambidextrous" (Forger 2020). Taghizadeh (2017) mentions that "choosing the best tools and methods to analyze them is the key issue to consider".

Use of ML or other AI methods have some problems that should keep in mind. AI solutions are easily a "black box" models for the user, when the user does not know how AI have reached the result (Li et al. 2018). Another challenge is that scientific researches are source of innovation for companies, but it is not known how to use the results in practical situations.

3.2 Use cases

When searching articles, used query words were "machine learning", "supply chain management" and "inventory management". Because using of ML in SCM has became more common in recent years, only articles that were published after 2015 were considered. Only those articles were selected which included case which was described completely. Articles were summarized with the focus on situation and problem, ML solution and benefits. Also new title for cases was added for quick overview. They were gathered in a table which is in Appendix 1. 21 use cases were selected and summarized for this thesis but more articles could have been found.

Overall view of the articles is presented in Fig. 3, where titles are listed to give quick information of the cases. New and consistent titles are made up, which follow same

structure informing the field of SCM and describing the situation to which the case is related to. The content of the articles is studied in the following sections, including the field of the use cases, used ML algorithms and finally, what benefits ML gave for the cases. Each of the wanted information was not received of every case and classification of the cases is not clear and one case can belong to several categories, so calculated results are not exact in the following sections.

ML use cases

- Pricing and demand in online fashion store for first exposure items (Simchi-Levi and Wu 2017)
- Forecasting and price optimisation in e-commerce marketplace (Simchi-Levi and Wu 2017)
- Demand forecasting of weather sensitive products (Taghizadeh 2017
- Predicting customer behaviour (Martínez et al. 2020)
- Demand prediction by identifying comparable products in order to predict demand for the new product h. (Mišic and Perakis 2020)
- k-means clustering in international location decision s. (Khalid and Herbert-Hansen 2018)
- Location of delivery trucks of a “buy online, pickup in delivery truck” retailer (Mišic and Perakis 2020)
- Designing service regions for an electric vehicle sharing service (Mišic and Perakis 2020)
- Modeling logistic routing in car terminal (Becker et al. 2016)
- Two approaches for inventory management (Mišic and Perakis 2020)
- Estimating product quality d. (Ko et al. 2017)
- Supplier selection model with gene expression programming in textile industry (Fallahpouret al. 2017)
- Supply chain partner selection model for agricultural products (He et al. 2015)
- Ranking model for supplier evaluation and selection (Tavana et al. 2016)
- Predicting demand after transportation disruption (Liu et al. 2016)
- Predicting flight diversions to help reorganizing delivery (Ciccioet al. 2016)
- Classifying RFID readings to the moved tags and the static tags (Ma et al. 2018)
- Predicting the temperature of food in supply chain for food quality and safety (Mercier and Uysal 2018)
- Simulation model of supply chain ordering management (Mortazavi et al. 2015)
- Production and inventory control policy for complex production environment (Wu et al. 2015)
- Approach for dynamic procurement to decide amount of products to order when there are different sources with different costs and lead times (Mišic and Perakis 2020)

Figure 3. ML use cases

The use cases are divided into different categories by the field. Results are presented as a mind map in Fig. 4. Prediction is the most common need where ML is used. Mainly continuous demand is predicted, but predicting is also used to other situations, for example different exceptions such as transportation disruption (Liu et al. 2016). Demand prediction is used for predicting the demand of new products based on demand of old products or improving current demand prediction with new information such as weather or other changing features (Mišić and Perakis 2020). Demand prediction is also used to pricing when pricing depends on demand (Simchi-Levi and Wu 2017). Exception prediction is used to help correcting the problems caused by an exception. In transportation disruption, new transportation has to be planned, and predicting disruption gives more time for planning when the disruption is predicted earlier (Liu et al. 2016).

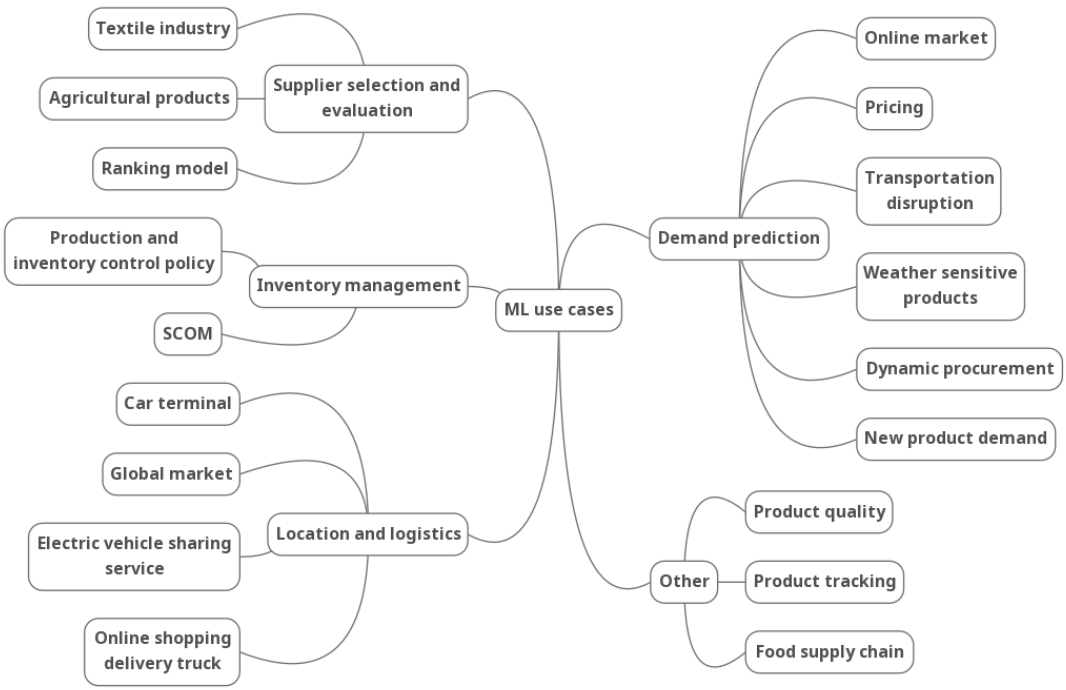


Figure 4. Categories of ML use cases

In SCM system, inventory control is covering almost 50% of the total costs, and in inventory control, supply chain ordering management (SCOM) is important. SCOM is "an integrated approach to determine the ordering policy of each supply chain actors", and decision makers try to decrease the costs of inventory and increase customer satisfaction. The simulation model of the system, including the retailer, the distributer, the manufacturer and the supplier, was made, and reinforcement learning algorithm was used. Different scenarios were used to solve out how they impact on supply chain, including total costs

and the average customers' waiting time. (Mortazavi, Khamseh and Azimi 2015) Efficient production and inventory control policy is needed in a complex production environment, where there are multiple working stations producing multiple intermediate components and end products. When producing different components and products, changeover times at work stations are causing costs and more time is needed. (Wu, Evans and Bae 2015) Besides these, different models to predict the amount to order to inventory are proposed. The models are based on empirical risk minimization , conditional demand distribution or ordering from different sources with different lead times and costs (Mišić and Perakis 2020).

Location and logistic decisions can be extremely complicated problems, so decisions cannot be made by human without calculated information. Location information is usually needed to decision making when new investments are wanted to make to new locations. In international location decision case, the model divided the countries into two clusters (Khalid and Herbert-Hansen 2018). Online shopping company uses ML to decide delivery locations predicting demand at a given location at a given time (Mišić and Perakis 2020). In a car terminal, where different treatments are made for cars, ML is needed to control and schedule car-flows so that treatments are done at the right time (Becker, Illigen, McKelvey, Hülsmann and Windt 2016). These shows that ML can be used in different ways within location and logistics problems.

Supplier selection and evaluation is one use case of ML. The problem in supplier selection and evaluation is personal subjectivity of decision maker. With ML it can be avoided. With ML, ranking model for supplier comparison is made. The idea of the ranking model is to select the wanted criteria to be compared and they will be evaluated for each supplier. The ML model evaluates the criteria, and then orders the suppliers. (Tavana, Fallahpour, Caprio and Santos-Arteaga 2016) The models are tested for textile industry (Fallahpour, Wong, Olugu and Musa 2017) and agricultural products (He, Ai, Jing and Liu 2015), but models can also be used to other ranking situations than supplier selection.

Other case examples are from different fields of SCM. In product tracking, ML is used to find false positive readings from RFID tags (Ma, Wang and Wang 2018). False positives occur for example when tag is too close of portal or read range is extended because of reflection from metallic objects. Problem is that tag is read but product is not loaded to truck, so incorrect invoices are sent. Product's quality management (Ko et al. 2017) is a key part of preventing company's reputation to declining and reducing warranty costs. A model was developed to estimate product quality using manufacturing, inspection and after-sales service data. In food supply chain (Mercier and Uysal 2018), temperature

of the products is important for food quality and safety. Temperature is measured with limited amount of temperature sensors. To receive information about the quality of supply chain, the researchers developed a model to predict temperatures using heat transfer model.

3.3 Machine learning algorithms in use cases

In this section, the used ML algorithms in use cases are considered in general. Used algorithms was not mentioned in each paper, and some papers had several methods. Though, different artificial neural networks (ANN) are the most used in these cases. ANN was used in seven cases. ANN is used in variety of different problems, for example demand prediction (Liu et al. 2016; Taghizadeh 2017), supplier selection and evaluation (He et al. 2015; Tavana et al. 2016), inventory management (Wu et al. 2015), logistics (Becker et al. 2016) and also in food supply chain (Mercier and Uysal 2018). The next common were support vector machine (SVM) and k-means clustering. Both were used in two papers. SVM was used in the RFID tag case (Ma et al. 2018) and predicting flight diversions (Ciccio, van der Aa, Cabanillas, Mendling and Prescher 2016), and k-means clustering in supplier selection (He et al. 2015) and international location decision case (Khalid and Herbert-Hansen 2018). Supplier selection and evaluation is the only field where especially one algorithm was used, as ANN was used in two out of three cases. Although, in one case gene expression programming was used instead of ANN, because it provides more visibility to model than ANN methods (Fallahpour et al. 2017).

Based on these findings, it is not possible to draw specific conclusions about what ML algorithms can be used in which situations in SCM. Algorithms need to be selected case by case, and different variables, such as amount of data or balance of classes, effect on which algorithm is suitable. The algorithm depends on whether the target is to classify, cluster or prediction. However, it can be said in general that the ANN is widely used and suitable for a wide variety of situations.

3.4 Benefits of machine learning models

The obvious benefits that are wanted in business is cost savings. In these cases, savings are received with good accuracies, over 80% (Ma et al. 2018; Martínez et al. 2020). The models are compared to similar models, and they received better results (Becker et al.

2016; Ma et al. 2018; Mišić and Perakis 2020; Taghizadeh 2017; Tavana et al. 2016). Results are also measured as growing of revenue, for example improvement of revenue for delivery truck company was 36% (Mišić and Perakis 2020). Demand grows when supply is offered at the right time and right place.

The product tracking model was more cost effective and had better performance than others (Ma et al. 2018). There was 12.7% improvement to throughput time in car terminal case (Becker et al. 2016). Flight diversions was predicted one hour before landing (Ciccio et al. 2016). Reorganizing trucks can be started earlier than before which can help preventing delays, costs and CO_2 emissions. Temperature model for food supply chain provided high accuracy average error being under 0.5K and gave more information and optimal location for sensors (Mercier and Uysal 2018). In the product quality model, the model found defective products with good performance (Ko et al. 2017). The model helps to reduce customer claim costs, and company's reputation is better, and customers are more satisfied.

When comparing to other models, the new model can be better on some areas but is not as good on the other area. In those situations, the entirety is decisive. Even though in production and inventory control policy model the needed time is longer and demand accuracy is not as good as in other models, the model reduced costs by 8% and intermediate components are produced more on time (Wu et al. 2015). Although the model is not better on every field, more important things can be emphasized. For example, if model is slower than other model, but it works fast enough in the situation, it can be used.

Besides predicting, ML is used to produce more information for decision makers. It helps in major investment decisions. The benefit of the simulation models is that different cases can be tested before execution. ML is cost effective method to test possible business plans. One important benefit is that there is not personal subjectivity in the results. Especially in supplier selection and evaluation, personal opinions can affect on decisions. Also, models can help to understand relationships between inputs (criteria) and outputs (performance) (Fallahpour et al. 2017). One of the supplier selection models provided better generalization ability than the model to compared (He et al. 2015).

When compared to other techniques, the benefit of ML is that it can be used for huge datasets and complex situations. Prediction models can be developed by adding new data and information to the model. For example, prediction model with weather data has better results than same model without weather data (Taghizadeh 2017). On the other hand, ML algorithms can help in situations when all information is not available (Tavana et al. 2016).

4 CASE STUDY WITH DATAROBOT

In this chapter, DataRobot platform is introduced. After that the case study is done using DataRobot.

4.1 DataRobot introduction

DataRobot is an end-to-end enterprise AI platform. Its purpose is to "unleash the full potential of human and machine intelligence". It offers automated tools to provide value from data. Automation is used to prepare, build, deploy and maintain the models. Also, user interfaces are AI-assisted, and feature engineering, model selection and tuning are automatic. DataRobot allows possibilities to do advanced regression, classification, time series and deep learning models. Data preparation is included in DataRobot. Data can be numerical, free-form text, image or geospatial data. DataRobot models are deployed, maintained and governed in the same place, and they can be connected to company's systems with application programming interface (API). (DataRobot 2021b)

DataRobot is developed so that it can be used by different user types. For example, DataRobot allows business analysts, executives and analytics leaders to work without data scientists, and it reduces the workload of software engineers and data scientists. DataRobot can be used in any industry, for example there are use cases for healthcare, manufacturing, financial services and automotive industries on DataRobot's website. It is possible to use DataRobot for example for predictions, optimize outcomes and support critical decisions. To help using the platform, DataRobot offers platform documentation, enterprise support, DataRobot university and community, which includes learning videos and possibility to ask questions. (DataRobot 2021b)

4.2 Case study problem

One of the problems in supply chain management is availability. The problem occurs when a consumer orders a product, but a vendor does not have it in the inventory at the moment, so it is out of stock. It is called product backorder if the consumer decides to wait until the product is again available. Reasons for product backorder are for example unusual demand, low safety stock or manufacturer or supplier problem. If the product backorder occurs and the company cannot handle it well enough in satisfying time, obvi-

ous consequence could be losing a customer or sales order, but also product backorder will affect negatively on company's revenue, share market price and customers' trust. (Kenton 2019; Lopienski 2019)

4.3 Data

The dataset is from Kaggle's data science competition (Kishore 2018). The dataset contains of training and testing data. Training data has 1 687 861 rows and testing data 242075 rows. Each row represents one product identified by *sku*, which is a random identification number for the product. The dataset includes 23 columns including the target column, which is a binomial class variable telling if the product went on backorder or not. The dataset is extremely imbalanced, because only 0.67% of products went on backorder. Imbalance of classes does it hard to predict. Besides identification and target, the dataset has information of products' inventories, demand forecast, actual sales and part risk flags. All the variables are introduced in Table 2.

The training data is added to DataRobot. DataRobot lists variables based on their importance in prediction, in other words "strength of relationship to target". The most important variables are all forecasts in the data which are for 3, 6, and 9 months. Each of the forecast variables has over 49% of the value of zero, meaning that there is not predicted demand for those products. The fourth most important feature is *national_inv*, which describes current inventory level of the product. Over 3000 products have negative inventory level and for 86 579 it is zero. It is not known what negative inventory level means, but it may occur because of human or machine error, or when a shipment is registered complete even when its yet to arrive (Islam and Amin 2020). Because there are so many negative values, it is probably because of the latter reason due to the marking policies. Histogram (Fig. 5) shows that when the *national_inv* is negative, proportion of product's that went on backorder is significantly higher being over 19% when *national_inv* is between -112 and -32. Otherwise, the proportion of backorders is under 1%.

Next are sales variables. Sales quantities are given for previous 1, 3, 6 and 9 months. DataRobot sorted *sales_3_month* as the most important of the sales variables, then 6 and 9 months, and sales for 1 month was the least important. Similarly to forecast variables, zero values also appear in sales variables. Sales for 9 months has the least zeros, 28% of products, but for 1 month there are even 45% zeros which means that 45% of product were not sold at all in previous month. Both forecasts' and sales' quantities vary between few and millions products, but small amounts, under 10 pieces, are more common than

Table 2. Features in the dataset and explanations

Feature name	**Feature description**	**Feature type**
sku	Random ID for the product	Numeric
national_inv	Current inventory level for the part	Numeric
lead_time	Transit time for product (if available)	Numeric
in_transit_qty	Amount of product in transit from source	Numeric
forecast_3_month	Forecast sales for the next 3 months	Numeric
forecast_6_month	Forecast sales for the next 6 months	Numeric
forecast_9_month	Forecast sales for the next 9 months	Numeric
sales_1_month	Sales quantity for the prior 1 month time period	Numeric
sales_3_month	Sales quantity for the prior 3 month time period	Numeric
sales_6_month	Sales quantity for the prior 6 month time period	Numeric
sales_9_month	Sales quantity for the prior 9 month time period	Numeric
min_bank	Minimum recommend amount to stock	Numeric
potential_issue	Source issue for part identified	Categorical
pieces_past_due	Parts overdue from source	Numeric
perf_6_month_avg	Source performance for prior 6 month period	Numeric
perf_12_month_avg	Source performance for prior 12 month period	Numeric
local_bo_qty	Amount of stock orders overdue	Numeric
deck_risk	Part risk flag	Categorical
oe_constraint	Part risk flag	Categorical
ppap_risk	Part risk flag	Categorical
stop_auto_buy	Part risk flag	Categorical
rev_stop	Part risk flag	Categorical
went_on_backorder	Product actually went on backorder. This is the target value.	Categorical

huge amounts.

For the next most important variable, DataRobot have found *sku* which is a random ID for the product so it should not have impact on prediction. *Sku*s have a gap on ID numbers between 2.4 and 2.7 million. DataRobot have found that when the *sku* is smaller, the proportion of products that went on backorder is higher. Though, the different is not big, as percentage of went on backorder varies between 0.1% and 1.1%. It is possible, that higher *sku* means newer product, but without more information, it is thought as random number. For further studying, *sku* should not be taken to prediction dataset.

After that there are different additional variables of products, such as transit time and amount of product in transit. The least important variables are part risk flags. *Perf_6_month_avg* and *perf_12_month_avg* are source performance for prior 6 and 12 month

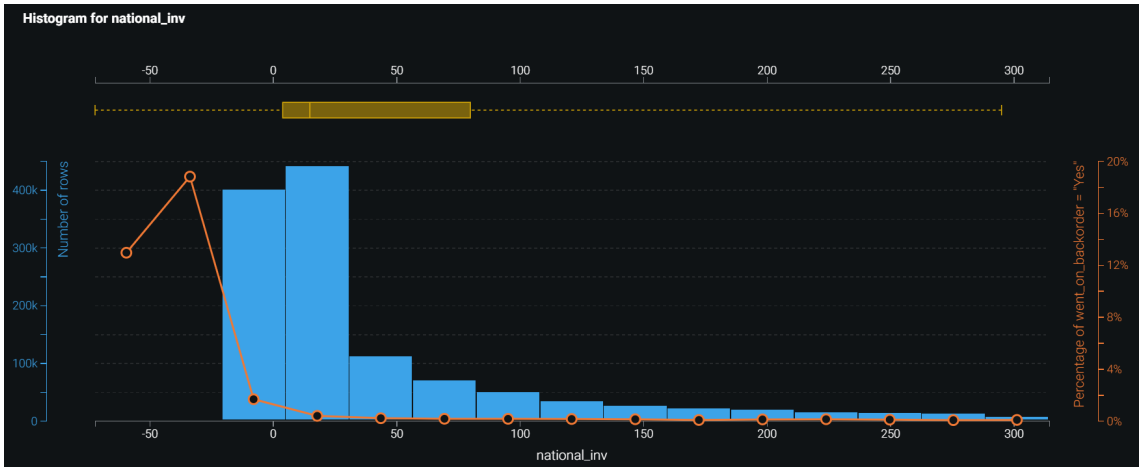


Figure 5. Histogram for inventory level

period. Range is from 0 to 1 but there are multiple -99 values, which means they are not evaluated. There are more higher values, and then the probability for backorder gets smaller when the source performance value grows. *Lead_time* describes how long is the transit time for the product. Lead time is between 0 and 52 so it is reported as weeks. Median for the transit time is 8 weeks. At maximum, proportion of backorder is 1.5% when the transit time is 0 to 1 week.

Amount of stock orders overdue is described with variable *local_bo_qty*. Mostly the value is zero and then the backorder percent is 1%. If there are stock orders overdue, the proportion of went on backorder is between 4% and 10%. *Min_bank* is minimum recommend amount to stock, and for 41% of products it is zero and the backorder proportion is under 1%. Otherwise, the backorder proportion is maximum being 1.2% when the recommended amount is 8-12. Any conclusions cannot be made with this variable. *In_transit_qty* describes how many quantities of product is in transit from source. 31% of the products are in transit, but it does not impact on the backorder as the backorder proportion is under 1% regardless if the product is in transit or not. *Pieces_past_due* means parts overdue from source and 21% of products are overdue. If a product is not overdue, it probably does not go on backorder as the probability is 0.6%. However, when there are pieces overdue, the probability for the backorder is higher, being at maximum over 9%.

The rest of the variables are part risk flags that are categorized as “Yes” and “No”. *Deck_risk* and *ppap_risk* are risks in 23% and 12% of the products, and *stop_auto_buy* is “Yes” even in 96% of the products. In each of these, the proportion of products that went on backorder are under 1%, both “Yes” and “No” categories. *potential_issue*, *oe_constraint* and *rev_stop* have only few “Yes” answers, so any decisions cannot be

made from those.

4.4 Modeling

Modeling with DataRobot was done with autopilot mode where it uses predefined models, which are the best to predict the target. Informative Features feature list created by DataRobot was used, where *potential_issue*, *oe_constraint* and *rev_stop* were left out. With these settings, DataRobot generated 59 models from the data. Modeling took time about 4 hours because the dataset is huge.

With the learning curves in Fig. 6, models can be compared. If the accuracy declines, DataRobot does not continue modeling it. In the diagram, models have split to two groups based on their accuracy. The lower group of models has higher accuracy, and the improvement of accuracies is higher when the used amount of data is bigger. Better accuracies are received with different gradient boosted trees classifiers and blender models. Random forest classifiers gave worse results. Based on this, gradient boosted trees classifiers and blender models are considered further. Speed vs accuracy comparison in DataRobot shows that gradient boosted trees classifiers predicts faster than blender models, but all of the models are fast enough.

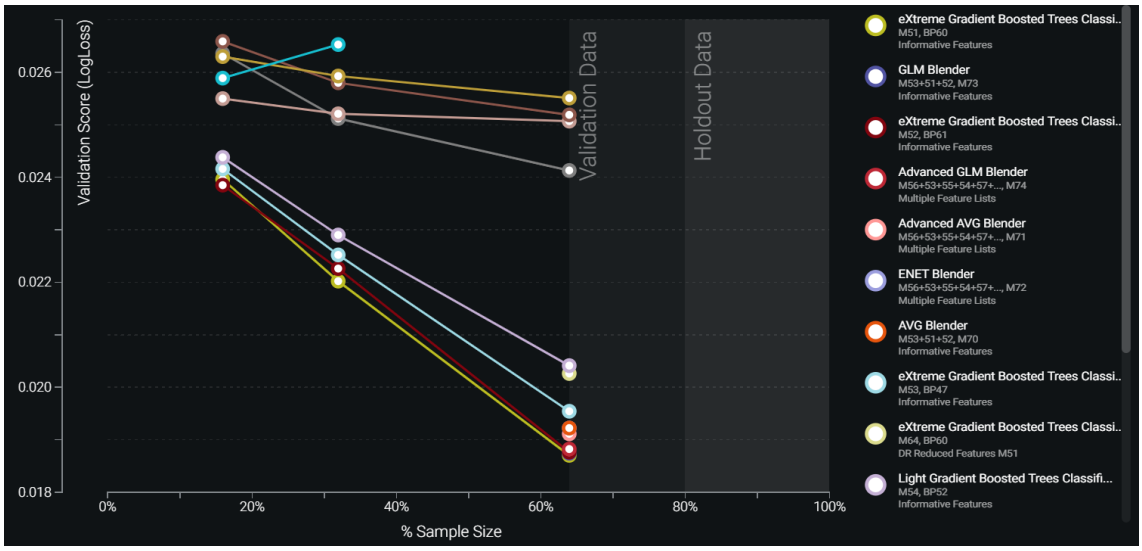
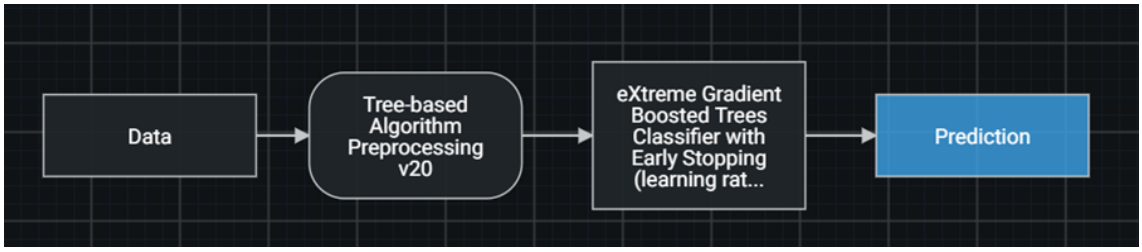


Figure 6. Learning curves figure for the created models

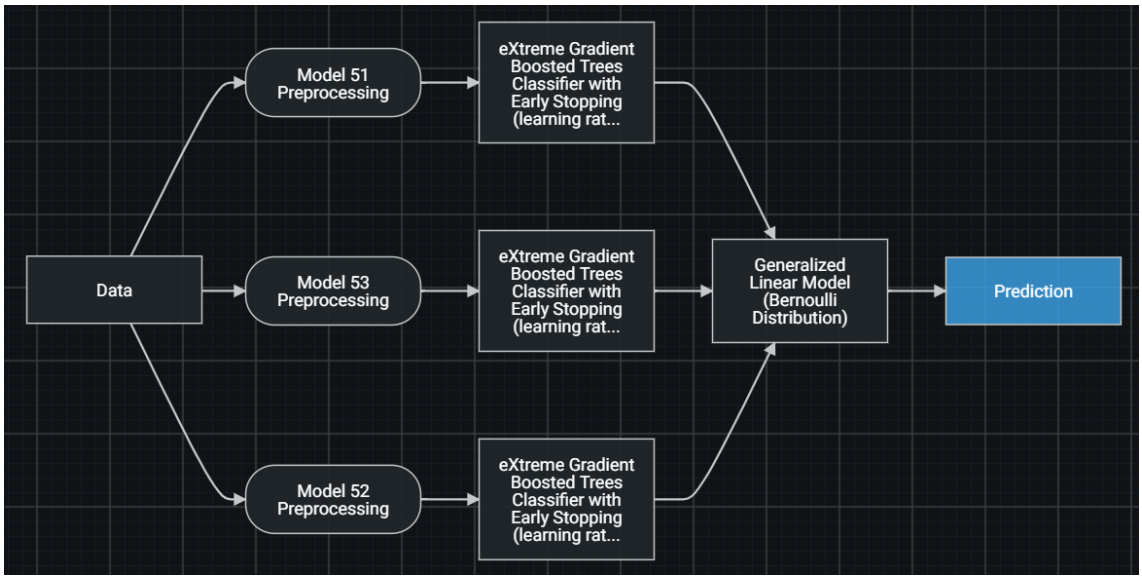
On the leaderboard tab, the models are listed based on accuracy metric, which DataRobot has selected to be logarithmic loss, which “measures the inaccuracy of predicted proba-

bilities" and it can be used in binary classification and multiclass predictions (DataRobot 2021c). The model "eXtreme Gradient Boosted Trees Classifier with Early Stopping (learning rate =0.08) (Fast Feature Binning)" (later EGBTC) is marked with badges "Recommended for deployment", which means that DataRobot recommends that model for deployment based on accuracy and complexity, and "Prepared for deployment" meaning that DataRobot has already released cross validation and holdout data.

The second best logloss value is received with GLM Blender algorithm (GLM). For the best model logloss value of validation data is 0.0181 and for GLM 0.0187. Because DataRobot predicts automatically, actual process and model may leave unknown for the user. That is called a black box. DataRobot tries to prevent the black box effect by showing the blueprint of the model. ML models may be hard to understand but the blueprint shows in the simplest way what the model does. In this case for EGBTC, the blueprint (Fig. 7a) is simple including only preprocessing and actual ML algorithm in addition to data and prediction, which are in every blueprint. The reason why the model is that simple, is that there are not many types of features in the data, which would need different processing, for example text variables. GLM is a combination of three models (Fig. 7b).



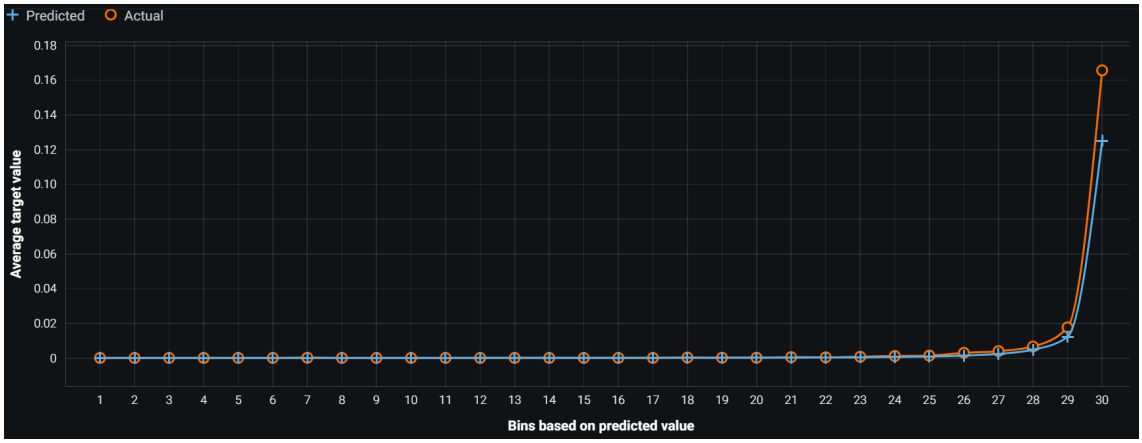
(a)



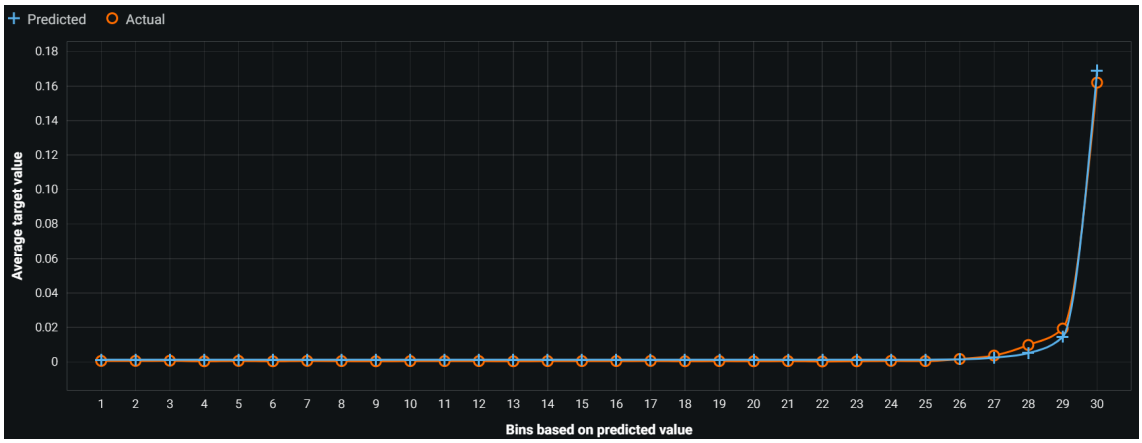
(b)

Figure 7. Blueprints of the models: (a) EGBTC; (b) GLM

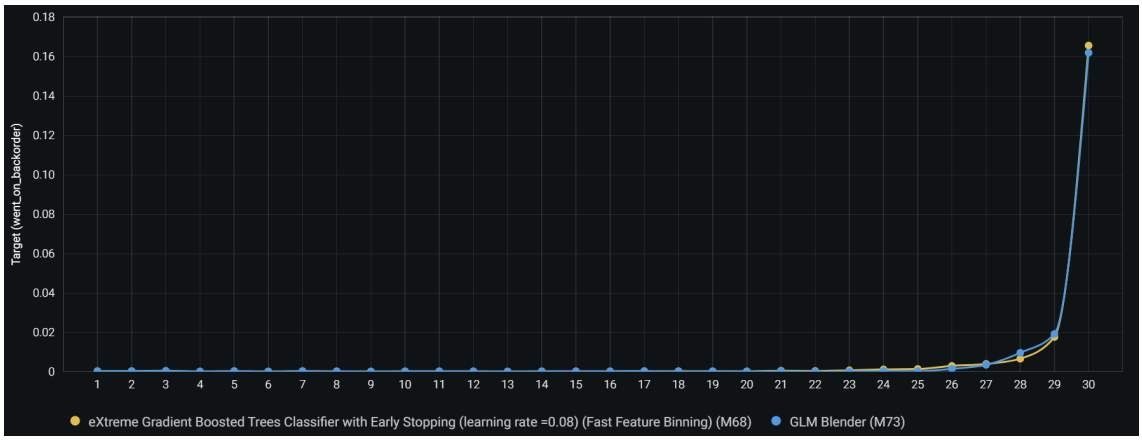
In a lift chart, data has been shared to bins and sorted from lowest to highest risk which means in this binary classification case from “No” to “Yes” results. Bins are created by ordering predictions in increasing order and then divided to equal sized groups. On the vertical axis is the average target value in one bin. The lift charts for the models are presented in Fig. 8a and 8b. Lift charts shows that the models can predict extremely well if the product do not go on backorder, but they do not recognize every product that went on backorder because actual line is above the predicted line at the right end of the chart. GLM model predicts little higher predictions on the right end. Predictions of the first model is under the actual and for the GLM otherwise, so EGBTC underestimates and GLM overestimates. In Fig. 8c, lift chart comparing of the two models is presented. There is a small variation on the right, but neither of the models is not significantly better.



(a)



(b)

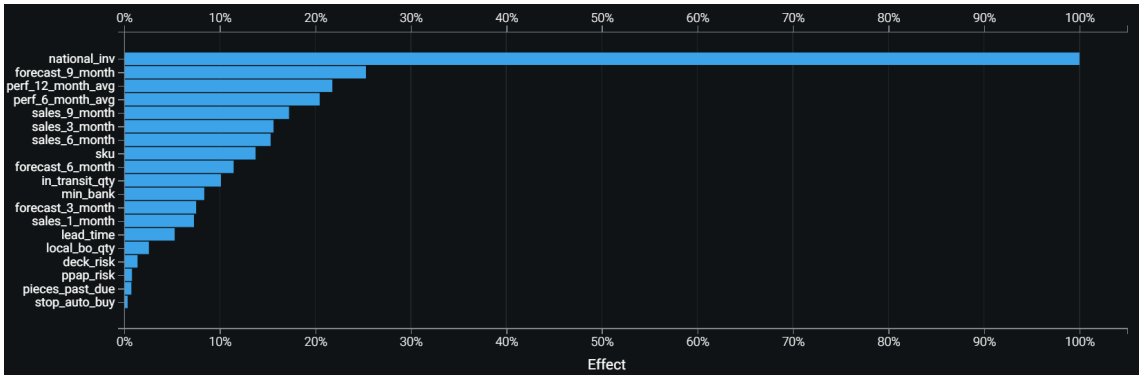


(c)

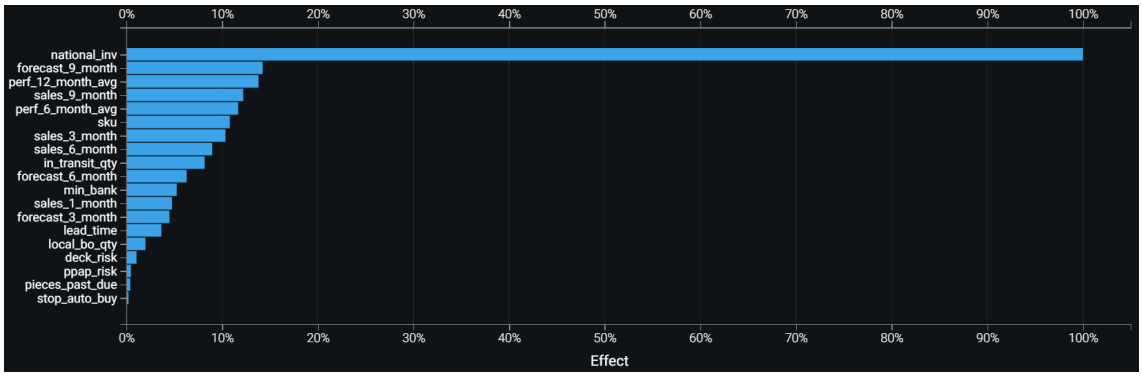
Figure 8. Lift charts for the models: (a) EGBTC; (b) GLM; (c) Both models

Feature impact chart describes which features are the most important in prediction model. Effect is proportional to the most important feature which is marked as 100% effect. Feature impact can be used for example to select the most suitable features to show in dashboard. Also, from the feature impact can be checked if the important features are reasonable. If not, should consider if the model is trustworthy. For both models, *national_inv* is the most effective feature (Fig. 9). Also, the next two features are same, *forecast_9_month* and *perf_12_month_avg*, but the effectiveness is smaller for GLM where all features are under 15% effective, comparing to EGBTC where the second important feature has 25% effect. Neither of the models do not have significant drop on the effects. Instead of it, effects are constantly descendent, so limit cannot be made to mark the most important features. It is reasonable that inventory level is important feature to detect if the product will go on backorder or not. In both models', part risk flags and transit time are the least important and do not have impact on prediction. Identification number for the product has over 10% effect in both models which is not reasonable as the identification number is given randomly for the products. It is surprising that source performance variables have such big effect on models because they were not important in feature importance, which describes how the variable correlates with target variable. Rest of the features do not have significant differences in order between the models. Overall, based on feature impact variables used in models should be considered more closely and models should be tested with another variable combinations. At least identification number need to be left out, and source performance and part risk flag variables should be considered.

DataRobot's feature effect graph shows how the model uses variables to predict the probability of product going on backorder. When comparing the feature effects between EGBTC and GLM models, there is not big differences. The shapes of the graphs are same, but overall GLM Blender model has slightly higher probabilities, even though all probabilities are very small. The feature effect graph for *national_inv* (Fig. 10) shows that probability of product going on backorder is higher when the inventory level is smaller. When the inventory level is over hundreds, the probability is almost zero. The shape of the graph is same for both models, but GLM has highest probability 0.065, when for EGBTC it is 0.045. The results are reasonable as inventory gets empty easier when the inventory level is lower. Sales for 3 months shows that probability of product going on backorder is highest when the sales is 500. This may mean that if the sales were higher than expected, it would run out easier. There is not reasonable reason why probability do not grow smoothly between 0 and 100 pieces. Only sales for 3 months have ascending trend, and sales for 6 and 9 months have descending trend. For 6 and 9 month sales the probability is higher when the sales amount is lower. It needs to be noticed, that there are most products predicted with small amount of sales, so there is more data of them for



(a)



(b)

Figure 9. Feature impacts for the models: (a) EGBTC; (b) GLM

the model to make assumptions. Each of the forecast variables have irregular graphs, and any conclusions cannot be made from those. Performance variables were found out to be effective in the models. From the feature effect graphs any conclusions cannot be made, because probabilities vary irregularly on the performance range. The most distinct graphs were *in_transit_qty* and *lead_time*. The graph of *in_transit_qty* is exponential decay. The probability is higher when there are 0 to 10 pieces in transit, and otherwise the probability is near zero. Need to notice that probabilities are still very small, maximum being under 0.009. Transit time has same highest probability as *in_transit_qty*, but the graph is linearly descending. To summarize, feature effect did not give new information about which events effect if the product will go on backorder or not.

Feature fit graphs shows how the model works on each variable. For the both models and each variables, feature fits are very exact. For example, feature fit for *forecast_9_month* is shown in Fig. 11. It can be seen that predicted values follow actual values well detecting lows and highs, also large changes starting from 1800 pieces. The y-axis is from 0.002 to 0.018, so the differences are not big. It is good for the model that predicted and actual lines cross each other multiple times. But the problem is the exactness of the prediction, because the model should not follow actual values so in detail. The predicted line should

Figure 10. Feature effect on EGBTC model for *national_inv*

be smoother, but now the model is too accurate, which is seen also in feature fit.

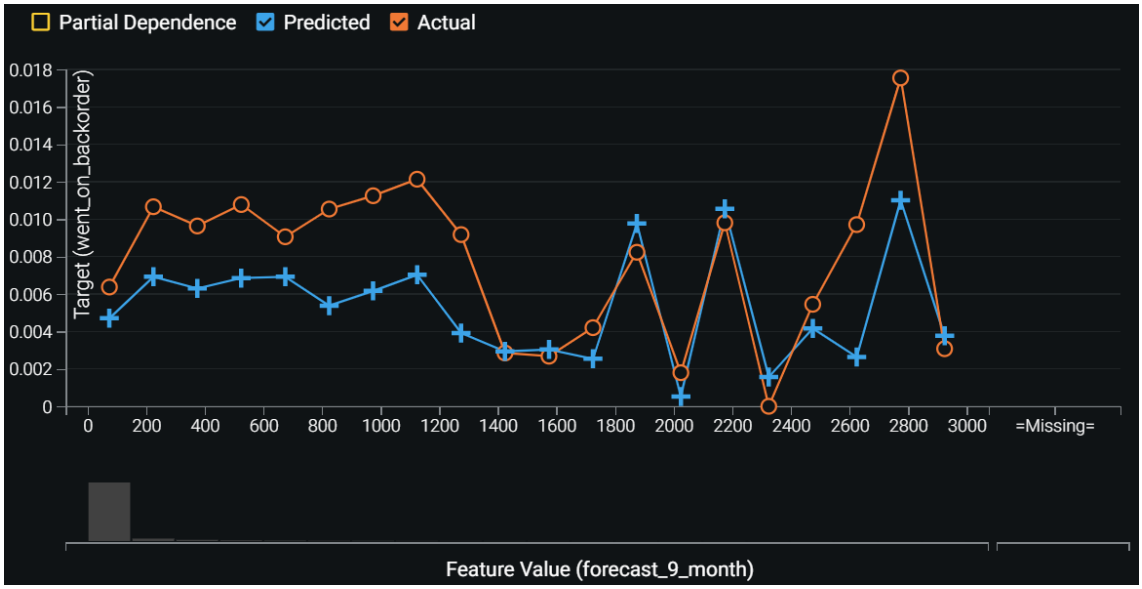


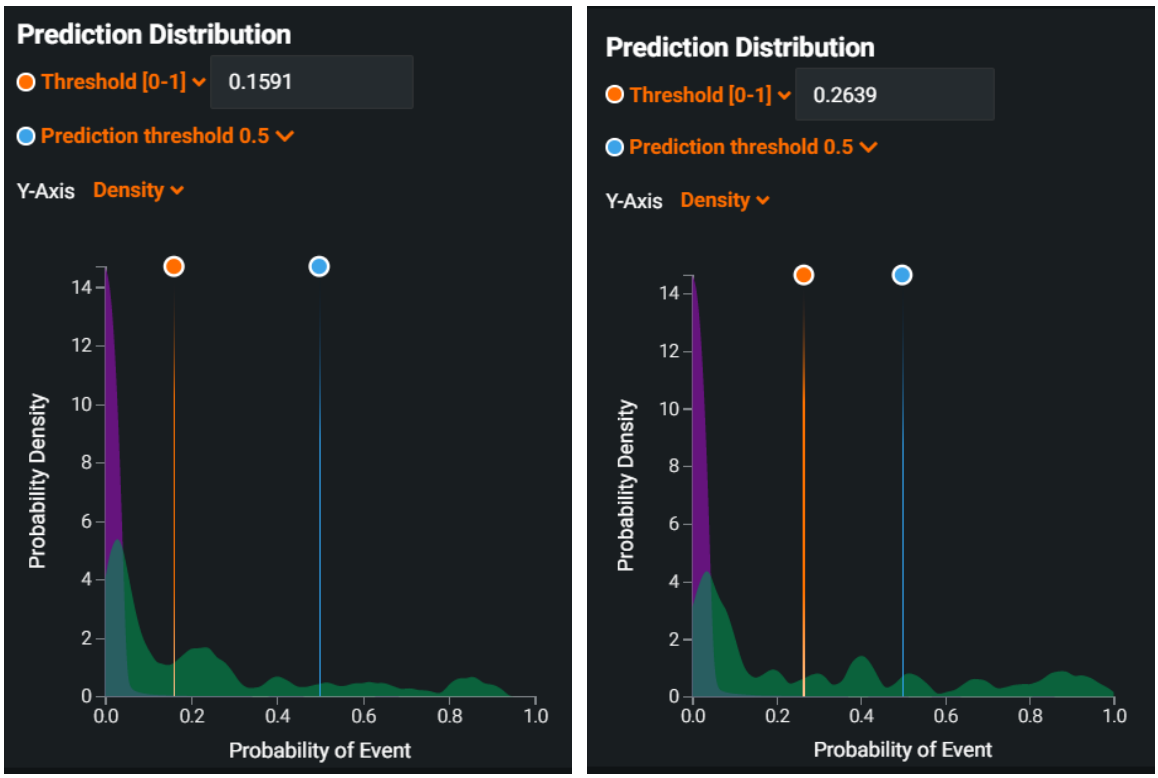
Figure 11. Feature fit on EGBTC model for *forecast_9_month*

Receiver operating characteristic (ROC) curve, prediction distribution, cumulative charts, confusion matrix and selection summary can be found on the ROC curve tab. On the prediction distribution, threshold can be changed, and it will change the results on the other charts.

The prediction distribution chart (Fig. 12) shows how the model has predicted and how the results have distributed. Datarobot can display density or frequency on the Y-axis. In this imbalanced case, density is better to evaluate the results. On the X-axis is probability of the event. Products that went on backorder are presented on purple color, and green color is for “No” answers. Threshold defines how predictions are read. Purple prediction

on the left side of the threshold is true negative and on the right side, it is false negative. Similarly, green prediction on the left side of the threshold is false positive and on the right true positive. Threshold is the line which defines how sensitively the prediction is decided to be “Yes” or “No”.

In the prediction distribution in Fig. 12, purple area, meaning negative predictions, is clearly focused on range 0.0-0.1. Green area is overall of the probability range. If it would be easy for the model to separate the “Yes” and “No” answers, the green area would be higher at the right end of the probability range. Now the highest probability density is on probability range from 0 to 0.1. This means that in that area it is difficult for the model to predict results.

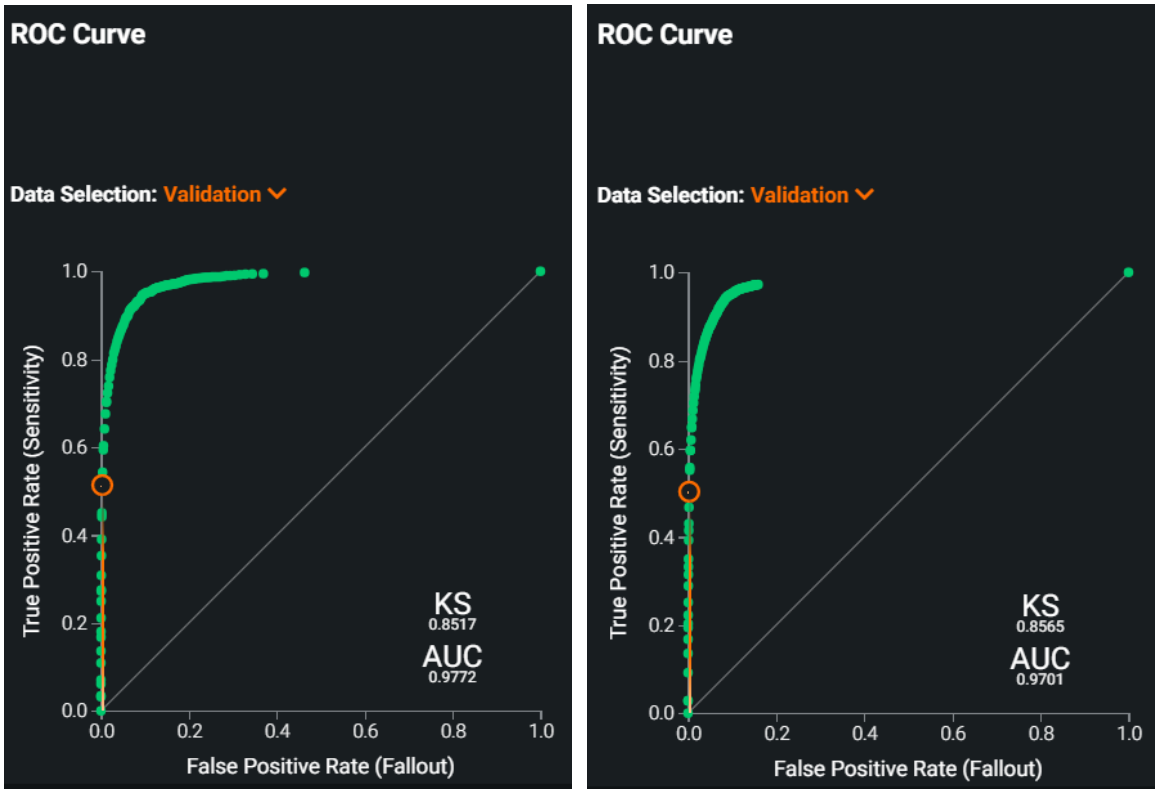


(a) (b)

Figure 12. Prediction distribution with original threshold: (a) EGBTC; (b) GLM

ROC curve can be used to compare the true positive rate and the false positive rate. The perfect situation, when there are not false positives, is when the curve touches top left corner. Area under the curve (AUC) describes summarized performance on every threshold. AUC value of 1.0 is the perfect situation when there are not wrong predictions. (DataRobot 2021a) ROC curve and AUC for the models are presented in the Fig. 13. The ROC curve curves near the top left corner in both models and AUC is over 0.97. Based

on the ROC curve and AUC, both of the models are too accurate.

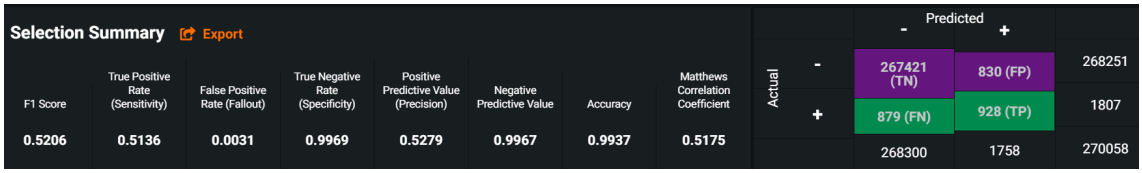


(a) (b)

Figure 13. ROC curves of the models: (a) EGBTC; (b) GLM

Selection summary (Fig. 14) shows confusion matrix and metrics calculated from confusion matrix. Values are updated as the threshold is changed. Accuracy is almost perfect in both models, which occurs because of the imbalance of the data. Huge number of true negatives compared to other values in confusion matrix causes high accuracy. Also, true negative rate is high. The models do not have troubles predicting negatives. Accuracy and true negative rate are not good metrics in this case. The target is to find the products that went on backorder, so it is important to predict true positives well. Metric for that is true positive rate (TPR), which calculates the portion of true positives to all positives (true positives and false negatives). Both models have TPR round 50% which means that the models find only half of the products that will go on backorder. Matthews correlation coefficient (MCC) is a metric that can be used for imbalanced datasets. Differing from other metrics, where range is from 0 to 1, the range of the MCC is from -1 to 1, where 1 is the best, 0 means halfway so that value should be over 0 so that the model works better than guessing (Chicco, Tötsch and Jurman 2021). Negative result means that prediction and actual are not in relation. Both models have MCC value 0.52.

Overall problem of this case is seen from the confusion matrix. Amount of true negatives is extremely big compared to others. The models know how to predict negatives so which products do not go on backorder. But more important is to predict the products that go on backorder, but it is not as easy for the models.

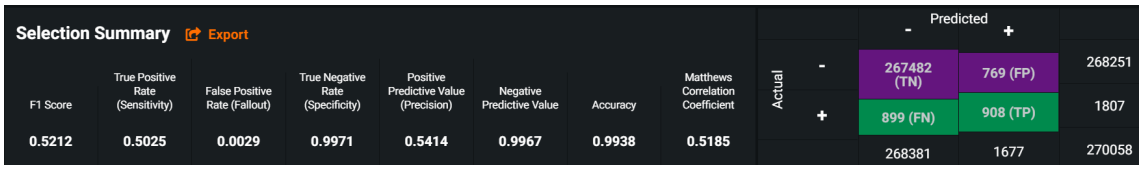


Selection Summary Export

F1 Score	True Positive Rate (Sensitivity)	False Positive Rate (Fallout)	True Negative Rate (Specificity)	Positive Predictive Value (Precision)	Negative Predictive Value	Accuracy	Matthews Correlation Coefficient
0.5206	0.5136	0.0031	0.9969	0.5279	0.9967	0.9937	0.5175

Actual \ Predicted	-	+	
-	267421 (TN)	830 (FP)	268251
+	879 (FN)	928 (TP)	1807
	268300	1758	270058

(a)



Selection Summary Export

F1 Score	True Positive Rate (Sensitivity)	False Positive Rate (Fallout)	True Negative Rate (Specificity)	Positive Predictive Value (Precision)	Negative Predictive Value	Accuracy	Matthews Correlation Coefficient
0.5212	0.5025	0.0029	0.9971	0.5414	0.9967	0.9938	0.5185

Actual \ Predicted	-	+	
-	267482 (TN)	769 (FP)	268251
+	899 (FN)	908 (TP)	1807
	268381	1677	270058

(b)

Figure 14. Selection summary for the models: (a) EGBTC; (b) GLM

4.5 Improving the model

Big differences were not found from the two models when compared. The models were not good enough, so better model should be found. DataRobot includes many possibilities to develop the models. There were 22 features in the data set, from which 19 were selected to Informative Features feature list. It was found out that not many features have significant impact on models. Next, different feature combinations are tested for EGBTC model, because GLM Blender cannot be retrained with new feature list.

Random number identification *sku* had too much impact for the models, so the model will be retrained without it. Inventory level was the most important feature in both models. Sales and forecast features have had more impact than for example part risk flags. Source performance features had impact for EGBTC model. Based on feature impact, other features are not needed for the model. Different combinations of the features are presented in Table 3.

The performance of the new models is quickly compared with TPR and MCC because they are useful in this case. The results are presented in Table 3. The best results are received with Features 4 list, which included all features which were taken to retraining.

The retraining with new feature lists did not give better results than original Informative Features list. Also, more features are needed for more accurate results.

Table 3. Results with different feature lists

Feature list	**Features**	**TPR**	**MCC**
Informative Features	All except *potential_issue*, *oe_constraint* and *rev_stop*	0.5136	0.5175
Features 1	*national_inv, forecast_9_month, sales_9_month, perf_6_month_avg, perf_12_month_avg*	0.311	0.2382
Features 2	*national_inv, forecast_3_month forecast_6_month, forecast_9_month sales_3_month, sales_6_month sales_9_month*	0.3282	0.273
Features 3	*national_inv, forecast_9_month, sales_9_month*	0.3116	0.2263
Features 4	*national_inv, forecast_3_month forecast_6_month, forecast_9_month sales_3_month, sales_6_month sales_9_month, perf_6_month_avg, perf_12_month_avg*	0.4261	0.4134

The biggest problem of this prediction is extremely imbalanced dataset. DataRobot offers tool to deal with imbalanced data. Before modeling, in advanced options section can be found smart downsampling, which "downsamples the majority class for faster model build time" (DataRobot 2021d). In this case, smart downsampling can be exploited to even the huge gap between false and positive classes. In smart downsampling, minority class stays same, but majority class will be reduced. Modeling will be made with quick modeling mode, and with 75%, 50%, 25%, 10%, 5%, 2% and 1% size of majority class. Comparison is again made with TPR and MCC. The results are looked from the first model based on weighted logloss by DataRobot, where DataRobot have already released holdout data. Results are presented in Table 4. The TPR varies between 0.31 and 0.41, and MCC is at maximum 0.43, so any of these is not better than the original model with the whole dataset.

Table 4. Results of smart downsampling

Majority class size (%)	TPR	MCC
75	0.37566	0.3719
50	0.3158	0.3473
25	0.4145	0.4304
10	0.3756	0.3324
5	0.4066	0.3042
2	0.3365	0.3104
1	0.3149	0.3294

4.6 Result evaluation

DataRobot has given threshold 0.1591 for EGBTC model and 0.2639 for GLM. With threshold, model can be specified for own needs. When predicting backorder, the target is to find products that would go on backorder and prevent it. Each action has a price. Presumably it is more expensive to correct the problems caused by backorder than to prevent product to go on backorder. This is the reason why it does not matter if the amount of false positives, that is the amount of products that were predicted to go on backorder but actually they do not go on backorder, grows if in that way it is possible to avoid false negatives that are the products that predicted to not go on backorder but they went on backorder. The threshold will be set to an appropriate point. To see how changing threshold effects to the results of the models, results of different thresholds are tabulated in Tables 5 and 6. Again, TPR and MCC are used as metrics. When threshold gets smaller, the TPR grows, but MCC declines, which was expected as negatives are not predicted as well. If the threshold is very small, the amount of false positives grows significantly, but false negatives are minimized. Then can be found even 88% of products that will go on backorder. To find the optimal point for the threshold, profit curve is examined next.

Datarobot offers profit curve, where various scenarios can be considered. The user can test how different inputs affect the end result. In the backorder case, the goal is to minimize costs, so all values are negatives and there are not profits in the model. Values are defined for each situation in the confusion matrix.

Possible costs are inventory replenishment cost, overstocking cost and backorder cost. Inventory replenishment is additional order is made beforehand to fill the inventory and to prevent backorder. The overstocking cost means that inventory replenishment is done but sale have been less than predicted when the inventory level is higher than optimal.

Table 5. Results with different threshold in EGBTC model

Threshold	**TPR**	**MCC**	**TN**	**FP**
			FN	**TP**
0.1591	0.5136	0.5175	267 421	830
			879	928
0.1	0.5938	0.5006	266 820	1 431
			734	1 073
0.01	0.8677	0.3013	255 758	12 493
			239	1 568

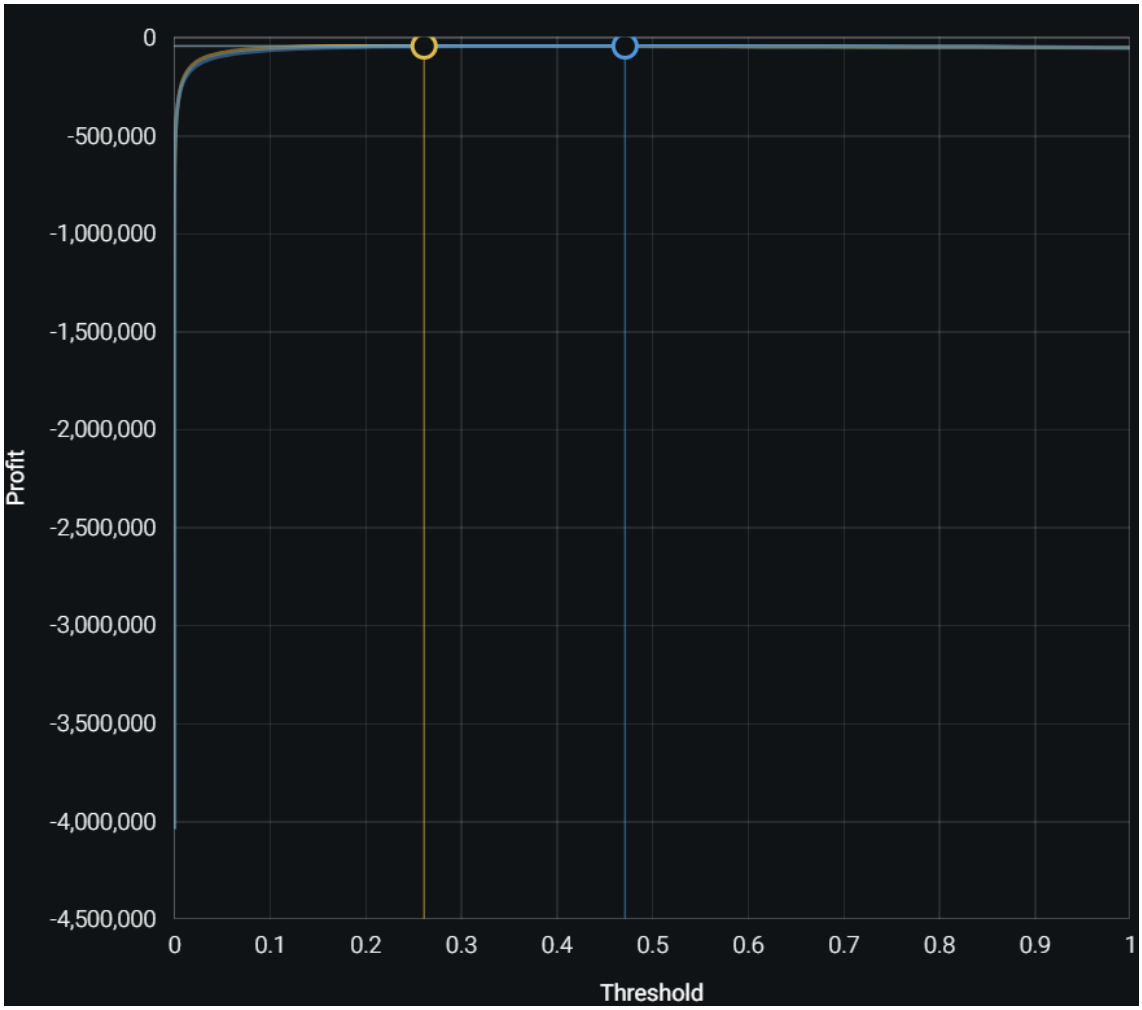
Table 6. Results with different threshold in GLM model

Threshold	**TPR**	**MCC**	**TN**	**FP**
			FN	**TP**
0.2639	0.5025	0.5185	267 482	769
			899	908
0.15	0.596	0.4991	266 791	1 460
			730	1 077
0.1	0.6497	0.464	265 953	2 298
			633	1 174
0.01	0.8794	0.2914	254 477	13 774
			218	1 589

Higher inventory level than needed occurs additional indirect inventory costs when more inventory space is needed and working capital is bigger. The highest costs occur when the product actually goes on backorder. Immediately occurring cost is urgent order to deliver product for customer as fast as possible. Other disadvantages such as customer dissatisfaction cannot be measured with profit curve. The following values are selected for the example: inventory replenishment cost 10, overstocking cost 5 and backorder cost 30. Unit for all values is currency, but it does not have an effect on this example, so it is not marked.

TN means that the prediction is right, and no actions are needed. The cost for TN is 0. FN is the worst situation, because any anticipation is not done based on prediction and product goes on backorder. The cost for FN is backorder cost. In TP situation, the model has predicted the backorder, so it is prevented by filling the inventory. The cost is then the inventory replenishment cost. If the prediction is FP, the backorder is prevented, but the ordered products are not sold, so the inventory level grows. The cost for FP is inventory replenishment cost and overstocking cost.

The results of the profit curves are presented in Fig. 15. There is not difference in the maximum profit, but the difference between thresholds is significant. Maximum profit is round -46k. The threshold for maximum profit is 0.26 for EGBTC model and 0.47 for GLM model. If the model was not used (actual positives affects backorder cost), total cost would be -54.21k, so the models offer financial benefits. The profit curves grow steeply when threshold is near 0, and otherwise the curves are relatively stable.



(a)

Selection Summary

MODEL NAME	PROFIT	THRESHOLD
eXtreme Gradient Boosted Trees Classifier with ...	-45,945	0.2617
GLM Blender (M73)	-45,755	0.472

(b)

Figure 15. Profit curve for the models: (a) Profit curve; (b) Description

The maximum profit is not the only metric to compare the models or to evaluate the results. Because the curve is stable, results can be optimized by changing the threshold. The target is to minimize both costs and FNs. When the threshold is smaller, more positives are detected. Different thresholds are tested, and results are presented in Fig. 16. Amount of FNs is marked with columns and costs with lines. Even though GLM model's maximum profit is received with higher threshold, it can perform almost as well as EGBTC model with lower thresholds. In GLM model, the costs grow significantly when threshold is under 0.15, and for EGBTC the same limit is 0.1. But above these limits, wanted relation between costs and FNs can be selected.

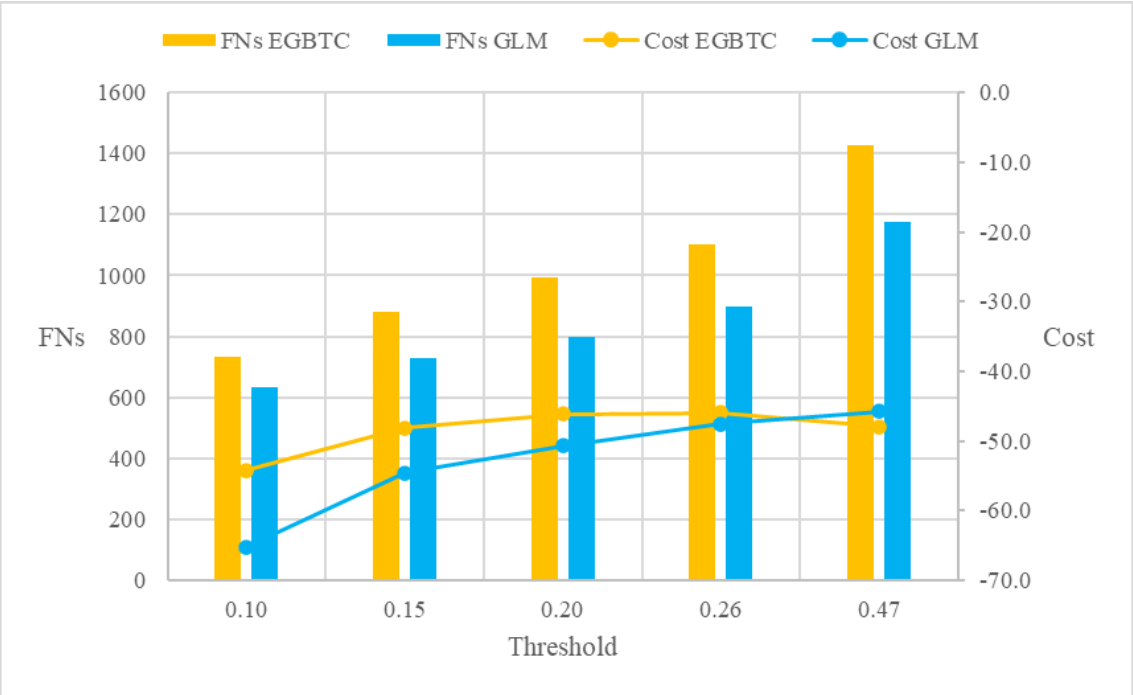


Figure 16. Cost and amount of FNs comparison between the models

The last step is to release cross validation and holdout data. The results are worse than with validation data. New testing data can be added to do predictions with it. With EGBTC model, testing data provides 0.32 TPR with 0.0027 threshold. The results of testing data are much worse. To achieve higher TPR, threshold should be extremely small, and then amount of FPs grows, and then costs grows. These results indicate that the model is overfitted from the training data. This also confirms that there are not clear signs which can predict if the product goes on backorder.

5 DISCUSSION

The purpose was to find out how ML can be used in SCM. The answers to research question was looked from literature and empirical study with DataRobot. The first research question was:

What SCM problems ML is suitable for?

The answer was found from literature review. Literature review showed that articles of ML in SCM is done and use of ML in SCM is diverse. Although, main focus is on the prediction. Besides demand prediction, there are other prediction problems and those are possible to solve with ML. In inventory management, ML is used to predict the amount to order with wanted conditions to optimize the prediction. The advantage of ML is that it can be used to solve complex problems such as production and inventory control policy where many things need to be considered at the same time. In location and logistics problems, ML can be used in different ways. It is possible to directly select the best locations, or to use clustering to group locations, so that human presumption does not affect on the results. Important factor in logistics is timeliness, which would be difficult without ML calculation. For supplier selection and evaluation, ranking models are developed, which can also be used for other ranking problems. Also in supplier selection and evaluation, the purpose is to avoid human subjectivity. Many other solutions were found. So, it is possible to utilize ML in many different problems in SCM, which were not found in this thesis.

The second research question was:

What benefits are achieved with ML in SCM?

The case articles introduced what benefits ML solutions are achieved. ML provided financial benefits and many concrete improvements to business. Received results were significant, as many models provided tens of percent improvement. ML solutions help to enhance customer satisfaction because better service is provided with more quality of products and selling in the right time and location. ML has important meaning for decision making, because it can provide new information compared to descriptive analytics. Especially, simulation models can have meaningful impact for decision making.

The case study concentrated on inventory management. The dataset was difficult to predict because of huge imbalance between the classes. The accuracy was almost perfect,

because the model predicted the negative class well, but predicting the positive class is more difficult. Specific reasons for backorder was not found. However, the model marked inventory level as the most important feature, so it is reasonable for company to consider inventory levels. The risk flag features were not found necessary. Even though, the results were not as good as wanted, it is possible to take the model to use, because there is not too high risks or costs. Another thing to consider is if the model is needed, because the amount of backorder is that small, but because the model predicts negatives well, the use of the model does not worsen the current situation.

DataRobot did not worked well enough with imbalanced dataset. To receive better results, more data preparation should have been done. DataRobot's methods were not enough for this dataset. However, DataRobot can work the best with binary classification problems, because most of the tools and charts are for binary classification. Important advantage of DataRobot is transparency to models, which is achieved with blueprints and how features affect on the model. All of the charts can be examined easily without more workload, which is not as easy when ML is coded.

Even though in this case study the results were not good, DataRobot can provide better results in other cases. Therefore, DataRobot offers good possibility for companies to start using ML. Using of DataRobot does not require knowledge of coding. There are not other costs than license of the software. Nevertheless, it is important to remember to evaluate the results that they are reasonable, and too high accuracy is not good. Overall, ML provides new possibilities in different areas of SCM. With ML software, a company can use ML with quite small investment. Also, testing ML do not have risks or consequences. In the future, research could be expanded both inside and outside DataRobot. DataRobot's other features (time series, multi-class classification) could be tested. Other ML software exist, so they could be studied, and features compared to DataRobot.

6 CONCLUSIONS

The goal of this thesis was to find out in which situations ML is suitable in the field of SCM and what are its benefits. The thesis was divided into two sections, literature review and empirical part. The purpose of the literature review was to find use cases that are already made, and to summarize the perceptions from the use cases. The empirical part focused on creating a case study with DataRobot and testing if DataRobot can be used in SCM.

In the literature review, 21 use cases were found from the literature and they were introduced. The use cases were divided into following groups: demand prediction, inventory management, supplier selection and evaluation, location and logistics problems and others. From the cases, it was perceived that ML's advantage is that it can be used to solve complex problems. The benefits that were received in the use cases were significant. The benefits were financial, and concrete improvements were perceived, such as a simulation model to help decision making and a model to improve the quality of products.

In the empirical part, DataRobot was used to solve product backorder problem. The goal was to predict if the product would go on backorder or not. The training dataset was extremely imbalanced, which caused difficulties for prediction. The model found very well the products belonging to negative class, but finding TPs were difficult. DataRobot's data preparation tools were not good enough to solve the imbalance problem. However, DataRobot can work better on other cases so it offers good possibility for companies to try ML easily as using DataRobot does not need huge investments, and the user does not have to have knowledge of coding.

REFERENCES

Alpaydin, E. (2014). *Introduction to machine learning*, Adaptive computation and machine learning series, third edition. edn, MIT Press, Cambridge, Massachusetts.

Baardman, L., Levin, I., Perakis, G. and Singhvi, D. (2018). Leveraging comparables for new product sales forecasting, *Production and Operations Management* **27**(12): 2340–2343.
URL: *https://doi.org/10.1111/poms.12963*

Ban, G.-Y., Gallien, J. and Mersereau, A. J. (2018). Dynamic procurement of new products with covariate information: The residual tree method, *Manufacturing & Service Operations Management* **21**(4): 798–815.
URL: *https://doi.org/10.1287/msom.2018.0725*

Ban, G.-Y. and Rudin, C. (2019). The big data newsvendor: Practical insights from machine learning, *Operations Research* **67**(1): 90–108.
URL: *https://doi.org/10.1287/opre.2018.1757*

Baryannis, G., Validi, S., Dani, S. and Antoniou, G. (2018). Supply chain risk management and artificial intelligence: state of the art and future research directions, *International Journal of Production Research* **57**(7): 2179–2202.
URL: *https://doi.org/10.1080/00207543.2018.1530476*

Becker, T., Illigen, C., McKelvey, B., Hülsmann, M. and Windt, K. (2016). Using an agent-based neural-network computational model to improve product routing in a logistics facility, *International Journal of Production Economics* **174**: 156–167.
URL: *https://doi.org/10.1016/j.ijpe.2016.01.003*

Biswas, S. and Sen, J. (2016a). A proposed architecture for big data driven supply chain analytics, *SSRN Electronic Journal* .
URL: *https://doi.org/10.2139/ssrn.2795906*

Biswas, S. and Sen, J. (2016b). A proposed framework of next generation supply chain management using big data analytics, *Proceedings of National Conference on Emerging Trends in Business and Management: Issues and Challenges* .
URL: *http://rgdoi.net/10.13140/RG.2.1.2183.4487*

Breiman, L. (2001). *Machine Learning* **45**(1): 5–32.
URL: *https://doi.org/10.1023/a:1010933404324*

Chakroun, A., Bouchti, A. E. and Abbar, H. (2018). Logistics and supply chain analytics: Benefits and challenges, *2018 Second World Conference on Smart Trends in Systems,*

Security and Sustainability (WorldS4), IEEE.
URL: *https://doi.org/10.1109/worlds4.2018.8611623*

Chicco, D., Tötsch, N. and Jurman, G. (2021). The matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation, *BioData Mining* **14**(1).
URL: *https://doi.org/10.1186/s13040-021-00244-z*

Ciccio, C. D., van der Aa, H., Cabanillas, C., Mendling, J. and Prescher, J. (2016). Detecting flight trajectory anomalies and predicting diversions in freight transportation, *Decision Support Systems* **88**: 1–17.
URL: *https://doi.org/10.1016/j.dss.2016.05.004*

DataRobot (2021a). Area under the roc curve. Accessed: 2020-01-12.
URL: *https://app.datarobot.com/docs/modeling/investigate/evaluate/roc.htmlarea-under-the-roc-curve*

DataRobot (2021b). Datarobot. Accessed: 2021-03-23.
URL: *https://www.datarobot.com/*

DataRobot (2021c). Optimization metrics. Accessed: 2021-01-12.
URL: *https://app.datarobot.com/docs/reference/opt-metric.html*

DataRobot (2021d). Smart downsampling. Accessed: 2020-01-20.
URL: *https://app.datarobot.com/docs/modeling/adv-opt/smart-ds.html*

Delage, E. and Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems, *Operations Research* **58**(3): 595–612.
URL: *https://doi.org/10.1287/opre.1090.0741*

Deloitte and MHI (2014). The 2014 MHI annual industry report - innovations that drive supply chains, *MHI,* Charlotte, North Carolina, USA .

Demchenko, Y., Grosso, P., de Laat, C. and Membrey, P. (2013). Addressing big data issues in scientific data infrastructure, *2013 International Conference on Collaboration Technologies and Systems (CTS)*, IEEE.
URL: *https://doi.org/10.1109/cts.2013.6567203*

Fallahpour, A., Wong, K. Y., Olugu, E. U. and Musa, S. N. (2017). A predictive integrated genetic-based model for supplier evaluation and selection, *International Journal of Fuzzy Systems* **19**(4): 1041–1057.
URL: *https://doi.org/10.1007/s40815-017-0324-z*

Forger, G. (2020). Ai starts to make some supply chain decisions, *Supply chain management review* **24**(2): 16–20.

Glaeser, C. K., Fisher, M. and Su, X. (2019). Optimal retail location: Empirical methodology and application to practice, *Manufacturing & Service Operations Management* **21**(1): 86–102.
URL: *https://doi.org/10.1287/msom.2018.0759*

Han, J., Pei, J. and Kamber, M. (2011). *Data mining: concepts and techniques*, The Morgan Kaufmann series in data management systems, Morgan Kaufmann.

Hazen, B. T., Boone, C. A., Ezell, J. D. and Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications, *International Journal of Production Economics* **154**: 72–80.
URL: *https://doi.org/10.1016/j.ijpe.2014.04.018*

He, L., Mak, H.-Y., Rong, Y. and Shen, Z.-J. M. (2017). Service region design for urban electric vehicle sharing systems, *Manufacturing & Service Operations Management* **19**(2): 309–327.
URL: *https://doi.org/10.1287/msom.2016.0611*

He, X., Ai, X., Jing, Y. and Liu, Y. (2015). Partner selection of agricultural products supply chain based on data mining, *Concurrency and Computation: Practice and Experience* **28**(4): 1246–1256.
URL: *https://doi.org/10.1002/cpe.3625*

Huan, S. H., Sheoran, S. K. and Wang, G. (2004). A review and analysis of supply chain operations reference (SCOR) model, *Supply Chain Management: An International Journal* **9**(1): 23–29.
URL: *https://doi.org/10.1108/13598540410517557*

Islam, S. and Amin, S. H. (2020). Prediction of probable backorder scenarios in the supply chain using distributed random forest and gradient boosting machine learning techniques, *Journal of Big Data* **7**(1).
URL: *https://doi.org/10.1186/s40537-020-00345-2*

Ittmann, H. W. (2015). The impact of big data and business analytics on supply chain management, *Journal of Transport and Supply Chain Management* **9**(1).
URL: *https://doi.org/10.4102/jtscm.v9i1.165*

Joshi, A. V. (2020). *Machine Learning and Artificial Intelligence*, Springer International Publishing.
URL: *https://doi.org/10.1007/978-3-030-26622-6*

Kenton, W. (2019). Backorder. Investopedia, Accessed: 2021-03-14.
URL: *https://www.investopedia.com/terms/b/backorder.asp*

Khalid, W. and Herbert-Hansen, Z. N. L. (2018). Using k-means clustering in international location decision, *Journal of Global Operations and Strategic Sourcing* **11**(3): 274–300.
URL: *https://doi.org/10.1108/jgoss-11-2017-0056*

Kishore, A. (2018). Predict products back-order to manage service level. data.world, Accessed: 2020-08-03.
URL: *https://data.world/amitkishore/can-you-predict-products-back-order*

Ko, T., Lee, J. H., Cho, H., Cho, S., Lee, W. and Lee, M. (2017). Machine learning-based anomaly detection via integration of manufacturing, inspection and after-sales service data, *Industrial Management & Data Systems* **117**(5): 927–945.
URL: *https://doi.org/10.1108/imds-06-2016-0195*

Li, Y., Jiang, W., Yang, L. and Wu, T. (2018). On neural networks and learning systems for business computing, *Neurocomputing* **275**: 1150–1159.
URL: *https://doi.org/10.1016/j.neucom.2017.09.054*

Liu, C., Shu, T., Chen, S., Wang, S., Lai, K. K. and Gan, L. (2016). An improved grey neural network model for predicting transportation disruptions, *Expert Systems with Applications* **45**: 331–340.
URL: *https://doi.org/10.1016/j.eswa.2015.09.052*

Lopienski, K. (2019). Backorder meaning: What does backorder mean for shippers? ShipBob, Accessed: 2021-03-14.
URL: *https://www.shipbob.com/blog/backordered-meaning/*

Ma, H., Wang, Y. and Wang, K. (2018). Automatic detection of false positive RFID readings using machine learning algorithms, *Expert Systems with Applications* **91**: 442–451.
URL: *https://doi.org/10.1016/j.eswa.2017.09.021*

Martínez, A., Schmuck, C., Pereverzyev, S., Pirker, C. and Haltmeier, M. (2020). A machine learning framework for customer purchase prediction in the non-contractual setting, *European Journal of Operational Research* **281**(3): 588–596.
URL: *https://doi.org/10.1016/j.ejor.2018.04.034*

Mercier, S. and Uysal, I. (2018). Neural network models for predicting perishable food temperatures along the supply chain, *Biosystems Engineering* **171**: 91–100.
URL: *https://doi.org/10.1016/j.biosystemseng.2018.04.016*

Mišić, V. V. and Perakis, G. (2020). Data analytics in operations management: A review, *Manufacturing & Service Operations Management* **22**(1): 158–169.
URL: *https://doi.org/10.1287/msom.2019.0805*

Mortazavi, A., Khamseh, A. A. and Azimi, P. (2015). Designing of an intelligent self-adaptive model for supply chain ordering management system, *Engineering Applications of Artificial Intelligence* **37**: 207–220.
URL: *https://doi.org/10.1016/j.engappai.2014.09.004*

Nguyen, T., ZHOU, L., Spiegler, V., Ieromonachou, P. and Lin, Y. (2018). Big data analytics in supply chain management: A state-of-the-art literature review, *Computers & Operations Research* **98**: 254–264.
URL: *https://doi.org/10.1016/j.cor.2017.07.004*

Ni, D., Xiao, Z. and Lim, M. K. (2019). A systematic review of the research trends of machine learning in supply chain management, *International Journal of Machine Learning and Cybernetics* **11**(7): 1463–1482.
URL: *https://doi.org/10.1007/s13042-019-01050-0*

Olson, D. L. and Wu, D. (2020). *Predictive Data Mining Models*, Springer Singapore.
URL: *https://doi.org/10.1007/978-981-13-9664-9*

Power, D. J. (2014). Using 'big data' for analytics and decision support, *Journal of Decision Systems* **23**(2): 222–228.
URL: *https://doi.org/10.1080/12460125.2014.888848*

Rebala, G., Ravi, A. and Churiwala, S. (2019). *An Introduction to Machine Learning*, Springer International Publishing.
URL: *https://doi.org/10.1007/978-3-030-15729-6*

Simchi-Levi, D. and Wu, M. X. (2017). Powering retailers' digitization through analytics and automation, *International Journal of Production Research* **56**(1-2): 809–816.
URL: *https://doi.org/10.1080/00207543.2017.1404161*

Souza, G. C. (2014). Supply chain analytics, *Business Horizons* **57**(5): 595–605.
URL: *https://doi.org/10.1016/j.bushor.2014.06.004*

Taghizadeh, E. (2017). Utilizing artificial neural networks to predict demand for weather-sensitive products at retail stores.

Tavana, M., Fallahpour, A., Caprio, D. D. and Santos-Arteaga, F. J. (2016). A hybrid intelligent fuzzy predictive model with simulation for supplier evaluation and selection, *Expert Systems with Applications* **61**: 129–144.
URL: *https://doi.org/10.1016/j.eswa.2016.05.027*

Wang, G., Gunasekaran, A., Ngai, E. W. and Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications, *International Journal of Production Economics* **176**: 98–110.
URL: *https://doi.org/10.1016/j.ijpe.2016.03.014*

Wenzel, H., Smit, D. and Sardesai, S. (2019). A literature review on machine learning in supply chain management, *in* W. Kersten, T. Blecker and C. M. Ringle (eds), *Artificial Intelligence and Digital Transformation in Supply Chain Management: Innovative Approaches for Supply Chains. Proceedings of the Hamburg Int*, Vol. 27 of *Chapters from the Proceedings of the Hamburg International Conference of Logistics (HICL)*, Hamburg University of Technology (TUHH), Institute of Business Logistics and General Management, pp. 413–441.

Wu, H., Evans, G. and Bae, K.-H. (2015). Production control in a complex production system using approximate dynamic programming, *International Journal of Production Research* **54**(8): 2419–2432.
URL: *https://doi.org/10.1080/00207543.2015.1086035*

Zhao, Y. (2019). Supply chain analytics, *International Series in Operations Research & Management Science*, Springer International Publishing, pp. 823–846.
URL: *https://doi.org/10.1007/978-3-319-68837-4*$_2$4

Appendix 1. Summaries of the found use cases

Pricing and demand in online fashion store for first exposure items Field: Pricing and demand prediction Algorithm: Regression trees Benefit: increase in revenue being approximately 10%	Rue La La is an online fashion business selling designer apparel and accessories. Rue La La offers 'flash sales', which are time-limited sales events lasting only 1 to 4 days. Rue La La needs solution to pricing and demand predicting problem, which occurs for new products that haven't been sold earlier. Significant portion of Rue La La's sales is these 'first exposure' items. Regression trees was used to predict demand, because regression trees work well data partitioning and also "it allows for a non-monotonic price demand relationship". Automated pricing decision tool was built to run every day and to give price recommendations to merchants. New pricing tool gave negative results only in low-priced products, where the price is less than $50. That problem was solved with maximum price increase of $5. As results researchers got total increase in revenue being approximately 10%. (Simchi-Levi and Wu 2017)
Forecasting and price optimisation in e-commerce marketplace Field: Demand prediction, pricing Algorithm: Unknown Benefit: increase of revenue and bookings	Groupon is an e-commerce marketplace where Groupon sells products or services from local merchants with discount. Groupon does the pricing for the products, but the merchant receives always the same amount. This way Groupon can affect on it's profit. The problem is first demanding and then pricing. The problem was solved by generating for the deal multiple demand functions. Product's selling period was divided to learning and optimisation. The demand function was chosen based on the learning period and the decision to increase or decrease the price was made on optimisation period. Totally the results were good as Groupon's revenue increased by 21.7% and total amount of bookings increased by 116%, though the revenue in two of five product categories decreased. (Simchi-Levi and Wu 2017)

(continues)

Appendix 1. (continued)

Demand forecasting of weather sensitive products Field: Demand forecasting Algorithm: Artificial neural networks (ANNs) Benefit: Little improve to forecast compared to forecast without weather data	The research used Walmart's sales data of weather sensitive products, such as umbrellas, bread and milk, to predict demand in different weather conditions. Demand forecasts can be made for certain products to find out specific models and attributes to effect on demand. Purpose is to predict needed level of inventory to avoid problems of inventory such as overstock or out-of-stock. The research used different algorithms to compare forecasting results. Random Forest (RF) was used as feature selection method. In prediction, different artificial neural network (ANN) methods were used, which were multilayer perception (MLP), time-delay neural networks, recurrent neural networks (RNN) and bagging. With MLP model gave most accurate results. In the research, it was found out that event, such as weekday, affected more on the demand of products than weather. Adding weather data to sales data improves the forecast but the difference is not significance. Though, using weather data may improve the model. (Taghizadeh 2017)
Predicting customer behaviour Field: Demand forecasting Algorithm: logistic Lasso regression, extreme learning machine and gradient tree boosting Benefit: 80% accuracy	Knowledge about future customer behavior is needed to plan manufacturing, inventory and point of sales. Researchers developed a dynamic and data driven framework for next month to predict if a customer will make purchase or not. Three ML algorithms, logistic Lasso regression, extreme learning machine and gradient tree boosting, were used and best results were received with the gradient tree boosting. Testing with over 10 000 customers and 200 000 purchases gave 80% accuracy and 0.949 AUC value. (Martínez et al. 2020)

(continues)

Appendix 1. (continued)

Demand prediction by identifying comparable products in order to predict demand for the new product	"At the core of many inventory problems - and operations management problems in general - lies the problem of demand estimation. The results of this estimation can then be used, among others, in order to address pricing, distribution, and inventory management questions. Baardman et al. (2018) develop a data driven approach for predicting demand for new products. The paper develops a machine learning algorithm that identifies comparable products in order to predict demand for the new product. The paper demonstrates both analytically and empirically, through a collaboration with Johnson & Johnson, the benefits of this approach." (Mišić and Perakis 2020)
Field: Demand prediction, inventory management	
Algorithm: machine learning algorithm	
Benefit: Unknown	
k-means clustering in international location decision	Companies that compete on global market face the problem of international location decision (ILD). Unsupervised machine learning method k-means clustering were tested to help decision making. Data of 24 indicators were collected from different sources for 94 countries. k-means clustering was made with 13 steps in MATLAB. The algorithm divided the countries into two clusters. The benefits of this method are that there is not any personal subjectivity in the results, and it can work with huge datasets with many parameters and variables. (Khalid and Herbert-Hansen 2018)
Field: International location decision	
Algorithm: k-means clustering	
Benefit: No personal subjectivity in the results, huge datasets	

(continues)

Appendix 1. (continued)

Location of delivery trucks of a "buy online, pickup in delivery truck" retailer	"Glaeser et al. (2019) considers a location problem faced by a real "buy online, pickup in store" retailer, that fulfills online orders through delivery trucks parked at easily accessible locations (e.g., at schools or parking lots). The retailer needs to decide at which locations and times to position its trucks, in order to maximize profit. To solve this problem, the paper first builds a random forest model (Breiman 2001) to predict demand at a given location at a given time, using a diverse set of independent variables such as demographic attributes of the location (total population, population with a post-secondary degree, median income, and so on), the retailer's operational attributes (such as whether or not the retailer offers home delivery in this location) and other location attributes (e.g., number of nearby competing businesses). Then, the paper uses fixed effects regression to account for cannibalization effects. Using the retailer's data, the paper shows how a heuristic approach based on greedy construction and interchange ideas together with the combined random forest and fixed effects model, leads to improvements in revenue of up to 36%." (Mišić and Perakis 2020)
Field: Location decision	
Algorithm: Random forest model, fixed effects regression	
Benefit: Improvements in revenue of up to 36%	

(continues)

Appendix 1. (continued)

Designing service regions for an electric vehicle sharing service	"In a different domain, He et al. (2017) consider the problem of how to design service regions for an electric vehicle sharing service. A number of car sharing firms, such as Car2Go, DriveNow and Autolib, offer electric vehicles to customers. Providing such a service in an urban area requires the firm to decide the regions in which the service will be offered, i.e., in which regions of the urban area are customers allowed to pick up and drop off a vehicle; this, in turn, informs major investment decisions (how large the fleet needs to be and where charging stations should be located). The main challenge in this setting comes from the fact that there is uncertainty in customer adoption, which depends on which regions are covered by the service. To solve this problem, the paper formulates an integer programming problem where customer adoption is represented through a utility model; due to the limited data from which such a utility model can be calibrated, the paper uses distributionally robust optimization (see, e.g., Delage and Ye (2010)) to account for the uncertainty in customer adoption. Using data from Car2Go, the paper applies the approach to designing the service region in San Diego, and shows how the approach leads to higher revenues than several simpler, managerially-motivated heuristic approaches to service region design." (Mišić and Perakis 2020)
Field: Location decision	
Algorithm: Integer programming problem	
Benefit: Informs major investment decisions, higher revenues	

(continues)

Appendix 1. (continued)

Modeling logistic routing in car terminal	Logistic routing options are complex, and short throughput times, high schedule reliability and low costs are wanted. Car terminal receives cars, makes needed treatments (such as refueling and washing) and parks cars and delivers cars. Model is needed to control and schedule car-flows so that treatments are done at the right time. Neural network model (NNM) creates the fastest car-flow. NMM worked better than previously used routing heuristics, the improvement of NMM is 12.7% in throughput time and compared to the best heuristics, improvement is 48%. (Becker et al. 2016)
Field: Logistic routing	
Algorithm: Neural network model	
Benefit: 12.7% in throughput time	

(continues)

Appendix 1. (continued)

Two approaches for inventory management	"Ban and Rudin (2019) consider a data-driven approach to inventory management. The context that the paper studies is when one has observations of demand, together with features that may be predictive of demand, such as weather forecasts or economic indicators like the consumer price index. To make inventory decisions in this setting, one might consider building a demand distribution that is feature-dependent, and then finding the optimal order quantity for the distribution corresponding to a given realization of the features. Instead, the paper of Ban and Rudin (2019) proposes two alternate approaches. The first, based on empirical risk minimization, involves finding the order quantity by solving a single problem, where the decision variables is the decision rule that maps the features to an order quantity, and the objective is to minimize a sample-based estimate of the cost. The second approach involves using kernel regression to model the conditional demand distribution, and applying a sorting algorithm to determine the optimal order quantity. The paper derives bounds on the out-of-sample performance of both approaches, and applies the approaches to the problem of nurse staffing in a hospital emergency room using data from a large teaching hospital in the United Kingdom. The paper finds that the proposed approaches outperform the best-practice benchmark by up to 24% in out-of-sample cost." (Mišić and Perakis 2020)
Field: Inventory management	
Algorithm: Kernel regression	
Benefit: Proposed approaches outperform the best-practice benchmark by up to 24% in out-of-sample cost	

(continues)

Appendix 1. (continued)

Estimating product quality Field: Product quality Algorithm: ML anomaly detection algorithms Benefit: The model found defective products with good performance. The model helps to reduce customer claim costs, and company's reputation is better, and customers are more satisfied.	Product's quality management is a key part of preventing company's reputation to declining and reducing warranty costs. A model was developed to estimate product quality using manufacturing, inspection and after-sales service data. The problem is imbalance between classes, so different ML anomaly detection algorithms were used to find suitable model. The model found defective products with good performance. The model helps to reduce customer claim costs, and company's reputation is better, and customers are more satisfied. (Ko et al. 2017)
Supplier selection model with gene expression programming (GEP) in textile industry Field: Supplier selection and evaluation Algorithm: Gene expression programming (GEP) Benefit: The model helps decision makers to compare suppliers and "simulate their behavior in the future". GEP based model can be used also in other decision-making situations to understand the relationships between inputs (criteria) and outputs (performance).	Supplier selection and evaluation is an important part of SCM. It is shown that with AI methods better results are received. The researched developed gene expression programming (GEP) model to evaluate suppliers by selected criteria. GEP is used because it provides visibility to model unlike AI methods (such as ANN and SVM) which are black box models for the user. The example was from textile industry and selected criteria were quality of materials, service, delivery, technical capability, cost and flexibility. The criteria and suppliers' performance were evaluated on a scale from 1 to 5. GEP was found to be more accurate than other AI methods. In textile industry case, quality of materials, service and delivery were the most effective criteria. The model helps decision makers to compare suppliers and "simulate their behavior in the future". GEP based model can be used also in other decision-making situations to understand the relationships between inputs (criteria) and outputs (performance). (Fallahpour et al. 2017)

(continues)

Appendix 1. (continued)

Supply chain partner selection model for agricultural products Field: Supplier selection and evaluation Algorithm: Back-propagation neural network, k-average method Benefit: Improvement of partner selection, the model's generalization ability is better	Supply chain partner selection is important as poorly made partner selection decision effects negatively to the whole supply chain. The researchers developed supply chain partner selection model for agricultural products. Partner selection analysis is made by analyzing selected criteria, which were partner's core ability, collaboration level, external environment, and internal performance. Back-propagation neural network is used for classification and k-average method for clustering. The results of the research were good as the analysis improved partner selection and the model's generalization ability is better. (He et al. 2015)
Ranking model for supplier evaluation and selection Field: Supplier selection and evaluation Algorithm: Artificial Neural Network Benefit: More accurate results than other model, and for situations, when all the information is not available	When the complexity of supplier evaluation and selection is low, decisions are based on experience and intuition, but it is not effective when the selection is not that simple. The proposed model can be used in ranking problems. The model consists of four phases, where firstly the criteria is defined, next Analytical Hierarchy Process (AHP) is used to prepare the data, and Adaptive Neuro Fuzzy Inference System (ANFIS) is used to recognize the most important criteria. Finally, Artificial Neural Network (ANN) is used to ranking. The model gave more accurate results than another model that it was compared. The model can be used in different decision making and comparing alternatives situations, but the researchers suggested the model especially for the situations, when all the information is not available. (Tavana et al. 2016)

(continues)

Appendix 1. (continued)

Predicting demand after transportation disruption Field: Demand prediction Algorithm: Grey neural network model Benefit: Provides information for decision making in production and inventory management	Transportation disruption is an unusual occurrence caused by accident such as natural disaster or social accident, which has significant impact for the whole supply chain, but especially for the first suppliers and their inventory. Market demand is difficult to predict after transportation disruption, but it is important for companies to predict the market demand to optimize inventory, purchase and production. The researchers developed grey neural network model to predict demand after transportation disruption. The research found out that improved grey model GM(1,1) model was more accurate than traditional GM(1,1) model. The improved model is feasible to be used to provide information for decision making in production and inventory management. (Liu et al. 2016)
Predicting flight diversions to help reorganizing delivery Field: Predicting flight diversion Algorithm: One-class SVM classifier Benefit: Prediction one hour before landing. Reorganizing trucks can be started earlier than before.	In supply chains that involve cargo airplanes, diversions have broad impact on supply chain. For some reasons, for example weather condition or technical problem, airplane has to land to another airport. This will occur delays, costs and CO_2 emissions as many trucks need to be reorganized. So that reorganizing could be done better, information of the diversion should get earlier. Researchers developed model to predict flight diversions automatically by detecting anomalous behavior of airplanes. Model uses one-class SVM classifier. The model got arguably high accuracy and it can predict the landing one hour before landing and two hours before originally scheduled landing. Reorganizing trucks can be started earlier than before. (Ciccio et al. 2016)

(continues)

Appendix 1. (continued)

Classifying RFID readings to the moved tags and the static tags Field: Product tracking Algorithm: Different ML classifiers (logistic regression, support vector machine, decision tree) Benefit: Average accuracy 93%, more cost effective and better performance than others	In supply chain, products are identified, tracked and traced with radio frequency identification (RFID). RFID tags are automatically read, but sometimes wrong tag is read when it is false positive. False positives occur for example when tag is too close of portal or read range is extended because of reflection from metallic objects. Problem is that tag is read but product is not loaded to truck, so incorrect invoices are sent. Different ML classifiers (logistic regression, support vector machine, decision tree) are used to classify RFID readings to the moved tags and the static tags. Average accuracy between different algorithms were 93%. When comparing to another researches, this method was cost effective and had better performance. (Ma et al. 2018)
Predicting the temperature of food in supply chain for food quality and safety Field: Food supply chain Algorithm: neural network Benefit: High accuracy was received average error being under 0.5 K, sensor location	In food supply chain, temperature of the products is important for food quality and safety. Temperature is measured with limited amount of temperature sensors. To receive information about the quality of supply chain, the researchers developed a neural network model to predict temperatures using heat transfer model. With neural network, high accuracy was received average error being under 0.5K. Also it was found that sensor should be located to near one of the bottom corners. (Mercier and Uysal 2018)

(continues)

Appendix 1. (continued)

Simulation model of supply chain ordering management Field: Inventory management Algorithm: Reinforcement learning Benefit: Comparing different scenarios and effect on supply chain, including total costs and the average customers' waiting time	In SCM system, inventory control is covering almost 50% of the total costs, and in inventory control, supply chain ordering management (SCOM) is important. SCOM is "an integrated approach to determine the ordering policy of each supply chain actors", and decision makers try to decrease the costs of inventory and increase customer satisfaction. The simulation model of the system, including the retailer, the distributer, the manufacturer and the supplier, was made, and reinforcement learning algorithm was used. Different scenarios were used to solve out how they impact on supply chain, including total costs and the average customers' waiting time. (Mortazavi et al. 2015)
Production and inventory control policy for complex production environment Field: Production and inventory control policy Algorithm: approximate dynamic programming (ADP) and Artificial Neural Network Benefit: Reduced costs by 8%, intermediate components are produced more on time	Efficient production and inventory control policy is needed in a complex production environment, where there are multiple working stations producing multiple intermediate components and end products. When producing different components and products, changeover times at work stations are causing costs and more time is needed. The developed model used approximate dynamic programming (ADP) and Artificial Neural Network (ANN). When compared to another policy, the model worked better in its entirety. Even though needed time is longer and demand accuracy is not as good, the model reduced costs by 8% and intermediate components are produced more on time. (Wu et al. 2015)

(continues)

Appendix 1. (continued)

Approach for dynamic procurement to decide amount of products to order when there are different sources with different costs and lead times	"Ban et al. (2018) consider the problem of dynamic procurement. In this problem, one has to decide how much of a product to order over a finite time horizon from different sources which have different lead times and different costs, so as to meet uncertain demand over the horizon. The problem setting studied in the paper is motivated by a practical problem faced by Zara, one of the world's largest fast-fashion retailers. The paper propose an approach that involves using linear and regularized linear regression to predict the demand for a given product, using the historical trajectories of other products. Then, the paper formulates a multi-stage stochastic programming problem, where the scenario tree is generated using the previously- estimated predictive model. Using real data from Zara, the paper shows how the existing approach, which ignores covariate information, can lead to costs that are 6-15% higher than the proposed approach." (Mišić and Perakis 2020)
Field: Dynamic procurement, inventory management, demand prediction	
Algorithm: Linear and regularized linear regression, multi-stage stochastic programming problem	
Benefit: Using real data from Zara, the paper shows how the existing approach, which ignores covariate information, can lead to costs that are 6-15% higher than the proposed approach.	