Using Deep Learning to Predict Back Orders
– A study in the Volvo Group Aftermarket Supply Chain

Jakob Bouganim
Konrad Olsson

Supervisor: Fredrik Stahre
Examiner: Magnus Berglund
External Supervisors: Bettina Linder & Baochang Wang
Abstract
The aftermarket holds a vital role in the Volvo Group value offer. Producing profitability by satisfying the customers needs for important spare parts, ensuring maximum uptime for the entire range of vehicles produced and sold. As the cost for keeping stock exponentially increases with a higher availability, the availability can never be 100%. This in effect means that there will be occasions where an order is placed on a part that is currently not in stock, creating a back order. And while not all of these back orders can be avoided completely, predicting them before they occur will allow for preemptive measures to be taken, potentially reducing lead times and costs. Deep learning is a sub-section of machine learning, the study of methods to make computers find complex patterns in data. Deep learning has had an increase in popularity as the computational power and available data has greatly increased in recent years and is something that Volvo sees potential in. This creates the aim of this study which is to develop a deep learning model to predict the occurrence of back orders.

In order to fulfill this aim, two main research questions were formed. The first research question intends to find underlying causes and factors that can explain the occurrence of back orders, in order to create the input features that the model can be trained on. This was initiated with a basis in literature, where a theoretical framework was created from different areas in the field of logistics as well as previous studies that combine logistics and machine learning. After this an empirical study was conducted where four previous initiatives from Volvo were found, that aim to explain the occurrence of back orders. As this was concluded, the findings were combined and synthesized into a list of factors that explain the underlying causes of back orders.

In the second research question the factors listed were translated into input features of the model, where all quantifiable factors that could be and located in the Volvo database were included. This created the data set used to train the deep learning model to predict back orders. After the feature creation was completed, the actual design and development of the model could commence. Based on literature concerning deep learning along with directives from Volvo, a deep recurrent neural network was developed. The exact size and shape of the model was varied and evaluated to find the best performance.

Evaluating the results showed several interesting findings. After training the model on one year of weekly data for 20,000 part numbers, the model proved to be skillful in predicting the occurrence of back orders. The model was able to predict 73% of back orders one week before they occurred (recall), and 72% of what the model deemed to be back orders were actual back orders (precision). The main challenges with predicting back orders were the imbalance between back order and a non-back order and the limit of one year of data. As the nature of back orders is that on average, only a few weeks per year will there be a back order on a given part, the training of the model becomes difficult. The difficulty with this imbalance is that the model is always less likely to predict a back order if the occurrence of back order itself is rare. The advantage of deep learning can be found with a large amount of data, and not being limited to one year of data is likely to produce better results. Despite these difficulties the model was highly successful in predicting the occurrence of back orders.
Acknowledgements

This master thesis was conducted at Volvo Group in Gothenburg over the spring term of 2019. This thesis is the final part in obtaining a Master of Science in Industrial Engineering and Management, with a specialization in logistics at Linköping University.

We would like to express our deep gratitude to all Volvo Group employees that helped making the journey that this project entailed immensely enjoyable and rewarding. We would also like to offer special thanks to our supervisors Bettina Linder and Baochang Wang for their great contributions and expertise that made this thesis possible.

Additionally, we would like to thank our supervisor at Linköping University, Fredrik Stahre, for always guiding us in the right direction. Your inspiring advice and ideas have truly been valuable for the outcome of this study.

Thanks are also due to our opponents whose critical but fair input truly aided the clarity of this study.
# Table of Contents

1. **Introduction** .......................................................................................................................... 1

1.1 **Background** .......................................................................................................................... 1

1.2 **Aim** .......................................................................................................................................... 2

1.3 **Study Directives** .................................................................................................................... 3

1.4 **Specifying the studied system** .............................................................................................. 4

2. **Current Situation** .................................................................................................................... 5

2.1 **Organization** ......................................................................................................................... 5

2.2 **Aftermarket (Service Market Logistics)** .............................................................................. 5

2.2.1 **Logistics Partner Agreement** .......................................................................................... 6

2.2.2 **Demand and Inventory Planning** .................................................................................... 7

2.2.3 **Dealer Inventory Management** ....................................................................................... 7

2.2.4 **Material Planning** ............................................................................................................ 8

2.2.5 **Product Segmentation** ..................................................................................................... 8

2.2.6 **Advanced Analytics** ....................................................................................................... 9

2.3 **Back Order** ........................................................................................................................... 9

2.3.1 **Reasons for Back Order Occurrence** .............................................................................. 10

2.3.2 **Back Order Recovery Team in Ghent** .......................................................................... 11

3. **Theoretical Framework** .......................................................................................................... 13

3.1 **Back Orders** .......................................................................................................................... 13

3.2 **Inventory Management** ........................................................................................................ 13

3.2.1 **Inventory Management Models** ................................................................................... 13

3.2.2 **Keeping Stock** ................................................................................................................. 14

3.2.3 **Forecasting** ................................................................................................................... 15

3.3 **Aftermarket Logistics** .......................................................................................................... 16

3.3.1 **Spare Parts** .................................................................................................................... 17

3.3.2 **Installed Base Information** .......................................................................................... 18

3.4 **Delivery Service** .................................................................................................................... 18

3.5 **Machine Learning** ................................................................................................................. 19

3.5.1 **Artificial Neural Networks** .............................................................................................. 19

3.5.2 **Deep Learning** ............................................................................................................... 20

3.5.3 **Recurrent Neural Networks** .......................................................................................... 21

3.5.4 **Convolutional Neural Networks** ................................................................................... 22

3.5.5 **Data and Choice of Features** ....................................................................................... 22

3.5.6 **Evaluating a Machine Learning Model** .......................................................................... 23

3.5.7 **Assessment Metrics** ....................................................................................................... 23

3.6 **Machine Learning in a Supply Chain Context** .................................................................... 25
1 Introduction

In this chapter a background for the study is presented that motivates the aim, which is presented and described. Later the directives, limitations together with the specification of the studied system is presented. Thereafter the purpose is broken-down into two overall steps, that provide an overview and guidance of the study. Lastly academic demands and research ethics that was used throughout the study is presented.

1.1 Background

The aftermarket is an important part of Volvo Trucks value offer. Their customers purchase contracts in conjunction with the purchase of a vehicle to ensure that service will be available in order to minimize the risk of a broken vehicle putting a stop to the operations of the customer. To cater to the demand of spare parts Volvo tries to keep a high availability throughout the supply chain. As the costs exponentially increase with increasing availability, a one hundred percent availability is impossible to deliver which causes shortages to occur. A shortage leads to a back order which causes Volvo to put a lot of effort in to ensuring that the part is delivered as soon as possible in order to ensure customer satisfaction and maximum customer vehicle uptime.

With a back order a lot of resources is put into delivering the demanded part, a back order recovery team searches for the part throughout the entire supply chain from production or a supplier, all the way down to another end customer where they might be able to buy back the part for a high price. This process thereby results in a high cost from the labor for finding and purchasing the parts and the express shipping needed to deliver parts in a timely manner. To be able to predict when a back order occurs or what causes them can thereby lead to a large cost reduction and higher customer vehicle uptime. While back orders make up a small portion of the total amount of orders, Volvo see this as an important area of improvement as the customers are more likely to react to these than when the parts are in stock as planned.

In recent years machine learning has seen many new areas of use, one if which is logistics and supply chain management. A typical use case for machine learning in supply chain management is to forecast demand. Using machine learning to forecast demand has shown great results, achieving higher accuracy than more traditional methods. Recently, computational power and technological advances has vastly increased the technology of deep learning, which has shown to surpass the results of more traditional machine learning methods as the amount of available data has increased. As a supply chain contains vast amounts of information being stored, this especially opens up great potential for deep learning to applied in this area.

Volvo do a lot of work, exploring machine learning in all parts of their operations and see this as an important part in ensuring future competitiveness. Volvo thereby see this as a crucial area to explore in order to improve the processes they use today by the use of modern technology. Deep learning is a relatively unexplored area at Volvo, and since deep learning has proved increasingly useful with increased computational speeds and available data they see this as a promising next step.

Combining a cost intensive part of the operations of Volvo, the back orders, and this new technology holds great potential. By predicting when a back order will occur it gives the business a possibility to take a more proactive approach and reduce the time and costs to solve back orders. This means an increased delivery service towards the customers and reduced costs for the business in whole can be achieved.
1.2 Aim

“The aim of this study is to develop a deep learning model to predict back order occurrence in the Volvo Group aftermarket.”

To clarify the central terms in the aim, here follows a description of these.

A deep learning model uses data to find patterns in data using several process layers with complex structures. The major part of findings patterns is done by a computer that carries out calculations. To allow the computer to find these patterns it needs to be fed input features and data for these features, so called training. In this study a deep learning model will be developed to achieve the aim. Developing a deep learning model include deciding on the appropriate network and model parameters such as size of the network and the number of training iterations.

Predict back order occurrence, where predicting means to forecast a future event. In this study the statement is concerned with whether a back order will occur or not. This prediction will be carried out by the afore mentioned deep learning model. In short, back order occurrence is when an order made in the supply chain and that order can't be fully met. For example, if the specific product is not currently in stock. That order is then sent back in the supply chain, and a so-called back order is raised.

Volvo Group aftermarket is the focus of this study. The term used by Volvo is Service Market, but the meaning is the same as the more general term aftermarket. In the automotive industry, the aftermarket is be the secondary market, concerned with the manufacturing, distribution etc. that take place after the sale of the vehicle to the consumer. Spare parts are a typical thing handled within the aftermarket.

In order to effectively fulfill the aim, it is helpful to break it down into several smaller parts as individual analysis of these parts will allow for a simpler overview and more clarification into what needs to be accomplished. The aim is broken down in to several research questions and smaller sub-questions, creating a process to fulfill the aim of the study. These questions aim to effectively gather all the information necessary to develop a working deep learning model that in the most effective manner possible can predict the occurrence of a back order in Volvos aftermarket. With an understanding of the Volvo organization and its processes, along with an understanding of the steps of creating a deep learning model, each question is chosen to serve as a practical step in creating a working model.

These practical steps start with a business understanding and mapping of the current operations, being useful for the later model formulating. To first get an understanding of the field of which the machine learning is to be applied is needed to be successful in the implementation of the model. (Shearer, 2000) The business understanding is seen as a vital part of the study as identifying the right factors greatly affects the end result, which motivates why it's a step on by itself and not a part of the model formulating step. The business understanding step focus on finding the factors correlating and explaining back order occurrence, for example inventory management decision factors such as price. This step is covered by the first research question.

Research question 1. What factors explain the occurrence of back orders in Volvo Groups aftermarket?
The model is central to the study, and therefore clearly formulating the characteristics of this is needed. This step comprises of deciding upon the final input features and model design decisions. This second step is covered by the second research question.

**Research question 2. How could a deep learning model be designed to predict back orders?**

These two steps are central and make up the backbone of the study, however beyond these the results will be discussed in a supply chain perspective. The effect and impact on the supply chain isn't thoroughly analyzed and therefore there are no research questions concerned with this. An overview of the steps that are included in the study and their corresponding research questions can be seen in Figure 1. A more detailed description of how the research questions with corresponding sub-question were motivated and decided on can be found in the Study Specificat.

![Figure 1 - Overview of the different fields in the study and their corresponding research question](image)

### 1.3 Study Directives

Used as starting point and guidance throughout the project lies the four directives given from Volvo. Giving guidance regarding delimitations and ways of carrying out the study.

The first directive is related to the geographical limitation of the study, where the directive states to involve only the “Ghent-EU” region, which consists of several European countries. Also related to the geographical limitation is the second directive, which dictates which part of the supply chain that should be studied. The directive reads that the scope also only concerns back orders that occur in the Ghent warehouse, meaning that back orders occurring down in the supply chain at regional and support distribution centers aren't considered. Back orders that occur in Ghent are often the ones that are the most strenuous and expensive, which in combination of the reduced complexity of the problem is the reason of this directive.

The third directive states a limitation regarding brands. This directive states that the project will only involve the Volvo Trucks brand, and no other products of Volvo group trucks product line will be considered. Examples of brands that this directive excludes are Volvo Penta, Renault Trucks, Volvo Construction Equipment. The fourth and last directive is relating to which type of model that should be developed. While other types of machine learning may be applicable and able to deliver a satisfactory solution to the problem, according to the fourth directive given by Volvo a model using deep learning will be developed. This limitation is based on that Volvo see deep learning as a promising area to explore.

A summary of the directives can be seen in the list below.

- Geographical market of Ghent-EU
- Only studying back order occurrences in the Ghent warehouse
- Studying Volvo Trucks brand
- Using deep learning
1.4 Specifying the studied system

To display and describe which part of the aftermarket supply chain of Volvo that will be in focus for this investigation, a clear delimation and definition of the studied system is made. This will be useful both for investigative work, and to make it easier to understand and grasp the study.

Given the directive regarding that the scope of the study is to involve the Ghent-EU region, it comes natural to only focus on this region. Given the directive that the study is only concerned with back orders that occur in the Ghent Central Distribution Center, Support and Regional Distribution Centers are not included in the studied system. Interesting data from downstream of the supply chain as for example demand, will be aggregated which enables to delimit this part of the supply chain. Upstream of Ghent, suppliers can be delimited from the studied system since the interesting data points like lead times and delivery precision are measured at Ghent and not at the specific supplier. Stock availability and other elements measured at each supplier is seen as redundant since it's covered by the elements measured at Ghent.

A visualization of the aftermarket supply chain can be seen in Figure 2, where the red dashed line indicates where the focus of this study will be. The arrows to and from Ghent are included in the studied system since the in and out-flow are seen as interesting for this study since these flows highly have an impact on the stock levels in Ghent.

![Figure 2 - Visualization of the supply chain of the Volvo aftermarket with the dashed red line indicating the studied system](image-url)
2 Current Situation
This chapter describes the current situation of the Volvo Group. The chapter is divided into sections describing the organization as a whole, as well as the departments acting in the area of service market logistics. Lastly, the current operations regarding back orders in the Volvo Group is described.

2.1 Organization
The Volvo Group AB, from here on Volvo, is one of the world’s leading manufacturers of trucks, buses, construction equipment and marine and industrial engines. The company as a whole employ close to 100 000 people and their products are sold to more than 190 markets worldwide. Volvo has production sites in 19 countries (“About us | Volvo Group,” 2019). Volvo is divided up into ten business areas consisting of Volvo Trucks, Volvo CE and Volvo Buses among others. Furthermore, Volvo is divided into three different truck divisions, Group Trucks Technology, Group Trucks Operations and Group Trucks Purchasing. There are also a number of support functions that operate cross business areas, such as Human Resources and Finances. (“Organization | Volvo Group,” 2019) The organization visualized can be seen in Figure 3, with the studied parts highlighted with blue. Volvo is also divided into 5 geographical markets, Europe being the largest regarding sales.

Figure 3 - Organization of Volvo Group AB with blue indications of the studied parts of the organization

The Service Market Logistics-unit (SML), is positioned within the Group Trucks Operations division and is responsible for the aftermarket. The goal of the operations of Service Market Logistics is to keep the optimal uptime of the Volvo vehicles and machines with no unplanned stops.

2.2 Aftermarket (Service Market Logistics)
After the sale of a truck, Volvo is obligated to keep spare parts for that truck for a minimum of 15 years. For Volvo this means a lot of trucks and brands that need spare parts at some point in time. For Volvo Trucks there is a total of around 120 000 active spare parts numbers today.

The supply chain of the aftermarket, from suppliers to the end customers is mostly controlled by Volvo which handles the flow of spare parts to the end customer. The suppliers deliver to Central Distribution Centers (CDC), which are large coordination centers that distribute parts downstream in the chain. In total, there are six Central Distribution Centers, one is located in
Ghent in Belgium. From the Central Distribution Center, the parts are either sent directly to the dealer, via a Regional Distribution Center (RDC) or a Support Distribution Center (SDC). Both the Support and Regional Distribution Centers are warehouse nodes closer to the dealer to shorten lead time and optimize the holding and transportation cost. The difference between the two is that Regional Distribution Centers often are placed further away from the Central Distribution Center in Ghent and acts as a smaller Distribution Center, while Support Distribution Centers are placed in more dealer-dense markets. Interaction between the users of the vehicles, the end customer, and Volvo is in almost all cases done via a local dealer. In Figure 4 below a sketch of the full Volvo aftermarket supply chain can be seen.

![Figure 4 - The supply chain of the aftermarket of Volvo Group](image)

The flow between the distribution centers consists of stock orders and back orders, where stock orders are the normal flow of goods planned per forecasts and back orders is the supporting flow, but more about this in 2.3. Overlooking the flow and stock levels are the two functions Dealer Inventory Management (DIM) and Demand and Inventory Planning (DIP), more on these in 2.2.2 and 2.2.3.

Volvo has a goal of being able to delivering 96% of all orders all over the world within 24 hours. To maintain this goal a safety stock is kept to safeguard against differences in the actual and forecasted demand. A software program is used to calculate the cycle service level, i.e. the number of stock outs during a cycle. Aspects such as price, economic order quantity and lead time is considered. Since keeping stock is expensive not all articles can have the wanted service level, which means there is a level of prioritization where some articles have a higher service levels than others.

### 2.2.1 Logistics Partner Agreement
Volvo and its network of dealers have implemented the Logistics Partner Agreement (LPA). The agreement regulates how spare parts are handled and distributed from Volvo and to the dealers. Dealers that sign the document have a closer cooperation with Volvo and data such as sales and stock levels are shared. The Logistics Partner Agreement implies that Volvo refills the dealers stock based on sales forecasts, much like a VMI-solution. VMI meaning Vendor Managed Inventory, being a supply chain integration where the supplier is responsible for the stocking decisions affecting the buyer’s inventory. In return, Volvo will buy back the parts that are not sold by the dealer.
With the Agreement Volvo can ensure the desired service level at the dealers and also better prevent articles with a low sales frequency from turning obsolete while in stock at dealers, since these are stocked higher up in the supply chain. The dealer is taking the stocking cost of the article and Volvo is taking the risk of the article not being sold or stock levels not being able to meet demand. Overall the Logistics Partner Agreement improves the stability of the supply chain and a better availability at the dealers. Another benefit of the Agreement is that Volvo can buyback a part from the dealer if it’s needed somewhere else in the supply chain, for instance in case of a back order, where the dealer can’t refuse to sell back the part more than a few times a year.

2.2.2 Demand and Inventory Planning
The Demand and Inventory Planning team (DIP) are liable for the forecasting process of the spare parts in the aftermarket supply chain. The focus area is the Central Distribution Center in Ghent, where the goal of 96% availability is the main goal that the team strive for.

The forecasting process, as can be seen in Figure 5, make use of the historical demand in order to via statistical automated methods together with additional information about sales data, give a forecast for the coming 4-weeks periods, which in turn gives a delivery schedule for the coming 12-month period. The large time frame is needed since the time from order to delivery vary from one week to half a year. The forecast is calculated and updated weekly. The delivery schedule is used to get an understanding of the stock levels in the Central Distribution Center.

There’s also a level of reactive approach where the gap between the forecast and the actual sales are monitored to intervene and adjust. The forecast for the different parts can either be over forecasted, meaning the forecast is higher than the actual demand, which introduces a risk of overstocking. It can also be under forecasted, which introduces an availability risk. The risk of overstocking and understocking are balanced in order to have a forecast that is as good as possible. Parts with lower variations takes a lower inventory to meet availability target.

There are also some initiatives in order to further improve the forecasting process. For instance, using data about the population of spare parts being used in trucks, and predicting when these will be changed in order to provide a forecast. This initiative is in the developing phase, and isn’t used in the forecasting today.

2.2.3 Dealer Inventory Management
For the dealers that have signed a Logistics Partner Agreement, decisions regarding stock holding and refill policies are made by Volvo and the team working with Dealer Inventory Management. These decisions are based on the characteristics of the article and the dealer. Several features are evaluated in order to decide the availability and achieve this to the lowest cost. The cost and order frequency together with the estimated demand makes the basis of the policy and on top of that a product segmentation based on customer criticality and the current life cycle of the article decide the wanted availability, also known as the service level. Beyond that, some features such as costs for stock shortages are also taken into consideration. New calculations are made continuously, which affect and updates the stock holding policies.
Based on the stock holding polices most of the orders are made automatically via integrated IT-systems when the stock levels reach the decided levels.

2.2.4 Material Planning
The Material Planning department is responsible for contact with the suppliers and make orders for the Central Distribution Center to meet the aftermarket need. Ordering is based on the forecast and delivery schedule from the Demand and Inventory Planning team. The material planners are responsible for getting back orders to the Ghent warehouse and to take proactive actions to try to avoid back orders.

2.2.5 Product Segmentation
The parts in the Volvo aftermarket have different characteristics, which is handled via a segmentation that can be seen in Figure 6. The two dimensions most seen is life cycle and customer criticality. Life cycle have an impact on the sales of the project, where initial and phase-out articles have a lower volume of sales. The customer criticality is both based on the how up-time critical it is for the functioning of the trucks, and a criticality from a business strategical point of view. The combinations of these two criticalities give the overall criticality. Price is also a factor considered in the segmentation. When it comes to the life-cycle segmentation, the number of years since the production ended for that truck comes into play, since spare parts only needs to be delivered 15 years after a sale. As seen in Figure 6, the number of segments is many. There are also four overall segments, Fast, Phase-in Medium, Phase-out Slow and Critical, that groups and covers all of the segments.

![Figure 6 - Segmentation on Life Cycle and Customer Criticality VTC/VC Ghent](image)

The segmentation has an impact on how the inventory is managed for that article, for instance an expensive part has lower stock to decrease the costs. Further, a part that is more critical to the customer needs a higher availability since stock outs are more ominous. Articles that are in the end of the life cycle, phase-out, have a decreasing demand and therefore the risk of over forecasting is greater.
2.2.6 Advanced Analytics

The Advanced Analytics team is embedded within the Service Market Logistics team and the supply chain optimization team. Their goal is to explore advanced analytical methods and tools in order to reach insights and apply these on the aftermarket of Volvo. Examples are applying machine learning in the forecasting or removing repetitive man work with automation. The Advanced Analytics team is working towards the whole aftermarket supply chain and collaborating on initiatives with different people to capture potential of existing data with new techniques to improve the supply chain efficiency. Today the part of machine learning, deep learning, is starting to be used in the ongoing initiatives, and the technology is seen as the next natural step in using and testing out to see what impact it can have on the optimization of the supply chain.

2.3 Back Order

Even though Volvo and the dealers work together and continuously improve the flow to maintain availability, shortages of stock occur. This can either be because of the stock policy, supplier issues or because of unexpected things in demand. When this happens and an order is placed there’s a stock out and a back order is raised. Meaning that an order of that part is sent upstream in the supply chain, towards suppliers and other warehouses in other geographical markets.

When an order is placed from downstream in the supply chain, either directly from a dealer or from a support or regional distribution center that order is in most cases met, and the parts are delivered in the supply chain. However, when the stock levels are insufficient in the Ghent warehouse, the order isn't met and a back order is raised. The back order gets the same priority as the incoming order.

A more detailed picture of the foregoing events in the Volvo aftermarket supply chain that triggers a back order in the Central Distribution Center in Ghent can be seen in Figure 7. This figure also showcases the impact on availability and solve time to get the article in place.

![Diagram of the steps superseding the occurrence of a back order with the corresponding aftermarket supply chain](image)

*Figure 7 – The steps superseding the occurrence of a back order with the corresponding aftermarket supply chain*
As can be seen in Figure 7 when the part is in stock at the dealer, Support or Central Distribution Center the availability is instant or can be solved within 24 hours. When the part isn't stocked in any of these points, the situation gets a bit more problematic and a back order occur of order and these are raised for the Central Distribution Center if they can't be fixed via a quick solution. If no quick solution can be found, the back order is raised in the system. Since the solution in these cases often is complicated, this often results in a slower recovery of the part. The aftermarket is important for Volvo, which lead to that the customer and their back orders are prioritized since customers value a high availability and the time a vehicle stands still, is very costly for the end customer. Therefore, a lot of resources are put into the process of getting the part into place.

Recovering a back order can be done in many different ways, depending on where it’s available. In cases when the needed part isn’t available anywhere near the Central Distribution Center Volvo goes through a rigorous process to locate and get the part shipped. Often the part is taken from far away in the supply chain. For example, directly from the supplier, from a production site or another dealer or warehouse in another geographical market. Another possibility that the Logistics Partner Agreement provides is the possibility to look at the stock levels at other dealers. The whole supply chain therefore comes in play when it comes to recover back orders.

Based on the importance of the part and the severity of it not being in place the order is assigned as "Stock Order", "Day Order" or "Vehicle off Road" (VOR). The back order gets the same label as the corresponding order. A Day Order has a lower priority and can often be solved via a shipment from the distribution centers upstream and the part can be in place within just a few days. In the VOR scenario the situation is a bit more acute and the vehicle may already be at the dealer awaiting the part. The vehicle can’t function properly without the needed spare part, therefore the term "Vehicle off Road". The severity decides the measures Volvo takes to get the part in place. A VOR takes more resources in form of time from operator and in delivery costs than a Day Order.

2.3.1 Reasons for Back Order Occurrence

Even though the inventory levels are tightly monitored and managed, back orders can never be fully prevented and therefore they will occur. The background and reasons for these stock outs and back order occurrence can be many. Generally speaking it can be a supplier issues, unexpected increases in demand and certain article characteristics that drives back orders. Spare parts that are overrepresented in the occurrence of back orders in the Central Distribution Center in Ghent often come from availability decisions. Since keeping stock is expensive, some articles are planned to have a lower availability.

The costs for back orders and stock outs are balanced against the costs of keeping stock, which leads to this overrepresentation for certain articles. For example, phase-out articles, parts that are in the end of the product life cycle and have a limited time frame that it's needed have an inventory policy of lower stock levels and back orders are planned for in a greater extent. The reason that the phase out has planned for a number of back orders is that the demand is decreasing and parts turning obsolete and not being sold should be avoided. Further, expensive parts are seen as overrepresented as back orders. Expensive parts tend to have lower inventory levels since the costs of carrying the articles. Related to demand, a forecast that is too low relatively the actual demand and a safety stock level that can't handle an unexpected increase will lead to stock outs and back orders. Also related demand and forecasts are first hits (new articles), that lack the historical data used for the forecast. Supplier related issued are often related to quality problems or that the supplier is having difficulties delivering in time.
2.3.2 Back Order Recovery Team in Ghent

A large number of back orders occur in the Central Distribution Center in since it’s positioned up in the supply chain, and the aggregated number of orders are large, which leads to a large number of back orders. The back orders that reach Ghent are often the ones that are not easily solved, and they are filtered through a service center which in some cases can find a simple solution first.

To give a perspective of the workload of the back order recovery team in Ghent. During the year of 2018 a total of 67 000 Vehicle of Road-back orders, the most critical order type, were managed, where 54 000 of these were related to truck and bus. Many of the back orders can be solved fairly quickly, for example if the part is available anywhere within reasonable distance to the Central Distribution Center or dealer, and therefore doesn’t cause that much problem. For the ones with a more difficult solution the time to be solved tend to longer and causing major problems for the end customers. Of the total number of back orders in Ghent 12 % are classified as VOR’s, the type of back order with the highest priority. Further, 13 % are day orders and 75 % are other, such as stock orders. Another attribute is that the VOR back orders in Ghent generally have a higher cost per order than the average back order.

When handling and recovering back orders the team use a framework called the Back-order recovery wheel, which can be seen in Figure 8. The wheel starts off with simpler and solutions with lower costs such as for example looking in the inbound flow and at reserved inventory within this flow and the flow for other brands. Then follows solutions such as checking dealer-to-dealer solutions and Regional Distribution Centers in other regions and technical alternatives are evaluated as well. After this the suppliers are investigated, other potential suppliers are evaluated and problems with the current supplier are trying to be solved. The time it takes to recover a back order vary from a couple of days to half a year, depending on the possible solution available. For instance, if the part isn't stocked anywhere and the supplier has problem, the recovery time can be very long.
For the back order recovery team working with the back orders in the Central Distribution Center Ghent that supplies the Ghent/EU region, the strategic focus for 2019 is a 100% avoidance of back order aging. By back order aging it means that it takes time for the back order to be solved, and therefore the back order age.
3 Theoretical Framework

This chapter contains the theoretical framework gathered from review of literature. The theoretical framework is divided into several categories and contains literature regarding logistics, machine learning and the combined usage of both. First general theory regarding inventory management is presented along with more specific theory relevant to this study, such as aftermarket logistics and back orders. This is where many of the theoretical factors are gathered. There is one section dedicated to machine learning, containing several parts, explaining the most important parts needed to make decisions regarding model design. Lastly to get an understanding for previous studies combining machine learning and logistics, a section is dedicated to this.

3.1 Back Orders

In the Postnord logistic dictionary (2011) a back order is defined as a special type of customer order, which is created when a complete delivery of a customer order can’t be met and the remainder of the order needs to be delivered at a later time. (PostNord, 2011) This definition is related to the definition of a stock out from Bowersox, Closs, & Cooper (2002). The definition reads, when a firm has no product available to fulfill the customer demand. There can be zero in stock for a specific article in a specific period, but the stock out doesn't occur until the customer desires a product. Olhager (2013) describes the effects of back order occurrence as costs for extra work, extra fabrication and a lower customer satisfaction. If these costs can be quantified they should be taken into account when deciding the economic order quantity. (Olhager, 2013) Also describing the costs in cases back orders are Langley, Coyle, Novack, Gibson, & Bardi (2013) that mentions the special ordering and transporting as drivers of costs. A back order often has a higher priority, needing a faster and therefore more expensive transportation. Also the shipment size tends to be smaller and the distance is farther which also contributes to an higher transportation cost. (Langley & Coyle, 2008)

3.2 Inventory Management

In this sub-chapter the theoretical framework relating to inventory management is presented.

3.2.1 Inventory Management Models

Oskarsson et al. (2013) state that there are three main questions within inventory management that needs to be answered:

1. When should products be ordered from a supplier/production/warehouse up the supply chain?
2. How much should be ordered each time?
3. How will uncertainties be safeguarded against?

The two first questions are related to cycle stock mentioned in chapter 3.2.1 and are connected with each other. The third question is related to the safety stock and is detached from the other two. (Oskarsson et al., 2013) To answer the first two questions, when and how much, the interval that orders are made and the quantity that is ordered each time, can either be fixed or varying. With a fixed order quantity and a fixed order interval the demand needs to be completely even and known, which is very rare. If the order quantity is fixed and the interval vary it's an order point system. In this system the economic order quantity (EOQ) is calculated and a new order is placed when the inventory level reach a decided order point. The EOQ is calculated with the Wilson-formula and focus on balancing the costs of ordering and holding stock. Then the order point is based on the lead time and the EOQ. Another alternative is to use a fixed interval and vary the quantity that is ordered each time. The quantity can for this be
decided either based on the economic order quantity or Lot-for Lot model can be used, which means the exact needed quantity is ordered. The Lot-for-Lot method means that ordering costs isn't taken into account, but there are no unnecessary holding costs. (Oskarsson et al., 2013; Jonsson & Mattsson, 2005)

The third question mentioned by Oskarsson et al. (2013) regarding how to safeguard against uncertainties, such as demand discrepancies and longer lead times. A suitably sized safety stock, which is mentioned in 3.2.1, will help avoid stock shortages that might occur in cases of uncertainties. To dimension the safety stock, the desired availability level is used in order to statistically compute the size of the safety stock. Important to point out is that safety stock only guards against random events. More regulatory patterns such as long going trends such as increasing demand, should be handled in other ways. (Oskarsson et al., 2013)

3.2.2 Keeping Stock
Stock is kept throughout the supply chain in order for its units to be able to be function more freely and be operated separately to increase the effectivity in the supply chain when it comes to costs and service. The reasons for not keeping stock is the costs for managing the warehouse and the costs for tying down capital in the form of stock. There are risks like for example obsolescence and damage during storage that regulates these costs. But there are also cost beneficial reasons that comes with keeping stock, for instance the ordering costs decrease if the number of order occasions are lower. A higher stock level leads to a higher availability and vice versa, therefore a decision on what availability level that is desired. Oskarsson, Ekdahl, & Aronsson (2013) splits up inventory into two major types, safety and cycle stock, which are visualized in Figure 9 below. Cycle stock is for meeting the demand, and the safety stock is to cover against uncertainties in demand or in the supply chain. (Oskarsson et al., 2013)

![Figure 9 - The two major types of inventory described by Oskarsson et al. (2013)](image-url)
Examples of uncertainties that Oskarsson et al. (2013) suggest that safety stock can safeguard against in order to prevent stock shortages include:

1. Delayed deliveries. A low delivery reliability from the supplier or the own production increases this risk.
2. Shortages in deliveries, which happens with a low delivery certainty. Shortages can be faulty products or wrong quantities.
3. An unexpected increase in demand from the customers.
4. The actual number of articles in stock is lower than according to the IT-system.

Robeson, Copacino, & Howe (1994) confirms that safety stock is a prevention of stock outs that occur because of the uncertainties. They divide the types of uncertainties into two types, listed below.

1. Demand uncertainty. Some degree of uncertainty will always exist, but the authors mentions working with the behavior of individuals in the supply chain, consolidating volume and improved forecasting as possible actions to reduce fluctuations.
2. Supply uncertainty. The authors’ further break down this type of uncertainty to supplier related and transportation related. To mitigate against these uncertainties the authors mentions a more comprehensive information exchange.

(Robeson et al., 1994)

Bowersox, Closs, & Cooper (2002) does a similar break down of uncertainties, but bring up performance instead of supplier related. With performance Bowersox et al. (2002) means the inventory replenishment time variations. So very much like Robeson et al. (1994) they mentions the uncertainties being either on the demand end, meaning that the even with good forecasting the actual demand during a replenishment cycle exceeds or falls short of the anticipated demand. Or being on the supply or performance end, meaning that a consistent delivery cannot be assumed. (Bowersox et al., 2002)

3.2.3 Forecasting
It’s hard to foresee the future but with forecasting methods, companies come pretty close. With forecasting you can foresee the demand so that it can be met through stock at the right place and time. (Oskarsson et al., 2013)

Robeson & Copacino (1994) segment forecasting into three different types:

1. Causation-based methods
2. Estimations from experts
3. Historical and computer-based methods

Oskarsson et al. (2013) further describes these different types with that causation-based methods use connection between a few variables have a direct impact on the demand. With expert estimation, people with great knowledge, such as salesmen, estimate the coming demand. The historical and data driven approach use the historical demand and patterns in the data are used to foresee the future demand. Examples of patterns that can be found are cyclic demand, more long up or downward trends. There is often also an element of random variations. (Oskarsson et al., 2013)

Important aspects to be aware of regarding of forecasting is that the forecast is not correct since it’s not the actual demand, and because of this a good forecast should include a measurement of the expected forecast error. Further aggregated forecasts give a more stable results, than for specific warehouses or products and that certainty of the forecast decreases with longer
forecast horizons. Lastly, information extracted from forecasts should never replace known information about the demand. (Olhager, 2013)

The characteristics of the demand can vary, and Olhager (2013) brings up five different components which contributes to the total demand.

- Trend. A gradual increase or decrease over time.
- Season. Patterns that reiterate on a yearly, monthly or weekly basis.
- Cycle. Pattern that reiterates during a longer time frame, economic situation etc.
- Level. The basic average general level.
- Randomness. Not all variations can be described by patterns.

These components are used in different ways in demand models and forecasting methods. (Olhager, 2013)

No forecasting method is perfect and it’s important to follow up the forecasting and measures its precision with the actual demand. (Olhager, 2013) brings up the forecasting error as a measurement of the forecasting precision, and if the forecasting error gets too large the forecasting method and the features should be evaluated and changed if needed. (Olhager, 2013)

3.3 Aftermarket Logistics

The logistics of an aftermarket, especially spare parts contrast those of different materials. Service level requirements are generally higher as the effects of a part shortage may have a severe financial impact. Furthermore, the demand is generally more sporadic which in turn also causes it to be harder to forecast. The price of individual parts is also likely to be very high. With high requirements as these, spare parts management is seen as an important area of research. (Huiskonen, 2001)

Supply chains are complex systems consisting of many actors and an intertwining flow, which cause disruptions at one point to give ripples on the water down the supply chain. (Samvedi & Jain, 2011) As supply chains are improved and more cost efficient, they also become more fragile and vulnerable to risks that lead to disruptions. Samvedi & Jain (2011) perform a computer simulation in order to evaluate the effects of disruptions in the supply chain. The results of the simulations show that the effect of a disruption decreases farther away from the disruption. The authors also state that they see risk management models as a way of being better prepared for disruptions, and in that way reducing costs. (Samvedi & Jain, 2011)

A similar conclusion is made by Lahiani, Apedome, Zhu, & Zhu (2018), but from a sourcing perspective. They see a more globalized world where manufacturers need to be more competitive and flexible. This results in suppliers that often are spread all over the world for reasons such as costs or access to unique products or technologies. (Lahiani et al., 2018) A global supplier often results in longer lead times, which in turn introduces a higher risk of inventory shortages. Lahiani et al. (2018) mention demand as an important feature to determine forecasts, and taking decisions in each step in the supply chain as important measures to avoid stock outs. The authors see that controlling risk management in the supply chain has a positive effect on the company’s operations, as they can apply solution more quickly than other companies. (Lahiani et al., 2018)
3.3.1 Spare Parts

Kennedy, Wayne Patterson, & Fredendall (2002) made an overview of literature of spare parts inventories and the unique aspects of these inventories were as follows. Reliability information to predict failure times are not available, there’s often a dependence between failures, demand is sometimes met via cannibalism of other units, costs of being out of a part is difficult to quantify. Obsolete machines that become replaced also needs to be considered, since some spare parts are made specifically for certain machines and it’s difficult to determine how many of these to stock. Lastly, components is more likely to be stocked rather than the complete part if the unit is expensive. Kennedy et al. (2002) point out that there are two major parts of maintenance, the scheduled and the unplanned repair. For the planned, it’s possible to schedule and order parts just in time. For the unplanned type, stock outs mean significant costs since the machine is standing still. The authors point out indicators such as machine failure, lead times, part use history, supplier reliability, stock-out objectives and inventory turn goals and to manage the inventory category-wise. Decisions that need to be made are where to place the spare part in the multi-echelon system and how to manage the risk with obsolescence. The risk of obsolescence is managed with inventory levels, balancing stock outs with inventory costs. (Kennedy et al., 2002)

Also on spare parts Bacchetti & Saccani (2012) investigate the gap between the models from literature and the practices in companies. The authors point out that service parts have grown into a major business, which introduces the need for spare part availability at the desired service level. Spare-parts inventory management is a complex matter due to the high number of parts, the presence of lumpy and intermittent demand patterns, the need of high responsiveness, and the risk of stock obsolescence. The inaccuracy of forecast demand, which partly exists because of new parts that are missing historical demand and failure data. Classification of spare parts is one approach to inventory management, since spare parts vary in for example cost, service requirements and demand patterns. Bacchetti & Saccani (2012) find the following spare parts classification examples from the literature:

1. Part cost/value
2. Part criticality
3. Supply characteristics / uncertainty
4. Demand volume/value
5. Demand variability
6. Part reliability
7. Life cycle phase
8. Part weight
9. Repair efficiency
10. Part specificity

Where 5-10 in the list above are the less common. (Bacchetti & Saccani, 2012)

When it comes to forecasting of spare parts Bacchetti & Saccani (2012) find that the different kinds of forecasting used in literature are time series, explanatory, a hybrid or some other method. The time series based consists of for example traditional methods, moving average or exponential smoothing. The explanatory methods take use of information such as failure data while the hybrids are machine learning methods in the form of neural networks and support vector machines. Other examples include advanced demand information such as order overplanning and early sales. (Bacchetti & Saccani, 2012) The gap between research and the companies investigated found where that the system perspective where missing since companies where making sub-optimical decisions because of the lack of information sharing. In
general companies used far more simple methods of forecasting and categorization, than the ones presented in research. (Bacchetti & Saccani, 2012)

### 3.3.2 Installed Base Information

The installed base is the whole set of systems or products sold by a company that still are in use. Dekker, Pinçe, Zuidwijk, & Jalil (2013) highlights this as information about the installed base, such as age, product-life cycle, is more useful and precise over time-based forecasting methods and provide great control of the service network. After a literature review over base information in forecast work Van der Auweraer, Boute, & Syntetos (2019) discuss that forecasting based on information of the installed based needs a lot of tailoring. They conclude in that in the last decade tools for data collection analytics has improved which the authors see as something that will drive more research within the area. (Van der Auweraer et al., 2019) The different uses that Dekker et al. (2013) identify are demand and return for each equipment, adapting to demand changes at both an aggregate and a disaggregate level as well as incorporating forecasts of part retrieval from equipment returns. Even though the information is useful, (Dekker et al., 2013) highlights the difficulties with managing it and that many companies weren't able to install it within their operations.

Chiang & Feng (2007) point out that information sharing in the supply chain is important and has an impact on costs, such as for example holding, back orders and ordering. Chiang and Feng (2007) also conclude in that the benefit is greater for the upstream supply chain members, especially in cases of higher supply uncertainty. That information sharing is important for the smooth operations of the supply chain is supported by Banerjee & Golhar (2017). Further they state that IT-system integrations between supply chain partners and internally lead to more effective decision making and better coordination. (Banerjee & Golhar, 2017)

### 3.4 Delivery Service

Oskarsson et al. (2013) indicate that the definition of logistics contains meeting the customers’ demand of delivery service. The term delivery service is then decomposed into several elements, giving a full reflection of a company’s delivery service. The delivery service elements are thus concerned with the actual delivery.

The delivery elements vary with different companies and industries, but six delivery elements highlighted by Oskarsson et al. (2013) can be seen in Table 1.

<table>
<thead>
<tr>
<th>DELIVERY SERVICE ELEMENT</th>
<th>DESCRIPTION</th>
<th>MEASURE POINT</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEAD TIME</td>
<td>Time from order to completed delivery</td>
<td>In-between supplier and customer</td>
</tr>
<tr>
<td>DELIVERY ACCURACY</td>
<td>The accuracy of delivery compared with the given lead time. Not delivering to late or too early.</td>
<td>Customer</td>
</tr>
<tr>
<td>DELIVERY RELIABILITY</td>
<td>Delivering the right product, in the right quantity and with the right quality</td>
<td>Customer</td>
</tr>
<tr>
<td>STOCK AVAILABILITY</td>
<td>The number of orders that can be delivered instantly from the supplier</td>
<td>Supplier</td>
</tr>
<tr>
<td>INFORMATION FLEXIBILITY</td>
<td>Providing information to the other part</td>
<td>Throughout</td>
</tr>
<tr>
<td></td>
<td>The ability of being flexible and meeting special demands from the customer such as faster transports etc.</td>
<td>Throughout</td>
</tr>
</tbody>
</table>
The authors also point out that there are some covariance, for instance a low stock availability affects the delivery accuracy. (Oskarsson et al., 2013)

3.5 Machine Learning

First explored by Arthur Samuel in his 1959 paper “Some Studies in machine learning Using the Game of Checkers” machine learning has many definitions. An often quoted definition is “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with the experience E.” by Tom M. Mitchell from his 1997 book *Machine learning* (Bell, 2014). Bell defines it as “with a computer running a set of tasks, the experience should be leading to performance increases”.

There are many different algorithms that can be used in machine learning which characterized by their output, usually is categorized into either supervised learning or unsupervised learning. Supervised learning uses labeled training data. Every training example has an input and an output, used for example to classify the input in to given categories. Supervised learning can generally be said to find the relationship between labeled input and output (Bousqaouï, Achchab, & Tikito, 2017).

Unsupervised learning does the opposite. The algorithm finds hidden patterns in a set of data and is commonly used for clustering data. (Bell, 2014); (Bousqaouï et al., 2017) Supervised learning is also referred to as classification learning, the learning scheme is presented with already classified examples as to learn a way to classify previously unclassified examples according to the different outputs or classes the classifier is trained with. (Witten, Frank, Hall, & Pal, 2016) These types of machine learning can be implemented through several different algorithms. An example is Support Vector Machines (SVM) that’s commonly used in Classification, Clustering and Regression (Murty & Raghava, 2016). It is a supervised learning method, best suited for high-dimensional, non-linear classification problems. Given a set of data where the data belongs to one of two categories, an SVM algorithm can determine which of these a given example belongs to. (Bousqaouï et al., 2017) Linear regression is another algorithm, commonly used and a well-known method to find the relationship between dependent and independent variables. It is widely used for prediction in fields such as economics and management. (Bousqaouï et al., 2017)

Decision Trees are graphs containing decisions and their respective outcome. Each node in a tree represents a question relative to a particular attribute. Random forest is a method of training multiple trees on different parts of the training data and randomizing subsets of features, predicting by averaging the predictions from the individual decision trees. This proves useful because while decision trees have low bias, they tend to overfit the training set. (Bousqaouï et al., 2017) Benefits of decision trees are their ease of use and white-box nature. Data requires little preparation, a working model can be created as long as the data is formalized into separated variables. The internal structure can be viewed, allowing for the validity to be tested with ease. (Bell, 2014)

3.5.1 Artificial Neural Networks

Mimicking the structure of animal brains an artificial neural network consists of several simple and connected processing elements, based on simple forms of inputs and outputs. Artificial neural networks (ANN) work best with large amounts of data and they produce good speed, making them appropriate for real-time scenarios. (Bell, 2014) In its simplest form, an artificial neural network is known as a perceptron. A perceptron consists of a single neuron with
multiple inputs and a single output, where each input is weighted and summed to be passed into an activation function. The most common activation function is the Sigmoid function which outputs a value between 0 and 1. (Mohammed, Khan, & Bashie, 2016)

![Figure 10 - A perceptron with a single neuron](image)

A multilayer neural network works in the same way that the single layer works, there are simply multiple layers containing neurons, known as hidden layers where the neurons pass the same activation function in multiple steps, allowing the network to fit more complex problems. (Bell, 2014) Multilayer perceptron is the most basic implementation of deep learning, further discussed in 3.5.2.

![Figure 11 - Multilayer perceptron](image)

### 3.5.2 Deep Learning

Deep learning works by the same principles as an artificial neural network, described above in 3.5.1. Deep learning is the general term for any neural network that has hidden layers between the input layer and the output layer. The number of hidden layers is commonly referred to as depth, which is where the notation deep learning comes from. (Ketkar, 2017) Different implementations of deep learning aimed for different types of problems are recurrent neural networks, discussed in 3.5.3 and convolutional neural networks discussed in 3.5.4.

The problem that a larger number of hidden layers solves, is the same as the key limitation of machine learning models in general, which is the feature engineering needed. (Ketkar, 2017)
Deep learning generally excels over traditional machine learning methods when the data set is of sufficient size. (Khan, Jan, & Farman, 2019)

To train deep neural networks, a technique called backpropagation is used. Backpropagation is divided into two phases, the forward phase and the backward phase. The forward phase takes the input data which is summed and then used in a multitude of multiplications and additions across the nodes and layers using the current weights, until the values exit the last layer to produce the output. The weights are what determine how much the multiplications and additions change the value between entering and exiting a node. The backward phase then compares the output value to the goal value and going backwards, updates the weights based on the learning rate and if the output was either larger or smaller than the goal value. (Aggarwal, 2018)

When training a neural network, the model aims to find the global optimum by updating the weights after each backpropagation, how much the weights are changed after each backpropagation is determined by the learning rate and the difference between the output and goal value, also known as error (Aggarwal, 2018). How the error is handled is decided by the optimizer. A commonly used modern optimizer is the Adam optimization algorithm, which is an extension to the older also widely used algorithm, stochastic gradient descent. (Brownlee, 2017)

A neural network also requires an activation function. In every node after the multiple inputs are summed and multiplied by the weights, the activation function is responsible for using this to create the output, both the output from one node into another and the last output from the final layer. A common activation function for the final layer is the sigmoid function. The sigmoid function transforms its input into a value between 0 and 1, which makes it the natural choice for the final layer of a binary classification. Other commonly used activation functions are the rectified linear unit (ReLU) and hyperbolic tangent (tanh). (Brownlee, 2019)

3.5.3 Recurrent Neural Networks

Recurrent neural networks (RNN) are neural networks that have feedback loops, adding a memory of previous decisions. (Ketkar, 2017) This feedback loop allows for the network to model dependencies among the data. (Ramakrishnan & Soni, 2018) An example for when this is frequently used is text analysis. When analyzing text the order in which the words occur is often of great interest, a model that sequences information thereby becomes useful (Aggarwal, 2018).

Long short-term memory (LSTM) is a recurrent neural network architecture. The benefit of recurrent neural networks is their ability to use contextual information, in practice however, the range of usable context is limited. Researchers find that the influence of a given input in a traditional Recurrent Neural Network tends to either get too large or disappear completely as it travels down the connections of the network. (Graves, 2012)

An LSTM layer is comprised of LSTM memory blocks, created to store information over long periods of time. The memory units inside the memory block holds information as long as the input gate remains closed, thus the memory cell be overwritten and the information stored for a later point in time.

An alternative to LSTM is the Gated Recurrent Unit (GRU), which was created with the same purpose as LSTM. In short, the difference is that the operations within the gated recurrent unit are slightly simpler and can therefore be faster to train. There is however no way of determining which one is better except for testing both and comparing results. (Nguyen, 2018)
3.5.4 Convolutional Neural Networks

A convolutional neural network is a neural network which has one or more convolutional layers. A convolutional layer takes an array of any dimension, flattens it into a vector if multidimensional and passes over the components to create a new smaller vector. (Ketkar, 2017) What separates convolutional neural networks from traditional machine learning is that when extracting features for example, traditional methods will have hand crafted features where as a convolutional neural network learns classes of objects. For images using a traditional artificial neural network the number of neurons would be high and the use of a convolutional neural network allows for fewer features and a deeper network that can be trained in a more efficient manner. (Aghdam & Heravi, 2017); (Ketkar, 2017) Convolutional neural networks were initially introduced for visual data processing like images and videos, they have however proved to be useful for almost any type of data (Khan et al., 2019).

3.5.5 Data and Choice of Features

The input used to train a machine learning model is a set of instances. Each is an independent example of the concept that is to be learned. Each instance is characterized by the values of each of the fixed, predefined attributes that make up the instance. A distinction can be made between different types of attributes, specifically attributes can largely be described as numerical or nominal. Numerical attributes measure numbers while nominal take on a predefined finite set of values that are distinct symbols. The value itself only serves as a label or name. (Witten et al., 2016)

When the final input features are to be selected, they could be so based on the usefulness. One method for analyzing feature usefulness for feature selection is Correlation-based feature selection, as described by Dong & Liu in their 2018 book. This is conducted by looking at feature relevance and feature redundancy. Feature relevance is based on the correlation between features and output class, and the feature redundancy is based on the correlation between different features. (Dong & Liu, 2018)

When a categorical input is used, these needs to be converted to numerical inputs to be able to use these as input to a machine learning model. To avoid ordering between the categories categorical encoding techniques is used. Potdar, Taher, & Chinmay (2017) describes a number of different categorical encoding techniques in their comparative study. One of which being the...
binary coding where categorical values are assigned to ordinal values, then binary values and split into different columns. (Potdar et al., 2017)

### 3.5.6 Evaluating a Machine Learning Model

As there are multiple methods for inferring structure from data, to determine which ones to apply to a particular problem there is a need for a systematic way to evaluate their performance to allow for comparison. Performance on the training set is not a good indicator for performance on a different independent data set, thus there is a need for methods of predicting performance in practice. With a sufficient amount of available data this is simple enough, train a model based on a large training set and evaluate it on another. High quality labeled data is however often hard to come by which creates a need for other methods which does not require vast amounts of data.

A common method for evaluating a machine learning model is Cross-Validation (CV). When employing cross-validation the available data is divided into two parts, a certain amount for training and the rest for cross-validation. There is however a drawback to this method, all different classes represented in the data set has to be in both the training and CV set to about the same proportion, the reasoning behind this is simply that a model trained with a data set cannot be expected to learn classes not present in the training set. A general way to combat this deficit is to repeat training and testing several times, each using a random sample of the available data as training data and CV-data. The standard method for applying this is called stratified 10—fold cross-validation. The data is divided randomly into ten parts, the data is trained on nine of the parts and then tested on the tenth. This is repeated ten times and the ten test error estimates are averaged to represent an overall error estimate. (Witten et al., 2016)

### 3.5.7 Assessment Metrics

When it comes to predictive analysis, de Santis, de Aguiar, & Goliatt (2017) press the importance of using the correct assessment metric. When working with a binary problem, all results of the model can be sorted into the confusion matrix seen in Table 2 which states whether the model predicted a positive or negative, and if this was correct or not. The number of predicted data points in each box is used in different formulas as assessment metrics. (de Santis et al., 2017)

<table>
<thead>
<tr>
<th></th>
<th>POSITIVE PREDICTION</th>
<th>NEGATIVE PREDICTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>POSITIVE CLASS</td>
<td>True Positive (TP)</td>
<td>False Negative (FN)</td>
</tr>
<tr>
<td>NEGATIVE CLASS</td>
<td>False Positive (FP)</td>
<td>True Negative (TN)</td>
</tr>
</tbody>
</table>

The most common assessment metric when it comes to predictive analysis models is the accuracy which can be seen in Equation 1, where the number of truly predicted examples are divided with the total number of examples.(de Santis et al., 2017)

\[
\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN}
\]

*Equation 1 - Accuracy rate based on data from the confusion matrix used as an assessment metric*

However, when working with imbalanced data sets where the number of positive data points are far less than the number of negative, the accuracy can be high even if all examples are predicted as positive. Other possible metrics are the precision and recall metric which can be seen in Equation 2 and Equation 3. The precision expresses the accuracy of a positive prediction, giving a measurement of the predicted positive examples that actually is positive.
The recall describes the sensitivity of the model in the way that it measures the models ability to pick up all actual positive examples. (de Santis et al., 2017)

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

Equation 2 - Precision based on data from the confusion matrix

\[
\text{Recall} = \frac{TP}{TP + FN}
\]

Equation 3 - Recall based on data from the confusion matrix

Witten et al. (2016) also highlight the precision and recall as useful metrics. They describe the metrics from an information retrieval perspective, where recall is the number of retrieved documents that are relevant divided with the total number of relevant documents. Precision is the number of retrieved documents that are relevant divided with the number of retrieved documents. (Witten et al., 2016)

Combining the precision and recall and finding the harmonic mean of the two can be achieved with the F-score, that can be seen in Equation 4. (Sasaki, 2007) This score combines the precision and recall measurements and therefore is a good measure when these needs to be balanced. Witten et al. (2016) also mentions the same metric, but under the name of the F-measure and point out that it’s usability when a single measure wants to be used. (Witten et al., 2016)

\[
F\text{score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]

Equation 4 - F-score used to finding the harmonic mean between precision and recall

de Santis et al. (2017) highlights another metric that measures the false-positive rate, which is number of false positives divided by the total number of negatives. The formula for the fall-out metric can be seen in Equation 5.

\[
\text{Fall out} = \frac{FP}{FP + TN}
\]

Equation 5 - Fall-out based on data from the confusion matrix

Using the fall out and precision the Area Under the ROC-curve can be calculated, per the formula in Equation 6. Whereas the ROC-curve is the Receiver Operating Characteristics-curve, which is a visualization of the trade-off between the precision and the fall-out. More specifically the area under the curve (AUC) is the probability to correctly identifying which one of the data points that is noise and which one that is signal plus noise. The area under the receiver operating characteristics curve (AUC\text{ROC}) can be used as a single measurement of which classifier that has the highest capability of evaluating which model that is best on average. (de Santis et al., 2017) If the area under the ROC-curve is larger, the model is better (Witten et al., 2016). An area of 1 means that the model perfectly classifies the classes and finds all true positives without any false positives.

\[
\text{AUC}\text{ROC} = \frac{1 + \text{Precision} - \text{Fall out}}{2}
\]

Equation 6 - The Area Under the ROC-curve
Further de Santis et al. (2017) mentions that several other metrics exist, but argues for that the AUC-metric is the one most applied in literature for benchmarking and assessment.

Further, the AUC-metric can be applied using the precision-recall curve (PRC). Plotting the precision and recall at different threshold values and calculating the area under this curve, \( \text{AUC}_{\text{PRC}} \). The advantage of using precision-recall curve as opposed to the ROC curve is that the ROC curve can give misleading results for imbalanced data sets. As often is the case, the 0-class is more frequently occurring and the ROC curve can give a misleadingly high performance from correctly predicting the 0-class. The precision-recall curve can however provide an accurate prediction of classification performance with a moderate to large class imbalance. (Saito & Rehmsmeier, 2015)

### 3.6 Machine Learning in a Supply Chain Context

Several examples of machine learning being effectively applied to improve supply chain solutions have been found. Carbonneau, Laframboise, & Vahidov discuss in their 2008 paper the application of machine learning techniques in forecasting demand in a supply chain. The purpose of their paper is to compare more “traditional forecasting methods” to modern implementations of machine learning, using neural networks and support vector machines, without participation or data from other stakeholders in the supply chain. Their findings show that while the machine learning methods performed well, the increase in accuracy over a simpler linear regression model was not large enough to warrant use of these complex methods. They suggest that using more data from further down in the supply chain can improve the results by providing information that explains the demand further.

Furthermore Shahrabi, Mousavi, & Heydar (2009) conducted a similar study. The authors applied neural networks and support vector machines to forecast the demand of car components and compared the results to traditional methods. They concluded that artificial neural networks can forecast demand more accurately than any of the methods tested. Chi, Ersoy, Moskowitz, & Ward (2007) arrive at conclusions similar to those of Carbonneau et al. (2008) . The purpose of their study is to model and optimize a vendor managed replenishment system using machine learning. They concluded that using machine learning for this purpose was highly successful and desirable as it can be implemented without deep statistical knowledge needed for other methods. They also comment on the fit of use in supply chain research which they deem appropriate due to its highly complex and information rich nature.

Supply chain management is become increasingly complex and as a response to this also more information intensive. Both business and academics seek ways to utilize this information in making more robust decisions and have concluded that machine learning can be of special interest in achieving this, and while the wide acceptance of machine learning as a tool, applications are still emerging within the supply chain field. (Priore, Ponte, Rosillo, & de la Fuente, 2018)

In their 2010 study, Yu & Chen proposed the use of an artificial neural network, with features based on external factors, for predicting the demand of parts in the auto aftermarket. The external factors used for building the model are divided into four categories each containing two or three factors. The first category is Economic environment, containing the gross domestic product (GDP), Consumer price index (CPI) and urban residents monthly disposable income per capital (UDI). The second is Auto industry containing vehicle production (VP) and average number of family cars per 100 households (AFC). Third is weather consisting of Average monthly temperature (AMT) and Average monthly rainfall (AMR). Last is Utilization rate of car, explained with Monthly highway passenger transportation volume (MHPT) and monthly highway cargo transportation volume (MHCT). Also mapped is the company sales records of
each auto part for each month, the assumption is that the sales of each part per month is influenced by these external factors. Yu & Chen concluded that a neural network based prediction, when tested in a real world scenario, has shown to be a promising tool in predicting demand. (Yu & Chen, 2010)

An alternative approach to inventory management using machine learning is highlighted by de Santis, de Aguiar, & Goliatt (2017) As a complement to the more traditional inventory management models they identify the materials that have a risk of back order occurrence and intercepting the back order, giving the business a reasonable time to react. In order to achieve the identification and allowing for an earlier interception the authors propose applying a supervised machine learning algorithm. (de Santis et al., 2017)

In their study de Santis et al. (2017) used data from 8 weeks and 21 input features, listed below.

1. Current inventory
2. Registered transit time
3. In transit quantity
4. Forecast sales for 3, 6 and 9 coming months
5. Sales quantity during the prior 1, 3, 6 and 9 prior months
6. Minimum recommended amount in stock
7. Parts overdue from source
8. Source performance during the last 6 and 12 months
9. Amount of stock orders overdue
10. General risk flags (a total of 6 unspecified risk flags)

The output from the model is a binary value, which is also used in a training purpose, telling if the product went on back order or not. With the model the authors achieved a precision of approximately 0.60, which indicates the ratio that the model can identify the positive class. The recall of their model is 0.2, which indicates that the model can predict 20% of the back orders. Given this precision and recall, the authors hypothetically make a situation where these back orders can be avoided which would increase the service level from 99.27 % to 99.42 %. (de Santis et al., 2017)

To summarize literature found in intersection between the fields machine learning and logistics is mostly within forecasting, and where machine learning is used to improve the performance of the forecast. The findings however include stock replenishment and back order prediction as well. A summary of the with the authors and their area of study is presented in Table 3 below.

<table>
<thead>
<tr>
<th>AUTHOR</th>
<th>AREA OF STUDY</th>
</tr>
</thead>
<tbody>
<tr>
<td>CARBONNEAU, ET AL. (2008)</td>
<td>Demand Forecasting</td>
</tr>
<tr>
<td>SHAHRABI, ET AL. (2009)</td>
<td>Demand Forecasting</td>
</tr>
<tr>
<td>CHI, ET AL. (2007)</td>
<td>Stock replenishment</td>
</tr>
<tr>
<td>YU &amp; CHEN (2010)</td>
<td>Demand Forecasting</td>
</tr>
<tr>
<td>DE SANTIS, ET AL. (2017)</td>
<td>Back Order Prediction</td>
</tr>
</tbody>
</table>
4 Study Specification

This chapter aims to further specify the study aim by describing two research questions that are further broken down into sub-questions.

4.1 Back Orders and the Aftermarket

According to Witten et al. (2016) who propose the use of the CRISP-DM model, a general methodology for data mining, it is important to first understand the field in which the machine learning model is to be applied, greater so than the actual technical knowledge of machine learning itself. This is due to the fact that to be successful in actually implementing a model, understanding and choosing the correct input and output is of greater importance than understanding how to implement it (Shearer, 2000). A good understanding of the current operations and what is the reason for back order occurrence can be seen right before a back order will help to understand what makes back order occur. In this study the Ghent Central Distribution Center is examined and the back orders occurring there. A back order occurs when an order of an article is placed to Ghent and that order can't be met in full, accordingly to the definition of a back order from PostNord (2011) and Bowersox et al. (2002) back orders are tightly connected with stock outs.

To understand what drives stock-outs, or back orders and what events and characteristics can be seen in the supply chain right before or at the time a back order occurs is an important step. Examples of reasons can be supplier related such as long lead times or supplier problems that result in delayed deliveries or quality issues. Other examples include discrepancies between the forecasted and actual demand or how the inventory of the article is managed, leading to stock-outs. Factors that impact the inventory management of the part include price, current life-cycle as well as criticality and reliability of the part (Bacchetti & Saccani, 2012). These examples, that are handled as factors that impact stock levels and back order risk. These factors are vital to deciding on the features used in the model.

Undertaking a rigorous effort to unravel which events that primarily lead to the occurrence of back orders is important. The reason of the importance is that these events and patterns will be of great use when it comes to deciding what features to take forth in the study and use in the model. As formulated by Jason Bell (2014), machine learning projects start with a question or hunch. For example, a perceived correlation between factors or an idea that something can be predicted. This leads to the first research question:

**Research question 1.** What factors explain the occurrence of back orders in Volvo Groups aftermarket?

This research question is divided into three sub-questions, where the first one is related to a more general approach to find useful factors that impact back orders from literature and earlier work. The reason for this general approach is to not get stuck in the reasons that Volvo sees as the ones with the greatest impact, but to seek for factors from several companies and researchers. This will allow for findings from sources unaffected by the data available through Volvo and avoid any bias inferred from the proceedings in the organization. Therefore, gaining insights from a wide array of sources and earlier studies will provide a thorough and general research of what drives back order occurrence and stock outs in inventories. Since stock outs are related to back orders, this motivates that literature regarding inventory management is seen as one of the fields of investigation. Furthermore, the patterns that are sought can be of a black box nature and thus largely incomprehensible when observed by a person (Witten et al., 2016). This motivates the first sub-question:
Sub-question 1.1. What factors of back order occurrence can be identified from literature?

While a general approach provides a more unbiased understanding of the field, a study applied on the studied company is needed to understand what distinguishes the aftermarket of Volvo from the general aftermarket and inventory supply chain. Further, to get a clear picture of the underlying factors of back order occurrence, a thorough analysis of the operations related to the aftermarket and distribution of spare parts according to Volvo's organization is necessary. This will allow for findings of relevant factors that can explain the underlying causes of back order occurrence that may not have been identified from literature. The factors that have been identified from the literature are used as a foundation and inspiration when seeking for factors at Volvo. This analysis also allows for the factors identified from literature to be put in to context of the Volvo operations and factors found at Volvo to be structured according to previous findings. This means that factors not seen as important at Volvo are still taken forward, reducing the bias. This leads to the second sub-question:

Sub-question 1.2. What underlying factors of back order occurrences can be identified from the operations of Volvo?

When these both sub-questions have been answered, the number of factors could be quite extensive and the factors from literature and Volvo need to be consolidated to get a list of factors that will answer to the first research question. A declaration over the factors found in literature that are applicable in the Volvo operations as well as the factors found as important from Volvo, even though the findings from the literature aren’t that clear. These factors will later be used as features in the model, which means that some kind of consolidation will take part, and the ones that best explain back order occurrence is taken forward. A large number of factors will be taken forward, which means that not that many factors will be sorted out. However, some kind of threshold is needed to sort out factors that for example has an unclear connection to the occurrence of back order. Therefore, the factors seen as candidates, are the factors that is seen as the ones that correlates best with and explain back order occurrence. This leads to the third and last sub-question.

Sub-question 1.3. Based on the literature and the Volvo organization, what factors best explain back order occurrence in the Volvo Supply chain?

4.2 Deep Learning to Predict the Occurrence of Back Orders

When sufficient understanding of the area of aftermarket logistics and back orders has been acquired, the focus can shift more toward the technical implementation of the proposed solution. Since the main aim of this study is to create a working model it is of great importance to understand the concept of machine learning, what it is and how it can be effectively implemented. The fundamental understanding within the field of back orders and the found factors will be of great use and will aid in the process of deciding upon relevant datapoints and features to feed to the model. This is necessary information as the selected datapoints dictates fundamental deep learning decisions, like selecting the most appropriate algorithm, and deciding upon input features. This is due to the fact that the performance of different deep learning frameworks depends on the characteristics of the input data (Khan et al., 2019). Given directives from Volvo the scope of algorithm selection will be in the realm of deep learning. This gives the question:

Research question 2. How could a deep learning model be designed to predict back orders?

How deep learning is applied can be divided into several steps as there are several parts that can be varied and thus need analysis before a model is created. This is discussed by Shearer
(2000) who suggests that after business understanding is reached, the phases for data understanding and preparation as well as modeling should follow. This means deciding on the features that should be used as input to the model, and what the output of the model should be.

The factors corresponding to the back-order occurrence need to be converted to features, being a numerical value, that is possible to feed to the deep learning model. To consider when going from factors to features include considering the number of features, in relation to the number of datapoints and the number of hidden layers in the model. If the number of features is too many, they should be reduced to the ones that best describe the output that are to be predicted. Another issue is the use of categorical data as a feature in a deep learning model, where the binary coding approach presented by Potdar et al. (2017) was To decide on the ones the most useful ones, with the feature selection described by Dong & Liu (2018). This way of selecting features is based on correlation of the feature and the backorder occurrence, which means that data point with categorical data needs to be transformed into several input-features and a data point with too many categories means a lot of input-features. All of these things need to be considered, which motivates and leads to the first sub-question:

**Sub-question 2.1. What features should be used for predicting back order occurrence?**

When the input and output features has been decided, the specification that needs to be done is deciding on which type of neural network from chapter 3.5 that is most appropriate for the given problem. The neural network works as the bridge between the input and output features, and as the architectures of deep learning a wide array of approaches to learning from data, it is important to choose the correct approach to a problem (Khan et al., 2019). Since the states in the supply chain are changing over time, and it’s seen as that factors impact themselves and other over time. For example, the demand for a certain week impacts the demand for surrounding weeks as well as stock levels for the surrounding weeks. This means the data is seen as sequential. To handle and use sequential data there is Recurrent Neural Networks (RNN) which use a feedback loop that allows for the network to model dependencies in the data. (Ramakrishnan & Soni, 2018) Therefore a Recurrent Neural Network is seen as the best type network to use in this study. However, the question regarding how to design and implement this Recurrent Neural Network remains, which motivates the second sub question:

**Sub-question 2.2. How could a Recurrent Neural Network be designed for predicting back order occurrence?**

Furthermore, one cannot be sure that a chosen model is optimal. This means that testing is required to make sure that required precision is reached. As described in 0 there are several possible ways of evaluating the performance of the model, and as described in 3.5.7 many assessment metrics exists. The number of non-back orders is far greater than the number of back orders, meaning the data is imbalanced. When working with an imbalanced data set de Santis et al. (2017) describe the two measurements precision and recall, that describe the accuracy and sensitivity of the model. Combining the precision and recall into a single measurement can be done with the F-score that balances the accuracy and sensitivity. (Sasaki, 2007); (Witten et al., 2016) Another metric that is useful is the Area under the precision-recall curve (AUC), as described in 3.5.7 uses the precision and recall at different thresholds to measure the models ability to correctly classify data points.

Given these metrics the performance of the deep learning prediction model needs to be decided to fairly evaluate the prediction. No single metric can be used, but several metrics is needed to fairly evaluate and assess the performance of the model. This gives the third sub-question:
**Sub-question 2.3.** What is the performance of the deep learning model when predicting back orders in the aftermarket of Volvo?

### 4.3 Overview of Research Questions

The aim of study has been divided into four separate areas in order to specify and clarify the purpose. For each area a research question has been formulated which when answered will represent a step in formulating the model. In order to further clarify, the research questions have been broken down into sub-questions. A compiled list of the research questions with their respective sub-questions is presented below.

**Research question 1.** What underlying factors best explain the occurrence of back orders?

- **Sub-question 1.1.** What underlying factors of back order occurrence can be identified from literature?

- **Sub-question 1.2.** What underlying factors of back order occurrences can be identified from the operations of Volvo?

- **Sub-question 1.3.** Based on the literature and the Volvo organization, what factors best explain back order occurrence in the Volvo Supply chain?

**Research question 2.** How could a deep learning model be designed to predict back orders?

- **Sub-question 2.1.** What features should be used for predicting back order occurrence?

- **Sub-question 2.2.** How could a Recurrent Neural Network be designed for predicting back order occurrence?

- **Sub-question 2.3.** What is the performance of the deep learning model when predicting back orders in the aftermarket of Volvo?
5 Methodology

This chapter describes the methodology used in this study. The chapter describes the phases that make up the study as well as a discussion on study credibility divided into validity, reliability and objectivity.

5.1 Methodology Outline

When deciding on a methodology that is to be used for a specific study, there is no scientific approach that relies only on one type of methodology. This creates a need for the author of a study to individually adapt and combine different methods to fulfill the correct type of data collection and to answer research questions. (Bell, 2006) The research methodology applied to this study is derived from the works of several authors. Oskarsson et al., (2013) present a model for conducting change in logistics operations while Kothari (2004) discusses a purely academic research methodology. This along with the CRISP-DM model by Shearer (2006), which provides a highly specific methodology for data mining, make up the foundation of the research methodology utilized in this study. And while these models differ in areas of study, they do share several main points than can create a logical methodology appropriate for this study.

Kothari (2004) discusses a research methodology specifically adapted to academic research. This is divided into seven steps that are executed sequentially and iteratively.

1. **Formulate research problem:** The research problem can initially be broad or general. Later when sufficient understanding has been gathered, the problem can be rephrased into meaningful terms relating to analysis.

2. **Literature review:** The second step is to conduct an extensive literature study in order to learn from previous findings in the area.

3. **Hypothesis:** Third, a hypothesis is formed which delimits the area of research and focuses the research on a narrower scope.

4. **Research design:** Fourth is preparing the research design in order to conduct the study in the most efficient manner possible, in order to yield the maximum amount of information. This also includes determination of the sample design, Kothari describes this as deciding what data to collect before any data collection is conducted.

5. **Data collection:** After this is completed the fifth step, data collection can commence. Through one of several methods, depending on scope, objective, financial and time resources as well as required degree of accuracy.

6. **Analysis:** Once this is completed the data can be analyzed in order to test hypotheses, a useful tool here is to classify the data in meaningful groups in order to more easily conduct analysis.

7. **Interpret results:** Once thorough analysis is concluded, the seventh and last step is to interpret the results of the analysis and attempt to generalize, also a part of this step is to finalize the written report or thesis.

Oskarsson et al. (2013) and their approach to conducting projects related to conducting change in logistics operations is practical and specifically aimed at projects related to logistics. The approach is divided into seven steps.
1. **Background**: The first step is to clarify any prerequisites such as project goals, what parts of the business are included in the study and what resources are available.

2. **Operational analysis**: The second step is to conduct a thorough analysis of the current operations in which is a necessary step in order to conclude with an improvement.

3. **Creating solutions**: The third step is to create suggestions to alternative solutions, i.e. changes that can be made to the operations in order to improve them, this is done closely to analyzing the operations and often done in parallel in such a way that they can be seen as the same step.

4. **Evaluate solutions**: The fourth step is to evaluate these different alternative solutions in order to find the most suitable one. This is done by assessing what changes these would bring and comparing the different solutions and the current operations by use of some certain metrics, such as total cost or service level.

5. **Choose the best solution**: Fifth, the best solution is chosen. This is done on the basis of the quantitative analysis as well as further qualitative analysis to ensure that the chosen solution is compatible with the organization as a whole.

6. **Implement solution**: As this is concluded, the sixth step is to implement these changes. This is by Oskarsson et al. (2013) seen as a difficult and time-consuming step, and it is of great importance that the proposed changes can successfully be implemented as a great solution is worthless if it is unable to be implemented.

7. **Evaluate changes**: Lastly it is time for evaluation, this can be done using the same metrics used when evaluating the solutions to see whether the changes gave the results expected.

As can be seen in these two approaches, they are created with the intent of conducting two different types of studies. One, an approach for academic research and the other, a practical approach to changing logistics operations. They do however share several key points that will be utilized in this study. Both highlight the importance of understanding the problem thoroughly and the same for the area of study. The proposition of alternative solutions and evaluation of these by Oskarsson et al. (2013) are in essence the same as hypotheses and testing of hypotheses as described by Kothari (2004).

In this study the hypothesis step from the Kothari (2004) methodology has been replaced by the study specification. Instead of forming a hypothesis there are several research questions that further specify the study that aim to answer the research problem.

As for the methodology proposed by Oskarsson et al. (2013). The focus of this study has not been to create a practical solution based on the study aim. The creating solutions part is however considered in the discussion portion of this study, where the effects of the study are explored to propose a practical use.
Furthermore, the CRISP-DM Model by Shearer (2000) is a general model for data mining. CRISP-DM structures the data mining process into six sequential and iterative phases.

1. **Business Understanding:** Step one, similar to the models previously discussed, is about business understanding and includes business objectives and background; requirements, risk, cost and benefits; data mining goals and success criteria; producing the project plan.

2. **Data Understanding:** The second phase is about understanding the data. The phase begins with an initial collection of data, describing and exploring the data and lastly verifying the data quality.

3. **Data Preparation:** The third phase is data preparation and contains all steps required to create the final data set from the collected data. The first step is to select what data will be used for analysis, based on its relevance to the data mining goals as well as data quality and constraints on data volume or types. Second is cleaning the data, selecting clean data or analytically filling missing data with estimations. After the data is clean the next step is data construction, creating new records and producing derived attributes, which are new attributes created from combining or dividing old attributes. After data construction, the data is integrated. The data intended for use often come from different tables and records and thus integrating them is a valuable step to ensure usability in a later stage.

4. **Modeling:** The fourth phase is Modeling, selecting appropriate modeling techniques and calibrating features to optimal values. As several techniques may be appropriate and may set requirements on the data, revisiting the data preparation phase may be necessary. An important step in this phase is model assessment, through the use of both data mining metrics and measures according to the business goals.

5. **Evaluation:** The fifth phase is Evaluation. This phase includes a more thorough evaluation of the model and ensuring it achieves business objectives.

6. **Deployment:** The sixth and last step is Deployment. The knowledge generated from the model has to be organized and presented in a way that the customer can utilize it according to their business goals.

Like the methods discussed by these authors, this study will be sectioned in to different phases that occur both sequentially and iteratively. In effect, this means that there will be a clear structure with a predetermined order in which these occur. Due to the nature of developing models for data mining however, as modeling and data processing can hardly be perfected in the first iteration, the results from these phases will be assessed, allowing for adjustments of the model until sufficient results are reached. In Figure 13 below a methodology outline can be seen, showcasing the phases and the steps taken during the study. The figure also consists of an indication of which sub-step the different sub-questions are responded to. The first step is to, like Kothari (2004) formulate a research problem and then further specify this after an initial literature review. After this the empirical study aimed to encompass the business understanding of CRISP-DM and operational analysis of Oskarsson et al. (2013). This is done in parallel with further literature review. After the combined knowledge from the empirical study and the literature review has been synthesized into factors, the data can be collected and prepared. After the data is collected, the modeling phase takes place. A model is developed and
iteratively tested to reach better performance. The phases and their contents are further described in separate chapters below.

![Methodology Outline](image)

**Figure 13 - Methodology Outline with indication of in which phase the sub-questions are responded to**

### 5.2 Introductory Phase

As discussed by the authors mentioned in 5.1, an important first step is to gather sufficient understanding and background to the problem posed for the study. This is what the introductory phase aims to accomplish. With a starting point in the field of back orders in the Volvo aftermarket, discussion along with directives from Volvo determined the aim of the study. In practice, this was carried out by conducting several meetings and discussions with employees with different roles working with the Volvo aftermarket. In this phase five hour-long meetings were conducted, two of which were with a data scientist and a supply chain and analytics expert in the Advanced Analytics team, one with a manager from the Demand and Inventory Planning department as well as two more with the aforementioned people together with the manager for the Advanced Analytics team. The discussions revolved around different proposed aims to determine the viability of different directions. Among these were the chance of success and business value. The problem of course should bring the maximum amount of value to Volvo, but the likelihood of success was also a factor in determining the aim. Also, of interest was the viability of applying deep learning to the different problems, as this was per request by Volvo, a requirement.
5.3 Specification Phase

The main focus of the specification phase is that it is the phase where the general study aim conceived in the first phase is further specified and broken down. As discussed by Kothari (2004) it can be useful to first form a general problem formulation, to later narrow it down. From the original aim, along with a deeper knowledge of the current operations, theory relevant to the area and directives from Volvo, research questions are formed. The research questions were then broken down further into more manageable sub-questions.

To realize operational changes, there is a need to thoroughly understand the current operations (Oskarsson et al., 2013). This means that to sufficiently specify the study it is necessary to describe and analyze the current situation. This part of the specification phase was initialized by an initial hour-long presentation of the Volvo aftermarket and related operations by the manager of the advanced analytics team. From this presentation, the different parts of the organization could be identified in order to judge how they relate to this study and what parts needed further analysis.

An initial literature study was carried out with the purpose of gaining insights regarding the steps needed to take in the creation and application of the deep learning, which was useful when deciding on the appropriate steps needed to take. To gain further knowledge of the business area this is the phase in which the initial literature study commences, as it is of great importance to fully understand the business in which a study of data mining is being conducted (Shearer, 2000). Further support for this can be found from Kothari (2004) who stresses the importance of studying previous findings in related areas. The results of the initial literature study were then used and impacted the formulation of the research questions. Several search keys were found to be relevant for use in this study, among them were: Supply chain, aftermarket logistics, spare parts logistics, machine learning, deep learning and inventory management. The search was conducted by combining these resulting in a large number of searches. Peer-reviewed journal articles were used in first hand, however in some areas such as the combination of machine learning and supply chain yielded very few results and thus conference proceedings were also used as a theoretical framework for this study.

The search engine used was LiU Unisearch. A more detailed description of the results and number of hits from different searches can be found in Appendix A – Literature Study. Furthermore, for areas that were deemed to contain generally accepted methods, such as a general foundation to machine learning or the fundamentals of logistics, books were used. The largest impact from the initial literature study came from increased general knowledge in machine learning and the data mining process from the CRISP-DM methodology presented by Shearer (2000). Specifically, the importance of business understanding before starting the modeling phase.

5.4 Empirical Phase

The primary goal of this phase is to find the factors relating to back order occurrence, and therefore answer to the first research question.

Research question 1. What underlying factors best explain the occurrence of back orders?

Furthermore, this phase also contains an element of conducting further interviews. The purpose of these interviews is two-fold. First and foremost, additional interviews will be needed to aid with the data collection. As the necessary data relates to different parts of the system, interviews with people in different parts of the supply chain will facilitate a better understanding of where relevant data can be found. The theory and information from the empirical study gathered in the specification phase, lays the foundation for answering the first
research question. The results retrieved are used to answer to two sub-questions in the first research question, related to factors that is correlated with the occurrence of back orders.

**Sub-question 1.1.** What underlying factors of back order occurrence can be identified from literature?

**Sub-question 1.2.** What underlying factors of back order occurrences can be identified from the operations of Volvo?

With these sources, both the general literature study and the specific empirical study on Volvo it’s seen as factors correlating with back order occurrence can be found in a good way. The literature and empirical study are described more in detail below.

5.4.1 Further Literature Review

This continuing literature review is more specified with a more concrete focus, which is to find underlying factors of back order occurrence in order to respond the first sub-question. However, the themes of the literature that were reviewed was similar to the initial literature study. In many cases the results and search terms overlapped. The search keys found relevant in this part of the study were: Supply chain, Aftermarket logistics, spare part logistics and the combination of these with machine learning. The search was conducted by combining these resulting in a large number of searches. Peer-reviewed journal articles were used in first hand, however in some areas such as the combination of machine learning and supply chain yielded very few results and thus conference proceedings were also used as a theoretical framework for this study.

The search engine used was LiU Unisearch. Furthermore, for areas that were deemed to contain generally accepted methods, such as a general foundation to machine learning or the fundamentals of logistics and inventory management, books were used. For example, general theory on safety stock. Safety stock is used to prevent stock-outs, to avoid back orders. This is a fundamental part of inventory management and citing books was deemed appropriate. The basis for many parts of machine learning, such as the fundamental understanding for recurrent neural networks is also fundamental knowledge and thus gathered from books. A more detailed description of the literature study with search terms and number of hits, can be found in Appendix A – Literature Study.

5.4.2 Empirical Study

This part of the empirical phase included holding semi-structured interviews with people in some way connected to the Volvo aftermarket Supply chain, with a focus on understanding the operations in order to identify what Volvo see as underlying factors for back order occurrence. The findings from answering the first sub-question were used as a point of direction when handing out the empirical study at Volvo. The empirical study used was carried out via interviews with employees throughout the Supply chain, most of which lasted roughly one hour. Interviewees included people from the functions, Demand and Inventory Planning, Dealer Inventory Management, Material Planning, Advanced Analytics and Back Order Recovery. After these interviews were conducted a total of four initiatives were found. The initiatives contribute in some way towards the collection of factors. Two of the initiatives were from Demand and Inventory Planning and two from Advanced Analytics. The analysis was first and foremost conducted by the use of semi-structured interviews, where the authors of the study had formulated questions relating to the specific area of the interviewee. The interviewee was also encouraged to freely present their area and any other information they might see as relevant to the study. From this a detailed description of the different functions and operations of the Volvo service market logistics could be formed.
When the two first sub-questions have been answered, and combining both the general approach of the literature study and the specific approach of the empirical study the third sub-question was answered.

**Sub-question 1.3. Based on the literature and the Volvo organization, what factors best explain back order occurrence in the Volvo Supply chain?**

The factors extracted from the literature and the empirical study constituted the basis, when answering to the third sub-question. To finalize the list of factors, a qualitative judgement over the collection from the two sources was made. This to find the factors that best explain the occurrences of back orders. This qualitative judgement means that the factors found are reviewed and judged on their importance in the both the literature and the Volvo operations. For example, factors found from the literature study are checked for relevance in the Volvo aftermarket, and if they are found relevant they are taken forward. In the same way, Volvo specific factors are checked for relevance in the literature and the studied system. Factors only specific to either the literature or Volvo operations, but seen as highly correlating to the occurrence of back orders are still taken forward. The main goal of this qualitative judgement is to remove the factors that don’t explain the occurrence of back orders good enough.

With this approach, not only looking at the operations of the studied company, bias of the findings can be reduced. At the same time the reasons found in the literature need to be in the context of the Volvo organization in order to extract data. This combination answers the research question in a holistic way.

### 5.5 Modeling Phase

In this phase the focus lies on the implementation of the model, with deciding on the features, creating the data set and lastly training and implementing the model. The focus of this phase is to answer to the second research-question.

**Research question 2. How could a deep learning model be designed to predict back orders?**

This chapter is further divided in the sub-steps taken.

#### 5.5.1 Data Gathering and Preparation

The first step in applying deep learning is to decide on in and out features and later gather and prepare the data. The first steps in this sub-step aims to answer to the first sub question.

**Sub-question 2.1. What features should be used for predicting back order occurrence?**

To reply to this question, first done according to the underlying factors previously identified from literature and empirical study in the empirical phase. This then poses the question, what steps are required to move from underlying factors to input features from actual data? The method for this is based on the *Data Preparation* step as described by Shearer, (2000) in the CRISP-DM model. Based on the factors deemed interesting and which of these factors can be identified in the available data, the data is selected and gathered from various databases used by Volvo.

Later the data is cleaned, which means making sure that all fields contain intended information and analytically filling or removing empty elements. Without clean data the results can be seen as unreliable (Shearer, 2000). After required data with correct fields are collected, the data can be integrated. This integrating step involves combining data from several tables and records to create a coherent data set usable when training the model. The data set contained one datapoint per feature, period and part number. The data was arranged with the features column-wise and with one column for part number and one for period. A majority of the features were retrieved
from the Volvo-database directly and some needed further processing. One of the data points were calculated and two were transformed to weekly data from monthly data. In this transformation an even distribution over the four-weeks in the period was used.

With creating the data set also comes the final decision regarding input features, through feature engineering. As discussed by Shearer (2000) selecting which data to use is determined by several factors, relevance to the data mining problem, data quality and more technical, data volume and type. As an increase in data volume will increase the runtime of training the model, it is of interest to prioritize data according to its relevance in case the volume has to be reduced to reach a restraint in runtime. Feature engineering involves several concepts revolving around analyzing the usefulness of features in order to select a set of features in order to make the size of the set computationally feasible. (Dong & Liu, 2018) The number of features were deemed as not being to many, which resulted in that further feature-analysis and feature-selection were not needed. If available, additional features could have been used. More features would however increase the runtime and necessarily improved the result. As for the out features, the problem has been formulated as a classification problem solved with supervised learning. This means that the model is trained to recognize the output presented in the training data set. In this case the output is a binary problem and the model is to present the output as either 0 – not a back order or 1 – back order. 52 weeks of data was used, from week 13 of 2018 to week 12 in 2019.

When deciding which part-numbers to include and use in the fitting and evaluation of the model, first only the active part numbers were used. Meaning the ones that still are seen as part numbers that have sales. Part numbers that fulfilled these demands were a total of 80 000 part numbers. To reduce the run-times for the training of the model, only the parts had a back order during the 52-week period were used in the runs. The reason for this decision was because it was seen as a good first step. If the model weren't able to perform well on this, it wouldn't perform well on a more representative dataset. This included in a total of 20 000 part numbers that met this demand. When the model parameter selection was evaluated the number of articles had be further reduced to decrease the run-time. Then from the part numbers with back order, 1000 were randomly selected and was used for smaller investigative training and evaluation runs of the deep learning model.

5.5.2 Deep Learning Modelling & Processing

After the in and out features are decided the question of the actual deep neural network design is posed. This sub-step is related to, and aims to respond to the second sub-question.

**Sub-question 2.2. How could a Recurrent Neural Network be designed for predicting back order occurrence?**

According to directives from Volvo, the model will be a deep recurrent neural network with several hidden layers, thus fulfilling the purpose of applying deep learning. Recurrent layers are useful when there is a temporal or spatial relationship between the data points (Ashwin, Maithili, Gokhale, & Chandavarkar, 2016). This is presumed from the fact that the underlying factors discussed in 5.7.1 at some point in time suggest the occurrence of a back order at a later point in time. The number of hidden layers and neurons will initially be decided from technical expertise by data scientists at Volvo and then further adjusted in later iterations based on results of assessment and run time restraints.

The Modeling Phase has the main goal of implementing an actual model and training it using the data set created in 5.4 Empirical Phase. By directives from Volvo, the model is to be implemented using a deep learning library as well as a library for evaluation metrics. The first step is to choose specifically what type of neural network the model is implemented with. The
basis for this is theory regarding deep learning and neural networks, as well as input from Volvo.

There are three model parameters that need to be considered, the number of nodes and layers in the neural network, as well as the number of epochs, how many times each data point is trained on. The method used when investigating the three model parameters, layers, nodes and epochs, is to first use a baseline with a low number on all three, then one by one the model parameters is increased in order to assess the impact that the model parameters had on the performance of the model.

5.5.3 Model Assessment

This sub-step is tightly connected with the previous since the assessment and modeling is done iteratively. The assessment aims to respond to the third sub-question.

**Sub-question 2.3. What is the performance of the deep learning model when predicting back orders in the aftermarket of Volvo?**

When a first version of the model is created, to ensure that the best possible results are achieved, the model is to be assessed. The model will be tested according to the method described by Witten et al., (2016) in 3.5.6, cross-validation. The data set is divided into two parts based on time, the model is then trained on the first 70% and evaluated on the last 30%. The assessment will be performed using two metrics, F-score and are under the precision-recall curve (AUC_{prc}), described in further detail in 3.5.7.

The first step of the model assessment is to test a different number of layers, nodes and epochs on a smaller set of part numbers to decide on what these should be when running on more part numbers. The assessment will take place after each run of the model where after both the data set and model will be adjusted iteratively in order to reach the best possible result. Lastly, the performance of the created model. As each run of the model is completed, the assessment will be conducted according to the chosen metrics, F-score, Area Under Curve, accuracy, precision and recall. This will give the performance for each iteration of the model. In total there were six iterations of the model. Four with 1000 part numbers and two with 20 000 part numbers.

5.6 Evaluation Phase

The evaluation phase aims to evaluate the results of the study, unlike the assessment performed in the modeling phase, this evaluation regards the actual business value and poses questions regarding how the metrics used for assessment will translate into actual practical improvements and with it, estimations on how a deep learning can be implemented in the Aftermarket Supply chain of Volvo. This phase does not explicitly answer a specific research question, but is rather focused with how the results relate to supply chain logistics in general. The phase attempts to relate the results from the model assessment to a real world scenario. Further, the results will be discussed and how certain factors relate and impact the results. Factors include time frame on the data used, using historical data and assuming an even distribution when converting demand and forecast data from months to weeks.

To conclude, the aim of this phase is to use the resulting evaluation metrics from the modeling phase and to connect this with the actual Volvo aftermarket to get an understanding for what the real-world effects would be. There is also a discussion, relating the discussed real-world to what impact these could have on the operations in the Volvo aftermarket.
5.7 Answering Research Questions

The research questions established and defined in the study specification in chapter 4, provide an outline to the study. In this section of the methodology a summary of the specification earlier given on how these research question will be responded to. A summary of the methods used to answer the sub-questions can be seen in Table 4 below.

Table 4 - Methods for answering sub-questions related to the both research question 2

<table>
<thead>
<tr>
<th>RESEARCH QUESTION</th>
<th>SUB-QUESTION</th>
<th>METHOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. - What underlying factors best explain the occurrence of back orders?</td>
<td>1.1 - What underlying factors of back order occurrence can be identified from literature?</td>
<td>Literature study</td>
</tr>
<tr>
<td></td>
<td>1.2 - What underlying factors of back order occurrences can be identified from the operations of Volvo?</td>
<td>Literature study Empirical study (unstructured meetings)</td>
</tr>
<tr>
<td></td>
<td>1.3 - Based on the literature and the Volvo organization, what factors best explain back order occurrence in the Volvo Supply chain?</td>
<td>Cross checking literature &amp; empirical study</td>
</tr>
<tr>
<td>2. - How could a deep learning model be designed to predict back orders?</td>
<td>2.1 - What features should be used for predicting back order occurrence?</td>
<td>Feature engineered factors from 5.7.1 Technical limitations and requirements</td>
</tr>
<tr>
<td></td>
<td>2.2 - How could a Recurrent Neural Network be designed for predicting back order occurrence?</td>
<td>Literature study Directives from Volvo Iterative model creation after assessment</td>
</tr>
<tr>
<td></td>
<td>2.3 - What is the performance of the deep learning model when predicting back orders in the aftermarket of Volvo?</td>
<td>Assessment according to assessment metrics</td>
</tr>
</tbody>
</table>
5.8 Study Credibility

When it comes to the credibility of a study, Björklund & Paulsson (2014) mentions three measurements that need to be taken into account to achieve a high credibility. These three dimensions of measurement are validity, reliability and objectivity. When conducting a study, one should strive for attaining an as high credibility via a high validity, reliability and objectivity. However the questions regarding this needs to be weighed and balanced against the consumption of resources. (Björklund & Paulsson, 2014) The three measurements are described more in depth below, together with the effect and impact these dimensions had on this specific study.

5.8.1 Validity

Björklund & Paulsson (2014) describes the validity as how well the subject measured is the subject that was supposed to be measured. Things that decrease the validity are systematic and methodological errors. (Björklund & Paulsson, 2014) Lekvall & Wahlbin (2007) describe validity in the same way, and also point out that it’s impossible to with definitely make sure that the result is valid. To accomplish this, another method of measuring that provides truly true results needs to be used, and then that method could be used. (Lekvall & Wahlbin, 2007) In order to increase the validity triangulation can be used, which means using several sources with different perspectives on the object of study. Other ways of increasing the validity when using interviews or surveys include clearly specifying the target group, as well as specifying clear and concise questions. (Björklund & Paulsson, 2014)

The literature study could have a higher level of validity, in the regard that some of studied aspects could have been triangulated better. Some of the investigated fields where quite narrow, and also relatively new which resulted in that the number of hits and valuable article could’ve been higher. In some cases when the number of hits was too low, the search was remade with a broader search term to generate more hits and triangulate better. When it comes to interviews, many of these took place in the study where some things still were unclear. This resulted in that preparatory work for the interviews and preparing clear and concise questions. Our approach was less time consuming, but leaves a bit to be desired when it comes to validity. When the effect is estimated it’s possible that the number of back orders that could be avoided were miscalculated. Even if the model could predict all of the back orders, it’s unreasonable to suppose that all of these could be avoided and a 100 % availability of VOR-orders in Ghent could be achieved. This means that there could be a slight systematic error when assessing the effect.

In a wider perspective, the most important question regarding validity is if the study has found the factors that best explain back orders. The factors that are deemed to best explain the occurrence of back orders are related to stock-outs and the uncertainties in the supply chain that cause these. This is seen as an area that is well understood in logistics research which reduces the risk that any important factors are missed, which provides high validity to this part of the study. To increase validity further, further quantitative analysis of the input-features can be done, this would ensure that the assumed correlations between the factors and the occurrence of back orders is true.

5.8.2 Reliability

The reliability is to which extent the measuring instruments would provide the same results if the investigation would be repeated. (Björklund & Paulsson, 2014) Lekvall & Wahlbin (2007) mentions examples as drivers for low reliability as situational factors such as disturbance in the interview environment and variations. The reliability is theoretically easier to investigate than the validity, and methods include in some ways dividing the results in different pools that are compared. (Lekvall & Wahlbin, 2007) An increased reliability can be achieved with control

41
questions in interviews and surveys, control questions that investigates the same aspects again. Triangulation can also be used to reach a higher reliability. (Björklund & Paulsson, 2014)

One of the most central parts of this study is the determination of factors and input features to the deep learning model. This study is affected by Volvo which means that if the operations of another company was studied the factors not gathered from literature could differ. The same factors from literature would likely have been picked by other authors as the area is well established. The factors from Volvo with little support from literature are however more likely to differ between authors as it is difficult to do this unaffected by personal opinions.

As for the reliability of the results from the deep learning model. As this study uses a well established library to create the deep learning model, anyone with the same data should be able to replicate the model and achieve similar results. This supports a high reliability of the results. The black box nature however, makes it difficult to understand exactly how the results are created. There is also always a small variance in exactly what the model learns from training, even when using the exact same data and model. This reduces the reliability somewhat. This can also affect what happens when new data is added. This study was limited to one year of data, if the next year of data is added the results are highly likely to be improved. There is however no exact way of knowing and albeit small, the possibility that the model performs worse if the next year of data is to be added exists.

5.8.3 Objectivity
The last dimension mentioned by Björklund & Paulsson (2014) regards the objectivity and is described as to which extent values affect the study. To increase the level objectivity in a study every decision made needs to be motivated and clarified in such a way so that the reader can take their own stand in regards of the results of the study. Further when earlier work in some way is reproduced and commented it’s of the highest importance to do so in an unbiased way, meaning no factual errors and no distorted selection of words. (Björklund & Paulsson, 2014)

Like in the reliability the decision of relevant factors from the literature and empirical studies. The authors are affected by Volvo and their view on certain things, which could affect the decisions of final features to use. For example, different interview subjects could propose different factors based on personal opinions or what department they work in. In order to minimize the impact of these all decisions have been clearly motivated in order for the reader to understand why these decisions were made. When conducting a literature study, there is a risk of only searching for facts from earlier work that support the own views, to avoid this a structured literature study was carried out of which the details can be found in chapter 5.3. Based on this it could be argued that the study has a relatively high level of objectivity.
5.9 Academic Demands and Research Ethics

To qualify as an academic report brings up a number of requirements that needs to be fulfilled, which are described below. (Björklund & Paulsson, 2014)

Relates to existing academic knowledge. This is considered with that the knowledge adds on to, and in certain cases replaces previous knowledge. This means that the authors should be aware of, use and establish results relating to already existing theories, models and theories. By using previous results, excessive efforts can be reduced and the knowledge contribution should be made in a way that some relate and deepens or increase the knowledge within that area.

Both general and theoretical interest. This means that even if the field of study is an in-depth study in an organization, there should be general questions in place which could be applied at a general problem for the product to be of academic nature. However, the studied field can never be fully general and to be able to hand out the study limitations and specifications to be able to fit it within the time frame.

Generally accepted scientific methods. What is considered as scientific method tends to change over time, but generally methods that are verifiable and possible to hand out by others tend to be a must when they are to be used in an academic context.

An internally logical functioning whole. This means that there should be a connecting thought throughout the whole material so the reader can easy follow and understand the connection between different parts.

Provides the reader the opportunity to adopt their own positions. By providing a thorough declaration of the material that is used as motivation for selection of method, theory and results the reader can come to her own conclusions. This relates to the need of showcase where models and theories from previous work is used and when it’s the conclusions of the authors.

Something that an academic study needs to consider is the research ethics. Research ethics is concerned with the relation between research and ethics and therefore places demand on the ethical demands on the researcher as well as the direction and implementation of the research. The researcher has a responsibility towards people and animals' taking part in the study, as well as other that might be affected positively and negatively of the results of the study. Therefore, it's important that the work is carried out without the manipulation of external forces and the researcher shouldn’t act within their own personal interests or interests of other stakeholders. (Vetenskapsrådet, 2017). The Swedish government agency Vetenskapsrådet (2017) summarize their recommendations within research ethics down to eight general rules, listed below.

1. You shall tell the truth about your research.
2. You shall consciously review and report the basic premises of your studies.
3. You shall openly account for your methods and results.
4. You shall openly account for your commercial interests and other associations.
5. You shall not make unauthorized use of the research results of others.
6. You shall keep your research organized, for example through documentation and filing.
7. You shall strive to conduct your research without doing harm to people, animals, or the environment.
8. You shall be fair in your judgment of others’ research.

(Vetenskapsrådet, 2017)

These above academic demands and research ethical guidelines has impacted and guided the authors in this study. Relating to the academic demands, all decisions regarding steps in the
study and assumptions have been clearly motivated. The literature used has been clearly accounted for and described, both in terms of contribution to this study and how it was found. This study is carried out in the intersection between machine learning and logistics, and even though the study focuses on the Volvo aftermarket, similar studies could benefit from the results and methodology used, proving the generalization of the study. More focused on the ethic pointers, the study has been carried out in a truthful and open manner, being transparent with that the study has been carried out in cooperation with Volvo, whom have a commercial interest in the results of the study. Further, previous research used has been clearly shown and motivated, as well as been judged fairly. The methodology used throughout the study is clearly accounted for, in order for the reader to easily understand how the study was carried out and judge the credibility of results.
6 Deciding Upon Factors Related to the Occurrence of Back Orders

This chapter aims to respond to the first research question.

**Research question 1.** What underlying factors best explain the occurrence of back orders?

Which means that a number of factors that correlate with and explain the occurrence of back orders will be identified and taken forward in the study. The thought process will be showcased and how the literature and empirical study contributed to the factor collection. Partitioning in the sub headings is corresponding to the sub-questions to this Research Question. So first of the factors extracted from the literature are clarified, grouped and presented. In the later sub-heading the ones from the empirical study at Volvo are presented. The four factor groups that are identified from the literature are used throughout the chapter as guidance and to further clarify the connection between the factors found from the literature and the studied company, and to provide a high-level point of view of the factors. These four factor groups intend to in a comprehensive way illustrate the different parts of the supply chain, and were selected in such a way. In other words, groups illustrating the upstream, focal point and downstream and one group for the factors that can’t be grouped in the other groups. Finally, the factors from these two sources, literature and Volvo, are coupled and synthetized into relevant factors that are taken forward to the later part of the study.

6.1 Factors from Literature

This chapter intends to answer the first sub-question of the first research question.

**Sub-question 1.1.** What underlying factors of back order occurrence can be identified from literature?

With the literature presented in Theoretical Framework as a foundation, this chapter will describe and report on the factors extracted from the literature within the logistic area. A short summary of the main topics, groups, is first presented, followed by a detailed list of all factors and from which authors it was extracted from in Table 5. The groups identified were demand, inventory, supply and article characteristics, in which the literature factors are grouped into below. The grouping was done to give a simpler overview and to easier find redundant factors. The four factor groups can be seen as an illustration of the supply chain, and its different parts.

**Demand**

The occurrence of back orders is dependent on the occurrence of stock outs. A stock out in effect means that the demand for a certain period was higher than anticipated. These so called uncertainties relating to the demand are highlighted by Oskarsson et al. (2013), Robeson et al. (1994) and Bowersox et al. (2002) who mention that some kind of unexpected increase in demand will always exist. Meaning that these authors see the variability in demand as something that needs to be safeguarded against. From another perspective Olhager (2013) stresses the importance of several components that contribute to the total demand and its variation over time. These are factors such as: Trend, a gradual increase or decrease over time and Season, a pattern that reiterates over months, years or weeks. These components aren't related to uncertainties, but are rather factors that should be considered when forecasting future demand since they impact the demand. More direct towards demand is Lahiani et al. (2018) that see demand as a very important factor when forecasts for aftermarket logistics are developed, and that the forecast is important to avoid stock outs. In a study conducted by de Santis et al. (2017) they formulated a machine learning model with the aim to identify materials prone to back order occurrence. In the study they used 21 input features with data from 8 weeks and achieved
a promising result with a high accuracy. The features included both the demand and the forecasted, as well as other features listed in other factor groups.

Dekker et al. (2013) highlight the use of installed base information, meaning the use of information about the products sold by the company that are still in use. In the case of this study this could be factors such as age of the trucks or date for last maintenance. Dekker et al. (2013) identify that the use of installed base information is related to demand and could be used to improve the forecast on an aggregate or disaggregated level, if taken advantage of the data in full. However, both Dekker et al. (2013) and Van der Auweraer et al. (2019) raise that the tailoring needed to take use this information is quite extensive and that new tools for data analytics will drive more research in the area.

**Inventory**

When it comes to stock and inventory the clearest connection with availability is the safety stock, as mentioned by Oskarsson et al. (2013) relating to how uncertainties can be safeguarded against. The uncertainties are what can cause an unexpected stock-out, which is a prerequisite for the occurrence of back orders. The point of keeping safety stock is to safeguard against these uncertainties. Robeson et al. (1994) and Lahiani et al. (2018) also highlight the importance of safety stock to safeguard against variability in demand. de Santis et al. (2017) uses the current inventory, together with decided stock levels as input features in their back order predicting model, and therefore the other dimensions of the inventory come in play. The two other main questions beyond how to safeguard against uncertainties lifted by Oskarsson et al. (2013) are when products should be ordered, how much should be ordered each time. Therefore, it’s found relevant to include the cycle and current stock mentioned by Oskarsson et al. (2013) and Lahiani et al. (2018) even though the connection with availability isn’t obvious, but motivated with the back order prediction model formulated by de Santis et al. (2017).

Kennedy et al. (2002) present an overview of logistics related to spare parts and in addition to the general factors from other authors, they present a perspective specific to spare parts. Maintenance requiring spare parts are either planned or unplanned. The planned maintenance allows for parts to be ordered just in time. The unplanned are where stock-outs and back orders are more likely to create significant costs. Thus, strengthening the idea that uncertainties in demand and delivery affect the occurrence of back orders.

**Supply**

The uncertainties mentioned earlier also exists within the supply, as mentioned by Oskarsson et al. (2013) and Robeson et al. (1994). Highlighted by Oskarsson et al. (2013) is delays and shortages in deliveries, corresponding to lead time and delivery precision. Robeson et al. (1994) divides the supply related uncertainties into supplier and transportation related. Bowersox et al. (2002) uses performance of the focal company instead of supplier related issues in their break down of uncertainties. However, in this study factors relating to the supplier are used rather than the warehouse performance since these are seen as difficult to quantify and use as a predicting factor. In the study by de Santis et al. (2017) where they attempted to identify materials prone to back order occurrence external factors related to supply such as lead time and a more general source performance.
**Article Characteristics**

Bacchetti & Saccani (2012) discuss the complexity of spare-parts inventory management and identify several factors that can be used to classify spare parts. Based on the characteristics of the actual part as well as factors such as demand and part life cycle. They argue that a classification according to several characteristics can be useful in predicting the demand of spare parts. So, the characteristics of the part can be related with the demand of the same. There are other correlations between the parts characteristics and factors, for instance the price affects the inventory decisions and levels Oskarsson et al. (2013) Since a correlation between the factors seen as relevant could be identified the part characteristics were included as factors as well. Even though for instance the stock level alone could be used as a factor, the price that affects that affects the stock level is included. This because the correlation could be greater with the price, than with the stock level.

Below in Table 5 all of the factors from the literature study are presented, grouped and marked with the corresponding sources and authors used.

**Table 5 - Compiled factors correlating with back order occurrence found from literature study**

<table>
<thead>
<tr>
<th>FACTOR GROUP</th>
<th>FACTOR</th>
<th>OSKARSSON ET. AL.</th>
<th>LAHIANI ET. AL.</th>
<th>KENNEDY ET. AL.</th>
<th>BACCHETTI &amp; SACCANI</th>
<th>SANTIS ET. AL.</th>
<th>DEKKER ET. AL.</th>
<th>VAN DER AUWERAER ET. AL.</th>
<th>ROBESON ET. AL.</th>
<th>BOWERSOX ET. AL</th>
<th>OLHAGER</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEMAND</td>
<td>Demand</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Forecasted Demand</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trend/seasonality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Demand Variability</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Installed Base Information</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INVENTORY</td>
<td>Cycle Stock</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Safety Stock</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Current Stock Level</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>SUPPLY</td>
<td>Lead time</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Delivery Precision</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Supply characteristics / uncertainty</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARTICLE CHARACTERISTICS</td>
<td>Part Criticality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Part Reliability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Part Life Cycle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Part Weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Repair Efficiency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Part Specificity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Price</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
6.2 Factors from Empirical Study

In this section focus on the specificity of the Volvo operations, and the sub-question that the chapter intends to reply to is as follows.

**Sub-question 1.2. What underlying factors of back order occurrences can be identified from the operations of Volvo?**

Studies of the Volvo operations has resulted in finding a total of four initiatives related to back order occurrence and stock levels. These initiatives give factors to take forward in the study. Beyond that some other factors where collected from general interviews with the employees. These factors are presented throughout the chapter, finalized with a compiled table as a summary. The factors extracted from the different initiatives are grouped accordingly to the groups found from the literature and overlapping factors with different naming are handled as one.

**Initiative 1 & 2**

First off, there are two initiatives in place where the back orders at the Central Distribution Center in Ghent is explored in order to draw conclusions in what drives back orders. The Demand and Inventory Planning team has developed and uses these, where one is for all orders, while one focus only on the VOR orders. These provide a good picture of the reasons seen as back order driving at Volvo. The output of the initiative is an excel-file with a set of rules for every possible reason in order to evaluate if that is the reason for the back order. The rules are checked one by one until a match is found with a waterfall logic. If no match is found the back order is marked as "unknown".

The two initiatives have very much the same reasons, with some discrepancies with how the reasons are traced and with naming. The second initiative concerned with VOR-orders have more reasons and many of them are very specific and based on comments from the material planner, instead of produced automatically from other data points. Reasons found from the initiatives include supplier related, such as quality issues or “no purchasing contract”, and mistakes made in ordering. Other reasons include capacity issues in the Central Distribution Center in Ghent, unexpected increases in demand and other specific reasons related to specific parts of the supply chain not studied, for instance the remanufacturing flow.

Not all the back orders can be identified and therefore some are marked as “unknown”, which means that these initiatives cannot be used as basis by itself. No extensive analysis of the frequency of these reasons was made as a relatively large number of back orders were to be taken forward, rather than using the ones seen as most common by the analysis from the Demand and Inventory Planning team.

**Initiative 3**

Another initiative that was used in the collection of factors was a machine learning initiative with the aim of forecasting the overall availability in the Ghent warehouse. Back order is, as earlier stated, closely connected with the stock levels and availability which motivates this initiative as relevant and worth using as a basis of factors. The machine learning model is using aggregated data for all part numbers and the total quantity, since the availability prediction is not made on part number level. Factors used in this initiative is mostly what the quantity is, both inbound and outbound of Ghent. The quantity is broken down to quantity that has been picked, are reserved in some way, the ordered value, the quantity shipped from the supplier and the number of late in-deliveries. Other factors used are the number of that are classified as “fast moving” and turnover rate.
**Initiative 4**

The fourth and last initiative that was used was a machine learning model for back-order prediction that was developed during a problem-solving team event with a short and fixed time frame. The model is predicting back order likelihood on a part number level. Some examples of the features used as input are the accuracy of the forecast, the delivery precision for that part along with lead time and the frequency of in-deliveries. Beyond these, the actual demand, the target availability and the price of the part were included.

**Other**

Beyond these four initiatives, a product group attribute for each part was found during the empirical study, which are accounted for below.

- **Product group**
  The parts are divided into several product groups depending on which kind of part it is. The motivation behind this factor is that there could be a correlation between the product group and characteristics that leads to the part being more or less back order prone.

Even though this factor wasn't identified from any of the four initiatives it's seen as a defining article characteristic, and the article characteristic is seen to have a correlation with the occurrence of back order occurrence, which motivates that it’s taken forward from the empirical study.

**Combining the initiatives**

If the factors from the initiatives are studied in regards to how they relate to the previous found groups, it can be seen that generally the factors from all of the four initiatives more or less can be categorized into each of the four groups. The first and second initiative have a very similar distribution over the groups, since they are very similar. The difference between these two is that initiative 2 has more factors in the article characteristics group. In general, the article characteristics group is the group that can be found from the fewest initiatives. The third initiative has most of it factors in the inventory group, since the focus of the initiative is the availability it’s reasonable. On the contrary the fourth initiative has only one factor grouped in the inventory group and most of the factors from this initiative are labelled as demand and supply factors.

Compiling and grouping these different initiatives and inputs to a single list of factors, which can be found in
Table 6. Some of the factors found are bundled up to get rid of duplicating and overlapping factors. For instance, the quality related factors and the current stock related factors are each bundled together into one factor each. Further factors that were bundled was order schedule related, supplier contract issues, and factors related with supplier issues. The remanufacturing flow related factors from the back orders of VOR-orders are bundled together as one, since the remanufacturing flow is not considered in this study. Output features from the two machine learning models are not included.
Table 6 - Compiled factors correlating with back order occurrence found from empirical study at Volvo

<table>
<thead>
<tr>
<th>FACTOR GROUP</th>
<th>FACTOR</th>
<th>INITIATIVE 1</th>
<th>INITIATIVE 2</th>
<th>INITIATIVE 3</th>
<th>INITIATIVE 4</th>
<th>OTHER FACTORS</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEMAND</td>
<td>Demand (Orders)</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Faulty Forecast</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Large Order</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Seasonality</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Passive part</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>First Hit / Slow Moving</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INVENTORY</td>
<td>Turnover Rate</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Current Stock</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parts Below Stock</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Part Blocked in Warehouse</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Late Binning</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Target Availability</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Manual Changes to Delivery Schedule</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Back Orders</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>SUPPLY</td>
<td>No Purchasing contract</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quality Related (QJ)</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Change of Supplier</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Late In-deliveries</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Supplier can't produce</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lead Time</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Delivery Frequency</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARTICLE CHARACTERISTICS</td>
<td>Cost of Part</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Product Group</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Segmentation (Fast &amp; Critical)</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Not intended as spare part</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Special Order</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Supersession Handling</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Remanufacturing Related</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>
6.3 Synthetization of Factors

Now that factors from both the literature and Volvo has been collected, these are to be synthesized down to a list reflecting both these parts. This chapter aims to do this specifically, and to respond to the third and last sub-question listed below.

**Sub-question 1.3. Based on the literature and the Volvo organization, what factors best explain back order occurrence in the Volvo Supply chain?**

As a foundation, factors from the literature and the empirical study as presented in Table 5 and
Table 6 above. The factors from these two different sources are merged and compiled into a single list of consisting of the factors that best explain back order occurrence. Our thought process is described and our decisions when factors are excluded or compiled is motivated below. Throughout the chapter the previous grouping of factors is used.

**Demand**

**Demand**

Vital factor identified in both literature and from the Volvo initiatives. This factor is one of the fundamental factors since demand in the form of orders is the direct reason for back orders, and an unexpected increase in demand leads to stock outs over time. Therefore, this factor is taken forward.

**Forecasted demand / Faulty Forecast**

Identified from the literature as vital to handle fluctuations in demand, and identified at Volvo as faulty forecasts leads to back orders. This factor is taken forward.

**Demand variability**

Tightly connected with the demand which is seen as an important factor. A higher variable demand is harder to predict which increase the risk of back orders. Therefore, this factor is also taken forward.

**Special & Large Order**

These factors are manually updated at the Volvo-initiative that they are and are therefore not seen as particularly interesting. Large order has a correlation, but this is seen as covered under demand and the factor special order are deselected since this study isn’t focused on these.

**Installed base information**

Even though seen as a factor that is correlating with back orders, since it provides information about spare part demand. However, the data available at Volvo is restricted and not adequate for this purpose, which leads to that this factor is not taken forward.

**Inventory**

**Cycle Stock**

Only found from the literature where the connection between availability and cycle stock is not totally evident, but is motivated through the back order prediction study by de Santis et al. (2017) and therefore the occurrence of back orders. Further the cycle stock is related with inventory management and stock levels in general. Therefore, the factor is taken forward.

**Safety stock & Parts below safety stock**

For the safety stock the connection with availability is clear, since the purpose for the safety stock is to safeguard against uncertainties. The factor safety stock can also be found in one earlier initiative at Volvo where "Parts below safety stock" was used. Here the factor "safety stock" is taken forward.

**Current stock level**
Identified from both literature and the Volvo operations, along with a connection with availability since. If the current stock level is relatively low to the cycle stock and safety stock the risk of stock outs and back orders is increased. Therefore, this factor is taken forward.

**Trend and Seasonality**

Trend and season are found from the literature as a factor correlating with the possibility to foresee the demand. From Volvo it’s only the seasonality that is seen as a factor. The trend and seasonality per part is handled within the forecast, but the factor seasonality can also be seen as relating to busy periods where all of the orders are hard to solve. Which leads to that the factor seasonality is taken forward, and not trend.

**Parts blocked in warehouse & First hit / Slow moving & Schedule not followed correctly & Not intended as spare part**

These factors found from the empirical study at Volvo are seen as factors found afterwards, and not factors that can be used to foresee the back-order occurrence. Therefore, these factors are not taken forward.

**Late Binning**

When a part hasn't been put on the shelf, it isn't available to be picked and to fulfill orders. That a part isn't available at the same moment it arrives in Ghent needs to be considered, which leads to that the factor is taken forward.

**Turnover rate**

This factor provides information about how long time it takes to go through the stock, which is seen as relevant as it indicates how fast moving the article is related to its stock level. Therefore, the factor is taken forward.

**Target availability**

The target availability set for each part is in a way the number of stock outs and back orders that is planned for. The correlation is very high and is taken forward.

**Supply**

**Lead time**

This factor has been found both from the literature and from the empirical study. Along with the motivation that longer lead times from the supplier lead to higher risk of back orders, since the time it takes from order to delivery is longer, the factor is taken forward.
Delivery precision

A supplier with low delivery precision increases the risk of getting late deliveries which introduce the risk of stock outs and back orders. From the Volvo late deliveries (Arrears) is seen as one of the factors of back order occurrence. The factor is taken forward.

Supply characteristics and uncertainty & No purchasing contract & Quality related

"Supply characteristics and uncertainty" is identified as a factor from the literature, and the corresponding factors found from Volvo are “No Purchasing Contract” and “Quality related”, and it’s these factors that are taken forward. No purchasing contract leads to stock outs if the problem persists, and quality related means that the part tend to have more problems than other parts which increases the risk of stock outs.

Change of supplier & Supplier can't produce

These factors from the empirical study are specific and not extracted from any specific data points but rather from manual updates by the material planner. This study aims to create a model that uses existing data points, which leads to that these factors aren’t taken forward.

Delivery frequency

A higher delivery frequency leads to that unexpected increases in demand can be mitigated against better since the number of in-deliveries is larger. At the same time a high delivery frequency leads to number of times where the cycle stock is zero is more numerous, which lead to that the inventory is sensitive to uncertainties. So even if it's uncertain if the risk decreases or increases with a higher delivery frequency, there is a basis to a correlation to back order occurrence. The factor is taken forward.

Article Characteristics

Price

The price is seen as one of the characteristics of spare parts and an expensive product has lower stock levels which is related to risk of back orders. Therefore, the factor is taken forward.

Part criticality

From the literature and from Volvo as the segmentation. The criticality impacts the stock levels at Volvo which motivates that the factor is taken forward.

Part reliability / Passive

A more reliable part has a lower demand, since the part doesn't need to be changed as much in vehicles. However, as identified from Volvo a part being marked as passive can get a higher demand than expected because of a faulty marking. Therefore, this factor, named as passive, is taken forward.

Part Life Cycle

When part is in a later life cycle a larger amount of back orders are planned for, which motivates that this factor is taken forward. From Volvo the life cycle is reflected in the segmentation.
Part Weight

Even though the weight of a part might impact how large the stock for this part will be, no clear indications of this could be seen at Volvo. Therefore, this part is not taken forward.

Repair efficiency & Part Specificity

Seen as possible factors from the literature, but no indications of being factors correlating with back orders at Volvo where found. The factors are not taken forward.

Supersession handling

That an article is superseding another might lead to that an order of the superseded article is placed which leads to a back order. However, this is seen as to complex and specific, which means that this factor isn't taken forward.

Remanufacturing related

These factors from the reasons of VOR-back orders where seen as to specific for the aim of this study, and factors directly concerned with the remanufacturing flow was not taken forward.

Product Group

Since the product group contains parts with similar characteristics the product group gives a lot of information about the part and the stock levels for the part would probably behave similar to those in the same product group.

The list of finalized factors that is seen best explain the occurrence of back order can be seen in Table 7 below. These are the factors taken forward in the study. As explained in 5.4.2, the final selection of factors was done on a qualitative basis. The selection was based on how much support was found in both literature and according to Volvo.
<table>
<thead>
<tr>
<th>FACTOR GROUP</th>
<th>FACTOR</th>
<th>LITERATURE REVIEW</th>
<th>EMPIRICAL STUDY</th>
<th>MOTIVATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEMAND</td>
<td>Demand</td>
<td>X</td>
<td>X</td>
<td>The direct reason leading to back orders.</td>
</tr>
<tr>
<td></td>
<td>Forecasted demand</td>
<td>X</td>
<td>X</td>
<td>A good forecast reduces the risk of back orders.</td>
</tr>
<tr>
<td></td>
<td>Demand Variability</td>
<td>X</td>
<td></td>
<td>Higher variability, harder to predict demand.</td>
</tr>
<tr>
<td></td>
<td>Seasonality</td>
<td>X</td>
<td>X</td>
<td>Some periods are busier than others, leading to problems handling all of the orders.</td>
</tr>
<tr>
<td></td>
<td>Passive</td>
<td>X</td>
<td>X</td>
<td>A passive part can be falsely marked, leading to unexpected demand.</td>
</tr>
<tr>
<td>INVENTORY</td>
<td>Cycle stock</td>
<td>X</td>
<td></td>
<td>Connection with availability.</td>
</tr>
<tr>
<td></td>
<td>Safety Stock</td>
<td>X</td>
<td>X</td>
<td>Connection with availability.</td>
</tr>
<tr>
<td></td>
<td>current Stock Level</td>
<td>X</td>
<td>X</td>
<td>Connection with availability.</td>
</tr>
<tr>
<td></td>
<td>Target availability</td>
<td>X</td>
<td></td>
<td>Lower target increases the number of back orders.</td>
</tr>
<tr>
<td></td>
<td>Late Binning</td>
<td></td>
<td>X</td>
<td>Not all parts in the Ghent warehouse have been put on the shelf.</td>
</tr>
<tr>
<td></td>
<td>Turnover rate</td>
<td></td>
<td>X</td>
<td>Indication of how fast moving the article is, related to its stock level.</td>
</tr>
<tr>
<td>SUPPLY</td>
<td>Lead Time</td>
<td>X</td>
<td>X</td>
<td>Long lead times gives higher risk of back orders.</td>
</tr>
<tr>
<td></td>
<td>Delivery Precision</td>
<td>X</td>
<td>X</td>
<td>Late deliveries increase the risk of back orders.</td>
</tr>
<tr>
<td></td>
<td>Delivery Frequency</td>
<td></td>
<td>X</td>
<td>Faster reaction to increased demand. Also more occasions with low stock level.</td>
</tr>
<tr>
<td></td>
<td>No purchasing contract</td>
<td>X</td>
<td>X</td>
<td>No contract leads to shortages of stock.</td>
</tr>
<tr>
<td></td>
<td>Quality Related (QJ)</td>
<td>X</td>
<td>X</td>
<td>Suppliers with a higher tendency of quality problems leads to higher risk of back orders.</td>
</tr>
<tr>
<td>ARTICLE CHARACTERISTICS</td>
<td>Price</td>
<td>X</td>
<td>X</td>
<td>Impacts stock levels and availability.</td>
</tr>
<tr>
<td></td>
<td>Part criticality</td>
<td>X</td>
<td>X</td>
<td>Impacts the stock levels, and therefore the availability.</td>
</tr>
<tr>
<td></td>
<td>Part Life Cycle</td>
<td>X</td>
<td>X</td>
<td>Later in the life cycle leads to a higher number of planned back orders.</td>
</tr>
<tr>
<td></td>
<td>Product group</td>
<td></td>
<td>X</td>
<td>Contains information about part characteristics.</td>
</tr>
</tbody>
</table>
Designing the Deep Learning Model

This chapter aims to respond to the second research question.

Research question 2. How could a deep learning model be designed to predict back orders?

This question is divided into three parts which will be individually handled in three sub-chapters. The first step is to create the features to be used for training the model, this is done on the basis of the factors decided on in chapter 5.9. The final input features are explained along with the reasoning behind their choosing. The next part is to describe the design and development of the model, based on theory collected about deep learning together with instructions from Volvo. Lastly the performance of the model is presented. A visualization of the model with the steps from features and model parameters, through the neural network to the output can be seen in Figure 14 below.

![Figure 14 - Visualization of the model and the different kinds of inputs and outputs](image)

**7.1 Deep Learning Model Features**

This section aims to answer the first sub-question of the second research question.

Sub-question 2.1. What features should be used for predicting back order occurrence?

Based on the factors synthesized in 6.3 the features are created from data from the Volvo database. Some factors are translated directly into their own features, some factors are merged into a single feature and two are completely excluded. The numerical representation and if the feature are dynamic or static, i.e. if feature-value is changing for each period or if it stays the same. Whether the feature is dynamic or static is stated in the summarizing Error! Reference source not found. below. The decision regarding if the feature is to be dynamic or static is based on how often the value of the datapoint change, and how the data is available from the Volvo database. The factor-grouping used previously is used for the features as well.

Demand

The actual demand for a part for the given week, represented by a positive integer. The variation of the demand feature over time is also comprises the demand variability and seasonality factors. The seasonality is however limited as the data set is comprised of data from one year. This datapoint is only available per month and was converted to weeks with an assumption of an even distribution. This means that the only true variability is between months which likely affects the results negatively.
**Forecasted Quantity**

The forecasted quantity feature is the forecasted demand for the following week, represented by a positive decimal number. Like the demand, this datapoint is only available per month and was converted to weeks with an assumption of an even distribution.

**Passive Part**

The passive part feature is represented by a 0 or 1, where a 1 means that the specific part is marked as passive. This datapoint is based on an assessment by Volvo, indicating if the part is will be needed to be changed a lot or not.

**Inventory**

**Safety Stock**

The safety stock feature is a positive integer which represents the safety stock level of each part number.

**Available stock**

The available stock feature is a positive integer representing the quantity of each part number on the shelves available for order in the current week.

**Receiving Quantity**

Receiving quantity is a feature which represents the quantity of inbound stock for the current week. Stock that is on the warehouse premises but not yet put on the shelves. It is represented by a positive integer. This also encompasses the late binning factor as the receiving quantity shows parts not yet available for picking.

**Economic Order Quantity (EOQ)**

The EOQ feature is a positive integer which represents the quantity of a certain part that is ordered each time an order is placed. This represents the cycle stock factor since this is the stock level that is ordered up to each time.

**Target Availability**

The target availability is a decimal number between 0 and 1 which represents the targeted availability for a certain part. The availability is decided on a segment-level and mapped based on the segment of the article.

**Supply**

**Lead Time**

Lead time is a feature represented by a positive integer which shows the, between Volvo and the supplier, agreed upon lead time for a specific part, in weeks. The lead time used is the total time until the part is put on the shelf, and not only until it has been delivered in Ghent.
Delivery Precision

The delivery precision feature is a decimal number between 0 and 1. The number represents the average delivery precision from a specific supplier over the last year. Extracted deliveries per week for each supplier, along with the number of which were too early or too late. These were then used to calculate the average percentage of deliveries on time.

Delivery frequency

Delivery frequency was not found in the Volvo database and thus not included as a feature.

Purchasing Contract

The purchasing contract feature shows if there is a purchasing contract in place or not for a specific part, represented by a 1 if a contract exists or by a 0 if one does not.

Known Quality Issue

Represented by either a 0 or 1, this shows if a part has been marked to have known quality issues.

Article characteristics

Price

The price feature is a positive integer which represents the unit purchasing price for the given part in SEK.

Segmentation

The segmentation is divided into four separate features each represented by a 0 or 1. Where a 1 shows that a part belongs to one of the segments fast, medium, slow or critical. The segmentation feature is extracted from the more detailed segmentation, so that the number of columns to keep the data set at a reasonable size. This feature captures the part criticality and part life-cycle factors.

Product Group

The product group factor is not included as a feature as the data found is categorical data divided into around 50 categories. Each category would have to be represented by its own column in the data set, the number of categories would then increase the size of the data set more than three-fold, which is not acceptable.

The features taken forward can be seen summarized in Table 9 below. The table shows which factor the feature cover and how the feature is represented in the model.
Table 9 - Compiled list of deep learning model input features

<table>
<thead>
<tr>
<th>FACTOR GROUP</th>
<th>FEATURE</th>
<th>NUMERICAL REPRESENTATION</th>
<th>DYNAMIC OR STATIC</th>
<th>FACTOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEMAND</td>
<td>Demand</td>
<td>Positive integer</td>
<td>Dynamic</td>
<td>Demand, Demand Variability, Seasonality</td>
</tr>
<tr>
<td></td>
<td>Forecasted Quantity</td>
<td>Positive Decimal number</td>
<td>Dynamic</td>
<td>Forecasted Quantity</td>
</tr>
<tr>
<td></td>
<td>Passive Part</td>
<td>1 or 0</td>
<td>Static</td>
<td>Passive Part</td>
</tr>
<tr>
<td>INVENTORY</td>
<td>Safety Stock</td>
<td>Positive integer</td>
<td>Static</td>
<td>Safety Stock</td>
</tr>
<tr>
<td></td>
<td>Available stock</td>
<td>Positive integer</td>
<td>Dynamic</td>
<td>Current Stock Level</td>
</tr>
<tr>
<td></td>
<td>Receiving Quantity</td>
<td>Positive integer</td>
<td>Dynamic</td>
<td>Current Stock Level, Late Binning Cycle Stock</td>
</tr>
<tr>
<td></td>
<td>Economic Order Quantity</td>
<td>Positive integer</td>
<td>Static</td>
<td>Target Availability</td>
</tr>
<tr>
<td></td>
<td>Target Availability</td>
<td>Positive Decimal number</td>
<td>Static</td>
<td>Target Availability</td>
</tr>
<tr>
<td>SUPPLY</td>
<td>Lead Time</td>
<td>Positive integer</td>
<td>Static</td>
<td>Lead Time</td>
</tr>
<tr>
<td></td>
<td>Delivery Precision</td>
<td>Positive Decimal number</td>
<td>Static</td>
<td>Delivery Precision</td>
</tr>
<tr>
<td></td>
<td>Purchasing Contract</td>
<td>1 or 0</td>
<td>Static</td>
<td>No Purchasing Contract</td>
</tr>
<tr>
<td></td>
<td>Known Quality Issue</td>
<td>1 or 0</td>
<td>Static</td>
<td>Quality Related (QJ)</td>
</tr>
<tr>
<td>ARTICLE CHARACTERISTICS</td>
<td>Price</td>
<td>Positive integer</td>
<td>Static</td>
<td>Price</td>
</tr>
<tr>
<td></td>
<td>Segmentation</td>
<td>1 or 0</td>
<td>Static</td>
<td>Part Criticality, Part Life Cycle</td>
</tr>
</tbody>
</table>

7.2 Model Design

This section aims to answer the second sub-question of the second research question.

**Sub-question 2.2: How could a Recurrent Neural Network be designed for predicting back order occurrence?**

As described in 4.2, the chosen model is a recurrent neural network. The architecture of a Recurrent Neural Network was deemed appropriate as the occurrence of back orders over time was seen as a time series. Time series are just the type of data that recurrent neural networks are created for and as such it was deemed most appropriate both by studying literature and by feedback from Volvo.

The model is comprised of a number of LSTM-layers and an output layer, which creates the final output. As discussed in 3.5.2, Long short-term memory (LSTM) is a recurrent neural network architecture aimed to allow for use with large time series. LSTM was selected partly due to its popularity in time series prediction, but also based in the input from Volvo where LSTM has previously been used successfully. As discussed in 3.5.2, an alternative to LSTM is the Gated Recurrent Unit (GRU). As these are both similar and there is no method for determining which is more appropriate, LSTM was arbitrarily selected as testing both to determine which is better is outside of the scope of this study. The use of an output layer is necessary in all deep neural networks for the model to give the correct output for the classification problem this study has been formulated as, classifying if a back order will occur or not for a given part.
Further design choices are activation function and optimizer. As creating an optimal neural network is not in the scope of this study, the activation function hyperbolic tangent (tanh) and optimizer Adam were arbitrarily chosen instead of other popular alternatives mentioned in 3.5.2.

When the types of the layers are decided the big design question for a deep neural network is the size. Depth, meaning the number of layers and width, the number of nodes in each layer. The neural network is being trained to approximate a function that will describe the different classes of the classification problem. A neural network with at least one hidden layer and enough nodes can approximate any function to any desired amount of error greater than zero (Goodfellow, Bengio, & Courville, 2016). For a function that can be learned using one hidden layer and several nodes, a solution with more layers and fewer nodes can be more efficient (Brownlee, 2018). And while there is no method for determining the optimal design, Goodfellow et al. (2016) state that a deeper network generally performs better than a wider one. As such, a common method is to systematically experiment with the number of layers and nodes. In this study the experiment was carried out via first a baseline iteration, and then followed up with three iterations where one of the parameters was increased at a time. After that two parameters were increased together to get an understanding of there was an aggregated effect. The different designs of the network have been tested and their performance along with an analysis of the results is presented in 0.

### 7.3 Model Performance and Results

This section aims to answer the last sub-question of the second research question.

**Sub-question 2.3. What is the performance of the deep learning model when predicting back orders in the aftermarket of Volvo?**

The answer to this question will come from detailing the results of several iterations of the training and evaluation of the model. First the connection between output and the performance of the model, followed by a description of the impact of the model parameters and finally the actual results of the model is described.

#### 7.3.1 Output and Performance Metrics

To facilitate a simpler understanding of the model and its results, this section aims to clarify the steps needed to go from the actual output of the model to the metrics measuring the results. Figure 15 below, is a visualization of the model and it's input in the form of model parameters and features and its output in the form of probability and binary prediction.

![Figure 15 - Visualization of the model and the connection to the performance metrics used in the assessment of the model](image)

The steps from probability output to binary and lastly to performance is divided into three sections below, and more thoroughly explained.
The Probability Output of the Model

The aim of the model is to create a binary prediction for the occurrence of back orders, the initial output however is not given in binary. The initial output from predictions created by the model is a decimal number between 0 and 1, indicating the likelihood that the given part number will have a back order the following week. The decimal number then has to be converted to either a 0 or 1 to give the final prediction of whether the given part number be classified as a back order or not. Whether a given decimal number output is classified to a 0 or 1 is decided by a threshold.

The Classification Threshold

The threshold is the limit for what numbers get converted to either a 0 or 1. The threshold is a decimal number between 0 and 1, where any output equal to or higher than the threshold is classified as a 1. Any number lower than the threshold is classified as a 0. In practice, this means that the threshold decides how certain the model has to be for a prediction to be considered a back order. A high threshold requires a high certainty, the output from the model has to be close to 1 for to be classified as a back order. In practice, this means that a high threshold is less likely to falsely classify a part number without a back order as having one. The drawback however is that this also makes the model more prone to missing back orders, due to the high certainty required to be classified as a back order.

This can be further illustrated using the error matrix displayed below in Figure 16. The error matrix contains all the different outcomes and compares the true values to the predicted values which produces four fields that describes how changing the threshold affects the results. The four classes are:

1. **True Positive** – The model correctly predicted a back order
2. **False Positive** – The model falsely predicted a back order
3. **False Negative** – The model falsely predicted a non-back order
4. **True Negative** – The model correctly predicted a non-back order

![Error Matrix](https://via.placeholder.com/150)

*Figure 16 - Error Matrix visualizing the output and the correctness of the predictions*
A high threshold requires a high certainty from the model that a back order is correctly predicted. This results in a smaller proportion of both true positive classifications and false positive classifications, and a larger proportion of false negative and true negative classifications.

A low threshold gives the opposite, the required certainty is low. This results in a larger proportion of both true positive and false positive classifications, and a smaller proportion of false negative and true negative classifications. The aim when deciding the threshold is to get many true positives with few false positives, which will also produce many true negatives and few false negatives. The effects from adjusting the threshold is visualized in Figure 17 which shows the effects of a lower and higher threshold.

![Higher Threshold vs Lower Threshold Error Matrix](image)

*Figure 17 - Error Matrix, with indications of how the proportions in the compartments are affected by the threshold*

**Performance Metrics Used**

The metrics are used to determine the performance of the model. The metrics most relevant to this study is precision and recall. Where in the context of this study, recall measures how many of the actual back orders that were classified correctly and precision measures how many of the instances classified as back orders were actual back orders. To connect the metrics to the error matrix, precision equals to the number of true positives divided by the sum of true positives and false negatives and recall equals to the number of true positives divided by the sum of true positives and false negatives. A visualization of this can be seen in Figure 18 below.
Furthermore, considering either precision or recall without the other gives little usable information. That is where the F-score, discussed in 3.5.7 finds it use. F-score is the harmonic mean of precision and recall and gives a single measure which can be easier to interpret compared to considering both precision and recall. The usefulness of the F-score comes from the fact that a model will not get a high F-score if either precision or recall is low.

Lastly the Area Under Curve (AUC) metric is a metric from plotting the precision and recall achieved at different thresholds. The Area Under Curve can be seen as a general measure of total model skill and is further discussed in 3.5.7.

7.3.2 Evaluating the Model Parameters

The first step in the modelling was concerned with deciding the impact different model parameters had on the output and the performance of the model. Model parameters investigated are being size of the neural network when it comes to number of layers and number of nodes per layer. The number of epochs that means how many times the model train on a dataset. In all of the iterations related with assessing the impact of the model parameters, the same randomly selected 1000 part numbers were used, with testing on a smaller subset of a 100 part numbers of these.

**Iteration 1 – Baseline Setup**

First a baseline run of the model was done, which can be seen as Iteration 1 in Table 8 below. As per the performance metrics, described more in detail in the chapter above, the performance can be seen and analyzed. The percentage of correctly predicted back order occurrences for this first iteration is at around 90%, given the accuracy metric. This might seem like a high accuracy, but since the dataset is imbalanced containing far more zeros and ones, the model needs to be assessed based on more metrics. The precision and the recall metrics, giving indications of how well the model can find the back order occurrences. From Iteration 1 in Table 8 it can be seen that a precision of 0.643, which mean that around 65% of the back order predicted actually are back orders. The recall being 0.310 indicate that around 30% of all back orders are predicted by the model. The F1-score give an indication of the harmonic mean between the precision and recall and no actual information can be extracted from this score.
### Table 8 - Complied list of deep learning iterations during model parameter evaluation with information about setup, performance and runtime

<table>
<thead>
<tr>
<th>Iteration</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Part numbers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td>Evaluating</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td><strong>Model parameters</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epochs</td>
<td>10</td>
<td>10</td>
<td>100</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>Layers</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Nodes</td>
<td>10</td>
<td>50</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td><strong>Metrics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.909</td>
<td>0.909</td>
<td>0.938</td>
<td>0.957</td>
<td>0.984</td>
</tr>
<tr>
<td>F1-score</td>
<td>0.419</td>
<td>0.398</td>
<td>0.630</td>
<td>0.793</td>
<td>0.182</td>
</tr>
<tr>
<td>Precision</td>
<td>0.643</td>
<td>0.660</td>
<td>0.853</td>
<td>0.811</td>
<td>0.481</td>
</tr>
<tr>
<td>Recall</td>
<td>0.310</td>
<td>0.284</td>
<td>0.500</td>
<td>0.776</td>
<td>0.112</td>
</tr>
<tr>
<td>AUC&lt;sub&gt;PRC&lt;/sub&gt;</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Misc.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Runtime</td>
<td>55m</td>
<td>59m</td>
<td>8h 34m</td>
<td>1h 47m</td>
<td>10h 55m</td>
</tr>
<tr>
<td>Platform</td>
<td>PC</td>
<td>PC</td>
<td>PC</td>
<td>PC</td>
<td>PC</td>
</tr>
</tbody>
</table>

**Iteration 2-4 – Adjusting Model Parameters**

In iteration 2-4 of the model, parameters were increased to see what the effect of increasing the different model parameters had on the performance metrics. In Table 8 above results from these runs can be found, where the colors indicate if the performance improved or got worse related to the baseline iteration. As seen in Table 8, it's clear that when the number of epochs is increased, as in iteration 3 it leads to a better performance seen to all of the metrics used. The same effects are reached when the number of layers is increased, as per iteration 4. Increasing the number of nodes, as seen in iteration 2, doesn't improve the performance. Finally, in iteration 5, both the number of layers and epochs are increased. As seen, the performance is increasing, but the precision and recall-metrics are decreasing. Another factor is the runtime, which in most cases is increases with an increased performance. In other words, when the number of layers and epochs increase, the runtime also increases.

This initial training and evaluation of the deep learning model and its model parameters is used when the model parameters are defined for when the model is trained with larger dataset and more part numbers. Since the both the number of layers and epochs gave a better result seen to performance but not in precision/recall. Both the smaller network from iteration 1 and the larger network from iteration 5 was used in the later training of the model. However not as many epochs as in iteration 5, since the runtime would increase too much.
7.3.3 Training on More Part Numbers
Initially when all 20 000 part numbers were to be trained on, a baseline run of the model with a smaller network was carried out on a laptop. This smaller network baseline gave an understanding of performance and runtimes for a basic setup of the model. The results from this run can be seen in Table 9 as Iteration 6.

Table 9 - Compiled list of deep learning iterations with more part numbers used, with information about setup, performance and runtime

<table>
<thead>
<tr>
<th>Iteration</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part numbers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td>20 000</td>
<td>20 000</td>
</tr>
<tr>
<td>Evaluating</td>
<td>20 000</td>
<td>1000</td>
</tr>
<tr>
<td>Model parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epochs</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Layers</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Nodes</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Metrics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.928</td>
<td>0.920</td>
</tr>
<tr>
<td>F1-score</td>
<td>0.722</td>
<td>0.708</td>
</tr>
<tr>
<td>Precision</td>
<td>0.729</td>
<td>0.674</td>
</tr>
<tr>
<td>Recall</td>
<td>0.716</td>
<td>0.745</td>
</tr>
<tr>
<td>AUC_{PRC}</td>
<td>0.59</td>
<td>0.60</td>
</tr>
<tr>
<td>Misc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Runtime</td>
<td>12h</td>
<td>42h</td>
</tr>
<tr>
<td>Platform</td>
<td>PC</td>
<td>PC</td>
</tr>
</tbody>
</table>

As seen, the baseline run in iteration 6 in Table 8 gave a better performance seen to all of the metrics used. Since more part numbers were used in the fitting and training of the model it’s reasonable that a better result was achieved. In iteration 7 a small increase in the number of layers and nodes was made, which resulted in slightly lower metrics in all cases, except the AUC-metric. An increase in the AUC-metric indicates that the model generally predicts and classify the states better, but the reason behind the lower precision and recall might be explained by the classification threshold, which impact on the outcome was more thoroughly explained in 7.3.1. This lower precision and recall could also be explained by the differences in evaluation. Iteration 7 was evaluated on 1000 randomly selected part numbers, the characteristics of these part numbers can determine the measured performance according to the metrics, to either produce a better or worse result compared to evaluating on all part numbers.
7.3.4 Model Results and the Volvo Aftermarket

The results of the model presented above, are metrics giving an indication of the performance of the model for different iterations. What the results of these model metrics mean for the supply chain is more difficult to define. Given the results of iteration 6 and a classification threshold of 0.5, the number of back order occasions that could be detected one week ahead would be 76 000. This based on the data used as training and evaluation. At the same time the number of falsely predicted back order occasions would be 28 000. The distribution of how the model has been predicted and classified the all of the data used in the evaluation can be seen in Figure 19 below.

![Figure 19 - Results from iteration 6 of model presented in error matrix](image)

The precision-recall curve gives an indication of how well the model can predict classes and in Figure 20 below, the precision-recall curve for iteration 6 in Table 8 is visualized. As the curve indicates, an increased threshold gives an increased recall, the precision slightly increases and then quickly declines after precision and recall both roughly reach 0.75. The Area Under the Curve (AUC) for iteration 6 is 0.59, suggesting that the model is indeed skillfully predicting the occurrence of back orders. What the optimal threshold is varies from different cases and how the output from the model is used, and is further elaborated on in the Discussion – Benefits and Difficulties.
As for precision and recall, precision measures how many selected items are relevant. A precision score of 0.729 means that, out of all predictions classified as a back order, approximately 73% of those were actual back orders. That means that around 27% of the true back orders predicted by the model would be false. Recall measures how many of the actual back orders that were classified correctly. The measured value of 0.716 means that roughly 72% of all back orders were predicted by the model. At the same time 28% of the actual back orders would be missed by the model. A visualization of the model’s ability to predict back orders can be seen in Figure 21 below.

Important to note is that the performance metrics are greatly affected by the fact that the model was trained on around 25% of the total part numbers. As the data set is highly imbalanced in regards back order occurrence, using all part numbers available would imbalance the dataset further, likely leading to a worse performance. Further discussion on this as well as how the supply chain can benefit from this information and what measures that should be taken with this information is found in Discussion – Benefits and Difficulties.
8 Discussion – Benefits and Difficulties

This chapter contains two separate sections. The first aims to discuss the benefits of predicting back orders with a potential use case that can be implemented at Volvo. The second sub-chapter discusses difficulties and limitations of the study.

8.1 The Benefits of Predicting Back Orders and the Impact on the Supply Chain

The performance and the results presented in Model Performance and Results is only a number of predictions and an indication of the performance of the model. Given the performance metrics, the model is assessed to be skillful in predicting back order occurrence since it’s able to successfully predict 73% of all back order occurrences one week ahead, and 72% of the back order predicted were actually back orders. This performance is not perfect, but proves that the model satisfies its aim and that it doesn’t provide random prediction output. The performance metrics in this study gives an indication that true information about 76,000 back orders during a year will be available to extract from the model. To compare with the results by de Santis et al. (2017) and their back order predictions, the precision of the model in this study (73%) is higher than the precision of their model (60%). The recall of this model (72%) also surpasses the recall of their model (20%). It’s difficult to with the current operations in the Volvo aftermarket since there are no process in place of this full extent.

As for the benefits of predicting the occurrence of back orders, one way to implement the prediction is to have a system that on a weekly basis predicts back orders for the following week to flag parts that have a high risk for back order occurrence. In other words, the model and its output will be used as support for the people working in the supply chain. A visualization of this can be seen in Figure 22 below. For example, this will allow the Back Order Recovery-team to act according to the logic of the Back Order Recovery-wheel with the benefit of taking more preemptive measures, a few examples of which being:

- Inbound flow
- Reserved inventory
- Other Brands

For example, if a part is identified to have a high probability of back order and none of the first solutions in the wheel are applicable. The part is not in the inbound flow, there is no reserved inventory or part from other brand, and so on. The member of the Back Order Recovery-team can preemptively contact the supplier to see if there are parts available for ordering, or search for a technical alternative.

If a part is found in reserved stock, or at a supplier and the likelihood of back order is deemed to be high, the part can be ordered or added to an existing order. If the lead time is short enough

![Figure 22 - Visualization of the model and its output & output and the connection with the supply chain](image)
compared to the lead time required by the customer, predicting the back order can avoid costly express shipping. If preemptive work has been carried out and a customer orders a part that is out of stock, depending on what steps were applicable, the customer can receive the ordered part earlier. In a best-case scenario, a part with a high likelihood of back order has already been stocked before the next order from a customer.

Another possibility is that the back order occurs during shipping of the part, which means that the customer will get the part earlier compared to if there was no prediction, this will also allow the Back Order Recovery-team to give the customer information regarding the part such as an Estimated Time of Arrival. However, even with accurate predictions the risk still exists that the part cannot be found anywhere in the supply chain. While this prevents any preemptive ordering, the customer can still benefit from quick information. Additionally, this information can in some cases be unnecessary as the occurrence of back order for some articles is seen as obvious. For example, if there is no supplier for an article, and no stock at all available there will be back orders, in these cases the information doesn’t help much. Nonetheless, the information might lead to an overall shortened solve time for back orders, which is in line with Volvos strategic goal of decreasing the back order ageing. A visualization of different possible actions that can be taken given a warning flag from the model can be seen in Figure 1Figure 23.

Threshold in Practice

The practical usage of the prediction actualizes the classification threshold and its impact on the distribution of true and false predictions. A high threshold will mean that the predicted back orders will have a very high likelihood of being an actual back order and avoid many false positives. The model will however be more likely to miss a larger amount of back orders. A low threshold will give the opposite effect. A large number of back orders will be caught, but so will a large number of false positives.

The threshold in the context of this study should be decided from two main factors. Firstly, the cost reduction of predicting the occurrence of a back order (true positive) and the cost increase of an incorrectly predicted back order (false positive). Secondly, the time available to the Back Order Recovery-team. The Back Order Recovery-team does a lot of work putting out the fires from parts currently on back order, and thus might not have time to investigate the predicted occurrence of future back orders. If the Back Order Recovery-team has a low amount of time available, the threshold should be set high. This would mean that they can use the model to preemptively solve some back orders, with a minimal risk of wasting time with solving false positives. When these two factors have been quantified and assessed, the precision-recall curve in Figure 20 in chapter 7.3.4 can be used to find the appropriate balance point.
In their study aiming to predict back orders, de Santis et al. (2017) discuss a scenario where all correctly predicted back orders could be avoided with a statement about the potential increase in service level. However, since the actual effects are to unclear, the potential in increased service level in actual numbers has been left out of this study, even though the potential exists. A potential increase of service means lower stock shortages costs. As mentioned, this model introduces these potentials even though the exact effects are difficult to assess. There will also be new costs such as maintaining the model and how the model will affect the express shipping costs is unclear. The focus of this study is not to explore the effects in detail, but it is clear that the information can prove to be useful for Volvo.

8.2 Process and Difficulties

As the study is concluded, three main difficulties during its course were identified. First off, the runtimes of training the model took over 40 hours for one iteration. Due to this, several of the iterations were carried out using a smaller subset of the part numbers, chosen at random. This allowed for testing the effects of adjusting the size and shape of the model without having an excessive runtime. A drawback however is that training on a larger data set is generally expected to produce a better result, meaning a model that gives predictions with higher accuracy.

**Runtimes**

The issue with the runtimes was largely caused by the characteristics of the problem of predicting the occurrence of back orders for a large number of part numbers using recurrent neural networks. Each part number is seen as a separate time series and needs to be trained in the model separately. This was done by iterating through a list of all part numbers, and for each iteration selecting a subset of the data set containing only data corresponding to a single part number. After the sub-set of data is created, the data is preprocessed and then used to train the model. This step with creating the sub-set slows down the model training by a fair amount but was deemed necessary as to separate each part number in training. The separating of training for a single part number is necessary as the model would interpret the data as a single time series and would use the previous states of one part number to train another part number, giving a faulty connection between part numbers. The data used was also limited to 52 weeks. The benefits of using deep learning can be found as the amount of data increases. If the model could be developed to run faster, with data from a larger time frame, the results should see significant improvements.

**Data Set Imbalance**

Another difficulty in this study is the imbalance of the data set. Out of the roughly 800 000 rows used, only 13.5% were of class-1, which compared to the ideal situation of a 50/50 distribution is considered a large imbalance. This means that the weeks where a part has a back order is outnumbered by close to 1:8. The problem with an imbalanced data set is that it greatly increases the difficulty of predicting the underrepresented class and changes the way you consider the accuracy metric. If class-1 has an occurrence of 10%, the model can predict class-0 for all cases and achieve an accuracy of 90%, which can seem like a good result but is highly misleading. Other metrics, such as the F-score are thus required to measure the performance of the model.

Furthermore, several iterations were done with a random sample of 1000 part numbers. This creates a risk that the sample does not represent the complete data set well and can thus further the risk of misleading results. Important to note is that the roughly 20 000 part numbers used were the part numbers that in the last 52 weeks had at least one back order. Using all the
80 000 part numbers would thus further imbalance the data set and would likely produce a classification with worse performance. Furthermore, not all features could be found on a weekly basis. Forecast and demand data was stored on a four-week basis. To combat this, the data was assumed to be evenly distributed over the four weeks and simply divided to match this. For example, if the demand for a certain part for the period of week 1-4 was 100, the demand for each of the weeks was assumed to be 25. As the variation in demand was seen as an important factor to explain back order occurrence, this is likely to have had a negative impact on the model performance.

**The Deep Learning Black Box**

Lastly the black box nature of deep learning is considered an issue for the sought business value from this study. The benefits of using a deep learning model is that the model can find connections between the features, that are incomprehensible to a human, thus allowing it to carry out estimations that are otherwise not possible. The problem is that the logic created by the deep learning model to predict the occurrence of back orders is highly complex and largely incomprehensible to a human.

While the model shows promising results that warrant further study, creating a change in the operations based on a deep learning model has problems. To realize an operational change, the change has to be thoroughly motivated which can be difficult when the underlying calculations hardly are understandable and thus hard to completely trust, and especially difficult to motivate to someone in the organization with a background not related to computer science or analytics. Even if estimations show that implementation of a deep learning model can produce business value, you risk losing the connection to the business area, logistics in this case. Even with thorough testing, the model can still have undiscovered flaws. If for example, the model causes an error which prevents customers from getting a part in time, it is difficult to explain why this happened if the calculation is so far removed from logistics that very few people in the organization understand exactly how it works.
9 Conclusions & Recommendations

This chapter presents a conclusion to the study aim by answering the research questions. The chapter also provides recommendations as to what the authors suggest should be the next steps for Volvo to further continue the study and reach an implementation. Lastly a section is dedicated to generalize the study findings.

9.1 Conclusions

The aim of this study is to develop a deep learning model to predict back order occurrence in the Volvo Group aftermarket.

The study successfully fulfilled the aim. The factors deemed to explain back order occurrences were used as input to a deep learning model. The deep learning model trained with these features successfully predicts back order occurrences in the Volvo Group aftermarket, with a precision of 73% and a recall of 72%. This is motivated by the answers to the two research questions, from which a comprehensive conclusion is reached. Detailed below are the answers to the research questions.

Research question 1. What underlying factors best explain the occurrence of back orders?

In order to find the factors that best explain the occurrence of back orders, a literature study and an empirical investigation were carried out. The literature review lead to findings in both logistics and machine learning, gathering a total of 18 factors that explain the cause of back orders. These 18 factors were grouped into four factor groups, demand, inventory, supply and article characteristics. The empirical study uncovered four initiatives that aimed to find the cause of back orders, and the factors from these were grouped to the earlier identified groups. A total of 28 factors were found from the empirical study.

Lastly, the factors accrued from the literature and empirical studies where compiled. Overlapping factors were combined and also grouped according to their main area. Some factors were also excluded based on not being available at Volvo or a low connection to back orders. After this compiling and grouping a total of 20 factors remained and formed the answer to the first research question. Table 10 below shows the 20 factors, with their corresponding factor group and an indication of it was found in the literature or at Volvo. This list of factors well covers the actual back order causes. The most important factors, uncertainties in demand and supply, are well understood and covered in literature as well as utilized by Volvo in their current operations. If Volvo was to measure the factors that were excluded based on availability, the results could see improvements.
**Table 10 - Compiled list of factors found best explain the occurrence of back orders**

<table>
<thead>
<tr>
<th>FACTOR GROUP</th>
<th>FACTOR</th>
<th>LITERATURE REVIEW</th>
<th>EMPIRICAL STUDY</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEMAND</td>
<td>Demand</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Forecasted Demand</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Demand Variability</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Seasonality</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Passive</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>INVENTORY</td>
<td>Cycle Stock</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Safety Stock</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Current Stock Level</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Target Availability</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Late Binning</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Turnover Rate</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>SUPPLY</td>
<td>Lead Time</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Delivery Precision</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Delivery Frequency</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>No Purchasing Contract</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Quality Related (QJ)</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>ARTICLE</td>
<td>Price</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>CHARACTERISTICS</td>
<td>Part Criticality</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Part Life Cycle</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Product Group</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

**Research question 2. How could a deep learning model be designed to predict back orders?**

The design of the model was divided into three steps. Firstly, the factors generated from RQ1 were used to create the input features. Some factors could be directly translated into features, where the exact factor could be found in the Volvo database and added to the data set. Factors not explicitly represented by their own feature, are in some cases deemed covered by another feature. Some factors were excluded in this step, either because appropriate data was not found or the factor was not viable as a quantitative feature. The exact effects the excluded features are difficult to assess, but it’s possible that the performance of the model would increase with more features.

The created model is a deep recurrent neural network, comprised of layers containing long short-term memory (LSTM) nodes. The size and shape, meaning the number of layers and nodes, was adjusted through experimentation. The final results regarding the model was a model that were able to successfully predict around 72% of the back order occurrences during a year (recall), which means around 76 000 yearly occurrences. And 73% of what the model deemed to be back orders were actual back orders (precision). It’s difficult to say if these results are good or bad, but compared to a study by de Santis et al., (2017) also predicting back orders, this model performs better, as seen in Table 11.

**Table 11 - Comparison of precision and recall metrics of model to a study by de Santis et al., (2017)**

<table>
<thead>
<tr>
<th></th>
<th>THIS STUDY</th>
<th>COMPARATIVE STUDY</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRECISION</td>
<td>73%</td>
<td>60%</td>
</tr>
<tr>
<td>RECALL</td>
<td>72%</td>
<td>20%</td>
</tr>
</tbody>
</table>
It was also found that more data in the form of part numbers and number of times each data point was fed to the model, increased the performance of the classification. At the same time run-times of training increased. Lastly, the importance of the threshold as this directly impacts the distribution of the correctly and falsely predicted back order occurrences.

As for findings beyond the research questions, conclusions also included potential usage and benefits of the predictions. A potential use case for Volvo identified in this study is to implement the predictions to be used by the Back Order Recovery-team. The team could use the information about back order occurrences for the following week to preemptively take measures. The team could follow the already existing logic of the Back Order Recovery wheel. The predictions would then work as a support to take decisions such as ordering a part to avoid a future back order.

9.2 Recommendations to Volvo

Results from the study showcase that this is an interesting and very promising way of taking advantage of the deep learning technology in the Volvo aftermarket Supply Chain. However, before the back order predicting deep learning model can be implemented, some things should be sorted out and further investigated. These further investigations that is recommended for Volvo to hand can be divided into three main areas.

Firstly, in this study the timeframe of the data was a year and the dataset used wasn’t fully representative of all the part numbers. So, further testing with a longer timeframe so that more long going trends and seasonal patterns can be identified. Using a representative dataset and make sure that the model performs well in this case is important as well. One approach to getting a representative dataset would be to use all of the 80 000 part numbers, which would increase the runtime for training the model. Especially if the timeframe of the data would be increased as well.

Secondly, the threshold of when a prediction is classified as a one should be evaluated and a favorable threshold should be decided. With favorable meaning finding the balance point between the potential cost reduction for predicting a back order and the cost increase for a false positive prediction. The time available for the Back Order Recovery-team to handle information also becomes a restraint that needs to be considered when the threshold is decided.

Third and last, the effects of this information should be decided and defined. This study had the main focus on the possibility of predicting when back orders and stock-outs were to take place. The effects of this prediction, together with the processes taking place after the prediction should be clarified and defined. With processes means which unit that should take advantage of the information, for example the Back Order Recovery-team, and how decisions are to be made regarding the output. For example, make orders of the part to suppress the costs of the back orders, or take a chance and wait for the back order to occur.

Apart from these three points of further investigative work, the deep learning technology is seen as beneficial in the aftermarket Supply Chain and that it could be used to provide valuable predictive information regarding both related to back orders, as well as other fields.
9.3 Generalizing the Study

As this study was carried out at Volvo and their European aftermarket, the results have been widely influenced by the operations of Volvo and analyzed mainly in a Volvo perspective. In a wider perspective, the methodology used in this study can be considered generalizable in large. In the first research question, both the literature study and empirical study could be applied to understand any state in the supply chain. The finalized factors explaining back orders found in this study is of a general interest as well.

If a similar endeavor with predicting back orders was carried out at another company with a similar supply chain, the results would largely depend on the available data. With data of high quality, a similar study on a different supply chain should see good results. However, in order to achieve this, it’s important to customize the features based on the factors that can be seen as correlating with the occurrence of back orders. In all different cases where machine learning where to be applied in a supply chain to predict states, the appropriate features need to be found and decided on. This study provides a useful framework in the methodology of going from a supply chain to input features to a machine learning model.

Further, we see the use of deep learning as a great tool to predict future states in the supply chain. Either as classification of states, or as prediction of volume. The capability of the computer model to take in a numerous number of input parameters from a longer time-frame, is seen as fitting as the supply chain contain a lot of datapoints with information. With this as motivation we see this as an interesting research area, whereas deep learning could be applied in several different ways in the supply chain, providing information about states in the future.

This study could be seen as an introduction to the intersection between supply chains and machine learning. The study provides information on how machine learning can be applied in a supply chain, as well as how to evaluate the results and the benefits that come with a successful implementation of a machine learning model. The study also covers and explains many of the machine learning specific decisions and aspects that need to be considered if machine learning was to be utilized in a supply chain. Making the study a good introduction to the machine learning technology and for someone who is curious about its potential and applicability on supply chains.
References


Graves, A. (2012). *Supervised Sequence Labelling with Recurrent Neural Networks*.


### Appendix A – Literature Study

<table>
<thead>
<tr>
<th>DATABASE</th>
<th>SEARCH TERM</th>
<th>NUMBER OF HITS</th>
<th>ABSTRACTS READ</th>
<th>DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNISEARCH</td>
<td>Machine learning AND Supply Chain</td>
<td>79</td>
<td>79</td>
<td>2019-01-29</td>
</tr>
<tr>
<td></td>
<td>Machine Learning</td>
<td>1121579</td>
<td></td>
<td>2019-01-29</td>
</tr>
<tr>
<td></td>
<td>Aftermarket AND Supply Chain</td>
<td>64</td>
<td>65</td>
<td>2019-02-12</td>
</tr>
<tr>
<td></td>
<td>Neural Networks and Machine Learning (Book)</td>
<td>5899</td>
<td></td>
<td>2019-02-14</td>
</tr>
<tr>
<td></td>
<td>Deep Learning</td>
<td>214199</td>
<td></td>
<td>2019-02-15</td>
</tr>
<tr>
<td></td>
<td>Deep Learning AND Python</td>
<td>991</td>
<td></td>
<td>2019-02-15</td>
</tr>
<tr>
<td></td>
<td>Convolutional Neural Networks</td>
<td>88093</td>
<td></td>
<td>2019-02-16</td>
</tr>
<tr>
<td></td>
<td>Back order AND Machine Learning</td>
<td>21</td>
<td>5</td>
<td>2019-03-05</td>
</tr>
<tr>
<td></td>
<td>Forecasting spare parts AND Logistics</td>
<td>265</td>
<td></td>
<td>2019-01-25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SEARCH TERM</th>
<th>HIT NR</th>
<th>AUTHOR</th>
<th>DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>MACHINE LEARNING</td>
<td>4</td>
<td>Bell, Jason. (2014)</td>
<td>2019-01-29</td>
</tr>
<tr>
<td>DEEP LEARNING AND PYTHON</td>
<td>1</td>
<td>Ketkar, N.</td>
<td>2019-02-15</td>
</tr>
</tbody>
</table>
Appendix B – Model Implementation

The model has been implemented using deep learning libraries TensorFlow and Keras. Other packages used are Pandas, Sklearn, Matplotlib and numpy.

Hyperparameters that were not varied between iterations:

- Batch size = 1
- Final layer is a dense layer with activation function sigmoid
- Activation function – tanh
- Optimizer – adam

Program Functions

Preprocess

Normalizes the data in the dataset to a value between 0 and 1

create_subset_pn

Creates a subset of the data set containing data for a single part number

Timestep

Adds a number of timesteps to the dataset, the timesteps dictate how far back in time the model can see.

train_and_val

Splits the data set into a training part and a validation part. Drops indexing columns, converts dataframe to numpy array.

create_predict_data

Same as train_and_val but does not create the training part. Used for evaluation.

create_model

Creates the model.

compile_neural_network

Compiles the model.

train_neural_network

Trains the neural network based on training data.

evaluate_neural_network

Evaluates the neural network based on the evaluation data.
import pandas as pd
import numpy as np
from sklearn.preprocessing import MinMaxScaler
from keras.models import Sequential
from keras.layers import Dense, LSTM
import time, datetimes
import matplotlib.pyplot as plt
from sklearn.metrics import accuracy_score, f1_score, precision_score, recall_score

# Load Dataframe from csv-file
df = pd.read_csv(r'dfnewtimestep.csv')
df.sort_values(by=['Part Number', 'Period'], inplace=True)
df.drop(columns=['Supplier No', 'Supplier No(t-1)', 'Supplier No(t-2)', 'Supplier No(t-3)', 'Supplier No(t-4)'], inplace=True)

# Create list of all part numbers that will be used in training (model-testing purposes)
part_numbers = df['Part Number'].unique()

# Select 1000 random part numbers used for testing
num_test_samples = 1000
np.random.shuffle(part_numbers)
test_part_numbers = np.random.choice(part_numbers, num_test_samples, replace=False)

# Define some variables for the model (epochs, layers in the model, timesteps in input, features)
n_epochs = 30
n_layers = 10
n_nodes = 10
timesteps = 5
n_features = int((len(df.columns) - 1) / 5)

# Declaring global variables to see what happens within functions (can be defined from functions)
glob_train_x = 0
glob_train_y = 0
glob_eval_x = 0
glob_eval_y = 0
pred_y_list = np.empty(0)
pred_y_list_perc = np.empty(0)
pred_y_list1 = []
pred_y_list2 = []
eval_y_list = np.empty(0)
precision_list = np.empty(0)

# Create training data and evaluation data from dataframe and store in numpy array
# Normalize data

def preprocess(df):
    features_to_normalize = list(df.columns[2:])
    if len(features_to_normalize) > 0:
min_max = MinMaxScaler(copy=False)
df[features_to_normalize] =
min_max.fit_transform(df[features_to_normalize])
return df, min_max

# create a subset dataframe for the part number as per input
def create_subset_pn(df, pn):
df_pn = df[df['Part Number']==pn]
return df_pn

# adding timestep to create the "weird matrix". adding and shifting columns
def timestep(df_pn, timesteps, n_features):
    features_to_shift = list(df_pn.columns[2:2])
    for i in range(1, timesteps):
        for feature in features_to_shift:
            col_name = feature+"(t-"+str(i)+")"
            df_pn[col_name] = df_pn[feature].shift(i)
    df_pn = df_pn[:-timesteps]
    df_pn = df_pn.iloc[4:]
    df_pn = df_pn.iloc[:-4]
return df_pn

# Splitting dataset into train and eval. Dropping part number and period column.
# Creating 3D-matrix to feed to the LSTM-model
# global reshaped is used for testing purposes
def train_and_val(df, timesteps, n_features):
    global glob_train_x
    global glob_train_y
    global glob_eval_x
    global glob_eval_y
    df['Period'] = df['Period'].astype(int)
df_train = df[df['Period'] <=
df['Period'].unique()[int(round((df['Period'].nunique()*0.7), ndigits=0))]]
df_eval = df[df['Period'] >
df['Period'].unique()[int(round((df['Period'].nunique()*0.7), ndigits=0))]]
df_train = df_train.drop(columns=['Part Number', 'Period'])
df_eval = df_eval.drop(columns=['Part Number', 'Period'])
train_x = np.array(df_train.iloc[:-1, 0:(timesteps*n_features)])
train_y = df_train['BO Balance'].shift(-1)
train_y = np.array(train_y.iloc[:-1])

eval_x = np.array(df_eval.iloc[:-1, 0:(timesteps*n_features)])
eval_y = df_eval['BO Balance'].shift(-1)
eval_y = np.array(eval_y.iloc[:-1])
glob_train_x = train_x
glob_train_y = train_y
glob_eval_x = eval_x
glob_eval_y = eval_y
train_x = np.reshape(train_x, (train_x.shape[0], timesteps, n_features))
eval_x = np.reshape(eval_x, (eval_x.shape[0], timesteps, n_features))
```python
return train_x, train_y, eval_x, eval_y

# same as train_and_eval, but without the splitting into train and eval sets
def create_predict_data(df, timesteps, n_features):
    df['Period'] = df['Period'].astype(int)

    df_eval = df[df['Period'] > df['Period'].unique()[int(round((df['Period'].nunique() * 0.7), ndigits=0))]]

    df_eval = df_eval.drop(columns=['Part Number', 'Period'])

    eval_x = np.array(df_eval.iloc[:-1, :].iloc[:, ::(timesteps * n_features)])

    eval_y = df_eval['BO Balance'].shift(-1)
    eval_y = np.array(eval_y.iloc[:-1])

    eval_x = np.reshape(eval_x, (eval_x.shape[0], timesteps, n_features))

    return eval_x, eval_y

# Creates the model, size is based on input
def create_model(n_layers, n_nodes, batch_size, activation_f, timesteps, n_features):
    model = Sequential()

    # Add a starting layer
    if n_layers > 1:
        model.add(LSTM(n_nodes,
                        activation=activation_f,
                        batch_input_shape=(batch_size, timesteps, n_features),
                        name='LSTM_layer_1',
                        stateful=True,
                        return_sequences=True))

        # Add additional layers
        for i in range(1, n_layers):
            if i < n_layers-1:
                model.add(LSTM(n_nodes,
                                activation=activation_f,
                                return_sequences=True))
            else:
                model.add(LSTM(n_nodes,
                                activation=activation_f))

        else:
            model.add(LSTM(n_nodes,
                            activation=activation_f,
                            batch_input_shape=(batch_size, timesteps, n_features),
                            name='LSTM_layer_1',
                            stateful=True))

    # Add an output layer (using sigmoid activation function for output)
    model.add(Dense(1, activation='sigmoid', name='output_layer'))

    return model
```

9
# Compile the neural network

def compile_neural_network(model):
    return model.compile(loss='binary_crossentropy',
                         optimizer='adam',
                         metrics=['accuracy'])

def train_neural_network(timesteps, df, part_numbers, model, n_epochs, batch_size=1, plot=True):
    # Start timing to measure training time
    start = time.time()
    # Create empty lists to store loss and error in
    train_history = []
    eval_history = [[]]  # Needed for stateful LSTM
    # Loop with that resets the states after each epoch.
    for pn in part_numbers:
        x = 0
        print("training pn: " + str(pn) + " with # of epochs: " +
              str(n_epochs)+"pn " +str(x) + "of"+str(len(part_numbers)))
        df_pn = create_subset_pn(df, pn)
        train_x, train_y, eval_x, eval_y = train_and_val(df_pn, timesteps, n_features)

        # Fit the model. Turn shuffle of to keep the time series sequence intact.
        result = model.fit(train_x, train_y, epochs=n_epochs,
                            batch_size=batch_size, verbose=0, shuffle=False)

        # Append the results to the history list
        train_history.append(result.history)
        epoch_eval = model.evaluate(eval_x, eval_y, batch_size=batch_size,
                                     verbose=0)
        eval_history[0].append(epoch_eval[0])
        eval_history[1].append(epoch_eval[1])

        # Reset the states (after each part number)
        model.reset_states()

    # End timing
    end = time.time()

    # Print the elapsed time when the training is done
    print('nTraining done. Time elapsed:',
          str(datetime.timedelta(seconds=round(end-start, 0))))

    temp_list = [[]]  # Needed for range
    for key in train_history[0].keys():
        for i in range(len(train_history[0])):
            temp_list[0].append(train_history[0][key][i])

    return temp_list
temp = train_history[i][key][0]
temp_list[l].append(temp)
l += 1

if plot == True:
    plt.subplot(2, 1, 1)
    plt.plot(temp_list[0], color='orange', label='training data')
    plt.plot(eval_history[0], '-r', label='evaluation data')
    plt.ylabel('Loss')
    plt.title('Loss')
    plt.legend()

    plt.subplot(2, 1, 2)
    plt.plot(temp_list[1], color='orange', label='training data')
    plt.plot(eval_history[1], '-r', label='evaluation data')
    plt.title('Accuracy')
    plt.ylabel('Accuracy')
    plt.xlabel('Part Number')
    plt.legend()

return
def evaluate_neural_network(model, df, scaler, timesteps, n_features,
batch_size=1, plot=True):
    global glob_pred_y
    global glob_actual_y
    global pred_y_list
    global pred_y_list1
    global pred_y_list2
    global eval_y_list
    global precision_list
    global pred_y_list_perc
    global df_metric

    #creates train/test-set for eval-pns again since it's not a global
    variable

    for pn in test_part_numbers:
        print("evaluating partnumber: "+ str(pn))
        df_pn = create_subset_pn(df,pn)

        eval_x, eval_y = create_predict_data(df_pn,timesteps, n_features)

        #getting non-divided-train-data

        # Make predictions on train and eval data
        pred_y = model.predict(eval_x, batch_size=batch_size)
        pred_y_list_perc = np.append(pred_y_list_perc,pred_y)

        pred_y= np.around(pred_y)
        train_acc = accuracy_score(eval_y, pred_y)
        train_f1 = f1_score(eval_y, pred_y)
        train_precision = precision_score(eval_y, pred_y)
        train_recall = recall_score(eval_y, pred_y)
        a = [pn,train_acc,train_precision,train_recall,train_f1]

        pred_y_list= np.append(pred_y_list,pred_y)
eval_y_list= np.append(eval_y_list,eval_y)
precision_list=np.append(precision_list,a)

precision_list=np.reshape(precision_list,(len(test_part_numbers),5))
pred_y_list1=pred_y_list
pred_y_list2=pred_y_list
pred_y_list = np.around(pred_y_list)

train_acc = accuracy_score(eval_y_list,pred_y_list)
train_f1 = f1_score(eval_y_list,pred_y_list)
train_precision = precision_score(eval_y_list,pred_y_list)
train_recall = recall_score(eval_y_list,pred_y_list)

print('Test Accuracy: %.3f' % train_acc)
print('Test F1: %.3f' % train_f1)
print('Test precision: %.3f' % train_precision)

# print('Training recall: %.3f' % train_recall)

output = pd.DataFrame({'Predicted y':pred_y_list_perc,'True y':eval_y_list})
output.to_csv('Bo_Prediction_output.csv',index=False)

# return train_acc, train_f1, train_precision, train_recall

df, min_max = preprocess(df)
model = create_model(n_layers,n_nodes,1,'tanh',timesteps,n_features)
compile_neural_network(model)
train_neural_network(timesteps,df,part_numbers,model,n_epochs,1,plot=True)
train_acc, train_f1, train_precision, train_recall = 
evaluate_neural_network(model, df, min_max, timesteps, n_features)

#Saves the Model after training
model.save('BO_precition_model.h5')