CHALMERS
UNIVERSITY OF TECHNOLOGY

Predicting Deviation in Supplier Lead Time and Truck Arrival Time Using Machine Learning

A Data Mining Project at Volvo Group

Master's thesis in Supply Chain Management
Supply and Operations Management

MENG HUANG

MASOOD BAGHERI

Department of Technology, Management and Economy
Division of Supply and Operations Management
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2019
Report No.E 2019:033

Master's thesis 2019:033

Predicting Deviation in Supplier Lead Time and Truck Arrival Time Using Machine Learning

A Data Mining Project at Volvo Group

MENG HUANG
MASOOD BAGHERI

Supervisor, Chalmers:	Joakim Andersson
Supervisors, Company:	Anton Ottosson & Martin Granic
Examiner, Chalmers:	Patrik Jonsson

Department of Technology Management and Economics
Division of Supply and Operations Management
Chalmers University of Technology
Gothenburg, Sweden 2019

Predicting Deviation in Supplier Lead Time and Truck Arrival Time Using Machine Learning
A Data Mining Project at Volvo Group
MENG HUANG
MASOOD BAGHERI

Master's Thesis 2019:033

Department of Technology Management and Economics
Division of Supply and Operations Management
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Printed by Chalmers Reproservice
Gothenburg, Sweden 2019

Predicting Deviation in Supplier Lead Time and Truck Arrival Time Using Machine Learning
A Data Mining Project at Volvo Group
MENG HUANG
MASOOD BAGHERI
Department of Technology Management and Economics
Division of Supply and Operations Managemen
Chalmers University of Technology

Abstract

The deviation in delivery performance from a company's suppliers directly affects the company's performance, causing availability loss for the customer orders and large costs for the rush transportation. If the deviation can be predicted in advance and used as deviation alerts, actions can be taken in advance either to prevent the deviation or decrease the impact of the deviation.

To predict the deviation in the supplier delivery performance from a buying company's point of view, this thesis work specifically focuses on the first two phases of a supply chain, namely supplier lead time from material suppliers and truck arrival time from logistics service providers (LSP). In order to examine the possible implementation of machine learning, a data mining project has been conducted at Volvo Group Service Market Logistics. The factors associated with deviation of supplier lead time and truck arrival time are identified, while the corresponding features are prepared under the constraint of the case company's data availability. For predicting deviation in the two phases, two machine learning models are constructed accordingly based on the characteristics of output and input features. The opportunities and obstacles along the data mining process in the case company are identified.

The results show currently in the case company, both generated machine learning models do not have enough predictive power in lead time deviation. This could be caused by the absence of some key features that have strong associations with deviation. However, the performance of the prediction model for truck arrival time is regarded to be improved to a deployable level when the desired features are constructed into the model by the case company. Future recommendations regarding constructing the desired features and improving the model performance are proposed. In comparison, predicting deviation in material suppliers' lead time could be practical when the buying company get more information sharing from material suppliers.

Keywords: Lead time deviation, Estimated time of arrival (ETA), Prediction, Delivery precision, Machine learning, Supplier evaluation, Spare parts, Automotive.

Acknowledgements

This master thesis was conducted during the spring of 2019 as part of the Supply Chain Management program at Chalmers University of Technology. This thesis collaborates with Volvo Group Service Market Logistics, where this data mining project was conducted.

First, we would like to thank our supervisors at Volvo, Anton Ottosson and Martin Granic. Their great support and experience guided us through the entire project. Their generous encouragement and compliment also made us confident towards the completion of this project. We would also like to thank all the involved people at Volvo for their time and insights contributed to this project.

Then, we would like to express great gratitude to Joakim Andersson. As our supervisor at Chalmers, his profound interests in our finding motivated our work and his feedback inspired us at every key point in this project. Further, we want to appreciate the insightful inputs from our examiner, Patrik Jonsson at Chalmers.

Finally, we thank our friends and family for their immense support. Meng would like to give a special thank to her girlfriend, Yannan Shen, who supported her dedicatedly from home. Her knowledge of statistics and academic writing also helped trigger a lot of thinking during this project.

Meng Huang and Masood Bagheri
Gothenburg, June 2019

Abbreviations

CART	Classification and Regression Trees
Catboost	Categorical Boosting
CDC	Central Distribution Center
CMP	Continental Material Planners
CRISP-DM	Cross Industry Standard Process for Data Mining
DDT	Door to Door
DIP	Demand Inventory Planners
EMEA	Europe, Middle East and Asia
ETA	Estimated Time of Arrival
FN	False Negative
FP	False Positive
FTL	Full Truck Load
KPI	Key Performance Indicator
LSP	Logistics Service Providers
LTL	Limited Truck Load
MMOG/LE	Materials Management Operational Guidelines / Logistics Evaluation
QPM	Quality Performance Measurement
RDC	Regional Distribution Center
SDC	Support Distribution Center
SEM	Supplier Evaluation Measurement
SLT	Supplier Lead Time
SM	Supplier Managers
SML	Service Market Logistic
SRM	Supplier Relationship Managers
TLT	Transportation Lead Time
TMC	Transport Material Coordinators
TN	True Negatives
TP	True Positives
PPM	Parts Per Million

Contents

List of Figures

List of Tables

1

Introduction

In this chapter, the theoretical background and company background of this thesis project is introduced, following by the aim of the project. The research questions are thereby formulated and the scope of the project is presented.

1.1 Theoretical Background

Spare part supply chain is a high-margin business bringing in high profits for the company. However, delivering spare parts is more complex than manufacturing the products, since a spare part supply chain has to cover the aftermarket service for all the products sold by the company. Customers also expect their things to be fixed quickly when they break down, while their demands are intermittent because the breakdown happens unexpectedly. These difficulties make only companies that provide the spare part efficiently can make revenues from aftermarkets (Cohen, Agrawal and Agrawal, 2006).

The supply chain management in a company should match the demand and supply (Jonsson, 2008). Forecasting the demand in order to mitigate the risks of uncertainty and availability loss of spare parts has received lots of research attention (Dekkeret al., 2013). The uncertainties also come from supply sides (Heydari et al., 2009), where deviation in lead time impacts the delivery precision and raises uncertainty on the supply. According to Ioannou and Dimitriou (2012), lead time has direct impacts on inventory and supply availability, and therefore the issue of managing lead time has also been consistently discussed in the literature since the late 1960s. To be specific, when a deviation occurs to the lead time, it results in the estimated time of arrival (ETA) being not accurate and further disturbing inventory planning. The inventory of spare parts is, therefore, going to fluctuate, causing stockouts when spare parts arrive late or inventory holding costs when they arrive early (Heydari et al., 2009). In particular, spare parts belong to maintenance inventories and the stockout costs of the spare parts could be significantly high (Kennedy, Patterson and Fredendall., 2002). Inspired by preventive and corrective maintenance (Mobley, 2002; Schmidt and Wang, 2018), if the deviation of lead time can be predicted beforehand, preventive actions can be adopted to minimize deviation, improving the accuracy of ETA and secure delivery precision. Corrective actions can also be scheduled to mitigate the impacts of the deviation. For instance, to diminish deviation, more attention can be put on monitoring the supply process where it is predicted to have deviated time of arrival and therefore the company can proactively take actions to avoid the deviation. To mitigate the impacts of deviated arrival time that could bring fluctuated stock level, inventory planning can be updated considering the deviation of ETA to ensure the availability of stock.

Overall, the successful prediction of the deviation on lead time can, firstly lead to a lower total cost, because the right information of arrival time contributes to having the right amount of spare parts in the inventory at the right time, saving both inventory holding costs and inventory shortage costs (Carbonneau, Laframboise and Vahidov, 2008). Secondly, it can improve customer satisfaction by securing their vehicle up-time with the availability of spare parts needed in the warehouse (Carbonneau et al., 2008). Therefore, costs saving and capability of fulfilling customer orders on time are the outputs of an accurate prediction of lead time deviation.

Since there are various companies cooperating in the supply chains, the performance from supplier companies is going to affect buying companies' performance. This is the case especially for manufacturing industries including automotive, who relies heavily on component suppliers (Krause, Handfield and Tyler, 2007). Therefore, it is beneficial to predict the delivery performance from the buying companies' perspective to secure their business operation.

Machine learning models are emerging to be used to predict suppliers' performance and predict the lead time or ETA in different transport modes, due to its ability to capture the pattern from complex relationship between input features and output performance (Witten et al., 2017). For example, predicting arrival time of truck in distribution are discussed (van der Spoel, Amrit and van Hillegersberg, 2017). Delay in passenger airplanes and freight trains (later than ETA) have also been predicted using machine learning from transport handlers' perspective (Belcastro et al., 2016, Takacs 2014, Barbour et al, 2018). However, for material supplier performance, existing literature only predicts supplier overall performance rather than specifically focusing on delivery precision (Jiang et al., 2013; Khaldi et al., 2017). For the transportation, the performance of prediction models varies with different input features. So far, we have not found literature that is based on input variables of organisation and human to predict truck arrival time with machine learning.

1.2 Company Background

Volvo Group (Volvo) Service Market Logistics (SML), as one of the departments in the case company where this project is performed, is responsible for the development and optimization of the spare part supply chain which strives for securing the availability of spare parts at the lowest possible costs.

To achieve this goal, the target of delivery precision performance from logistics service providers (LSP) in SML is 97%. It means 97% of transportation delivery shall not arrive late on each node. However, due to the fact that lead times are negotiated with their suppliers and set in the planning system for a longer period of time since the cooperation starts and there are various uncertainties in supply process, the deviation occurs frequently in lead time. For the spare parts of Volvo truck in Europe in 2018, around 37% delivery does not meet the ETA at their central distribution center (CDC) according to predefined transportation lead time (TLT). Among the deviation, 27% of them arrived earlier and 10% of them arrived later than ETA. Previous than that transportation delivery goal, the target for the material suppliers' delivery precision is 95%, which means 95% of the orders from material suppliers shall not be ready later than scheduled. However, for the previous performance in 2017 and 2018, merely 77% of them does not have deviation in supplier lead time (SLT) and was dispatched on time, with 9% of them dispatched earlier than scheduled, and the remaining 14% dispatched later than estimated.

This big share of deviation could directly bring fluctuation in inventories. Spare parts arriving earlier than estimated are bringing extra tied-up capital, inventory costs and disturbing the work schedule in warehouses, while late-arrived spare parts could either cause extra delivery costs in recovering the back-orders by expediting logistics using air transport, or become excess inventory and end up being scrapped because of missing out to supply the demand. As it is important for Volvo to fulfil customers' demand at a lower cost, there is a need for predicting lead time deviation for monitoring the delivery precision performance on their material suppliers and logistics service providers (LSP) in order to proactively checking ETA of spare parts and take actions.

In Volvo, the importance of big data is increasingly raising attention. More and more data are collected and analyzed. These new data resources combined with advanced analytic methods are creating new opportunities to reap the fruits of data mining to benefit business. Volvo has realized the power of machine learning models in prediction and has been initiating data mining projects to explore its possible usages and potential benefits. Therefore, this study targets on predicting deviation of lead time on its suppliers of material and transportation by implementing machine learning.

1.3 Aim

The aim of this thesis is to evaluate whether and how machine learning modelling can be implemented to predict lead time deviation from buying companies' suppliers of material and logistics, under the consideration of achieving benefits of a prediction model in the current stage of the case company Volvo SML.

To achieve the aim of implementing machine learning models to predict lead time deviation, the first research question is to investigate how the company can utilize the lead time prediction. This question sets the business goal and answers potential benefits of this data mining project.

RQ 1: What are the benefits of predicting lead time deviation for buying companies?

The second research question is to investigate the factors that are associated with deviation from the buying company's perspective. These factors are the basis for features construction for machine learning modelling.

RQ 2: What are the factors that could be associated with lead time deviation perceived by buying companies?

However, only the factors that can be represented with available data in the company's database can be analyzed and constructed into the prediction model. This research question reflects the limitation existing in the case company for the construction of the model and contributes to set the data mining goal of this project.

RQ 3: Which data are available to be used as features when building the prediction model of lead time deviation at Volvo SML?

The fourth question is to develop a prediction model by testing different machine learning strategies and algorithms. The modelling process is based on Volvo's situation considering the benefits that the company can practically achieve in the current stage. The results of modelling will be also examined and interpreted regarding their usability.

RQ 4: How should the prediction model be built using machine learning considering the practicality of use in the current stage at Volvo SML?

1.4 Scope

In order to fulfil the aim of this thesis project, a certain scope is needed. The scope of the thesis is focusing on the spare parts that belong to Volvo truck in European region. Further, for the scope of lead time, the chosen phase will be examined from the moment that Volvo places orders to its material suppliers and shipped by LSP until they arrive at the CDC in Ghent, Belgium. The reason for choosing this inbound flow is because it currently suffers from the largest deviation and this flow is at the beginning of the supply chain which has cascade effects on later processes. In this project, this lead time is named inbound flow lead time and it consists of two phases which are supplier lead time (SLT), and inbound transportation lead time (TLT). The SLT is the time taken by the material suppliers to get ready for ordered parts. The TLT is the time taken by LSP from consignors (material suppliers) to consignee (CDC Ghent). The deviation in TLT results in the deviation in arrival time. In line with the company's measurement system, the deviation for SLT is measured in the weekly basis, while the deviation for TLT is measured in daily basis. This means deviation of SLT beyond one week and deviation of TLT beyond one day is counted as 'Late' or 'Early'.

In our scope, the transportation mode is regarded as road transportation with trucks, since the delivery within Europe is mainly adopt trucks only with the exemption of the cross-docking shipments from Sweden to CDC Ghent which are transported via sea. This sea flow is not considered in the prediction model.

The information used in the project is limited within the case company. The data related to deviation in scope are not including the suppliers' solely owned information such as production information in material suppliers, and fleet management information in LSP. No external data is used.

2

Literature Review

In order to support the analysis and discussion by providing theoretical resources and domain knowledge for machine learning, a literature review is conducted in this chapter. It is divided into two parts with the first part reviewing spare parts supply chain, previous work and current state of predicting lead time deviation, while the second part including the last two sections is introducing machine learning.

2.1 Frame of Reference

This section introduces the frame of reference which helps to present the context of spare part logistics and the application of machine learning in the area of supplier evaluation and ETA prediction. They are corresponding to the subjects of this project.

2.1.1 Spare Part Logistics Context

The requirements for planning spare parts logistics are different from the logistics of other material from several aspects (Huiskonen, 2001). Firstly, the service requirement of logistics is high due to the remarkable costs and penalties for spare parts shortage. However, the demand for spare parts is sporadic and hard to predict which bring high risks of late delivery. Secondly, due to the decrease of the buffers of time and material in the supply chain and production systems, streamlining the spare parts logistics is under the pressure (Huiskonen, 2001).

Most papers are addressing these requirements by focusing on the inventory management of spare part locally rather than considering the whole supply chain (Zanjani & Nourelfath, 2014). However, inventory optimization often has strict assumptions and difficult to apply. There is a need to increase the collaboration between different actors to plan spare parts logistics to deal with the special requirements of spare part logistics (Huiskonen, 2001).

2.1.2 Supplier Evaluation

One aspect of collaboration for today's supply chain management is to maintain a long term relationship with suppliers by having a fewer number of suppliers with reliable performance. Hence, it is important to evaluate the suppliers' performance effectively in order to maintain the right suppliers (Ho, Xu and Dey, 2010). Since automotive companies are especially dependant on their sub-component suppliers, their performance is much affected by their supplier performance in delivery time, reliability and flexibility, according to Krause, Handfield and Tyler (2007). It means if a supplier improves its production time then its industrial customers could get their order faster as a consequence. Therefore, to evaluate the performance of their suppliers is very important for buying companies' performance.

As a multiple criteria decision-making problem, supplier evaluation can have several quantitative and qualitative criteria. The relationship between these criteria and supplier performance could be complex (Rezaei, Fahim, and Tavasszy, 2014). While existing papers mainly discuss supplier evaluation for the purpose of choosing the right supplier, which belongs to a pre-evaluation at a strategy level, very few papers are focusing on adopting post-evaluation at an operational level ((Khaldi et al., 2017). Only Khaldi et al. (2017) adopt artificial neural network algorithm to evaluate and predict the hospital's suppliers performance from their transactional contracts and paperwork of delivery articles including delivery delays, the number of partial deliveries, turnovers, amount of orders. The output of the prediction model is the efficiency score of suppliers. Jiang et al. (2013) conduct an experiment to forecast new suppliers' classification in terms of their performance and efficiency. They train the support vector machine model with the input of cost reduction performance, price, delivery, quality.

For predicting supplier's lead time deviation, in essentials, it is a supplier evaluation task which focuses specifically on suppliers' delivery precision performance. Delivery precision or delivery reliability refers to the ability to delivery according to schedules or promises (Sarmiento et al., 2007). The higher the delivery precision, the lower the deviation of lead time. This research has not been performed previously to our best knowledge.

2.1.3 ETA/Lead Time Prediction

For TLT prediction, there are literatures developed in each transportation scenario, such as train, road and flight. However, according to a literature review conducted by Van der Spoel, Amrit, and Hillegersberg (2017), there is very few literature predicting arrival time focusing on trucks. Therefore, this study considers to learn from the practice from each mode of transportation, one up-to-date paper is chosen and described for a review and summarized into Table 2.1.

Van der Spoel, Amrit, and Hillegersberg (2017) state that unlike the travel time which may be well predicted by using weather and traffic information, the truck arrival time could be much affected by human and organizational factors such as planning departure time. That means there is the difference between predicting lead time and arrival time. The result of lead time prediction cannot be directly applied to arrival time prediction without considering planning departure time. They test it by predicting arrival time only using those weather and traffic information. The response output is classified by the tardiness of trucks arriving at the distribution center. The classes are roughly from very early and slightly early to very late and slightly late. They test a set of algorithms such as random forest. Finally, the result is as estimated. The prediction power of the developed models for arrival time is not satisfying since human and organization factors are not included as features.

Belcastro et al. (2016) predict flight delays by focusing on weather condition since the weather is the cause of delay for more than 1/3 of the flights. They have high precision and recall score up to 86% for a large delay threshold to be 60 minutes. The threshold means when a flight arrives more than one hour later than the ETA, this flight is counted as 'late'.

Barbour et al. (2018) predict the travel time of a freight train in real time in order to generate ETA. A full network state information from transportation handler including physical train characteristics and train crew information are the input for having regression results. Compared to the current analytical method calculating the travel time only considering the network topology and traffic through particular routes, they manage to improve the performance by over 60% using random forest.

Table 2.1: Review of predicting ETA/machine learning with machine learning

Author(s)	**Subject**	**Classification/ Regression**	**Input data**	**Model**	**Remark**
Van der Spoel et al., (2017)	Truck arrival time at Distribution center	Classification	Traffic information, Weather information	M1 ensemble, Random Forest...	Low prediction power 72% accuracy
Belcastro et al.(2016)	Flight delays	Classification	Weather Condition Flight information	MapReduce	Accuracy 85.8% Recall 86.9%
Barbour et al. (2018)	Freight Train Arrival Time (travel time)	Regression	A full network state including physical train characteristics, train crew information	Random forest, Support vector regression, Neural network	maximum predictive improvements of over 60% using random forest compared to the current method

2.1.4 Conclusion from Frame of Reference

From the frame of reference, we can conclude that implementing machine learning model on predicting suppliers' delivery precision is an unexplored topic. Existing literature only implements machine learning to predict the overall performance of suppliers based on multi-criteria. Therefore, it remains to explore whether supplier delivery precision can be predicted with machine learning models from the buying companies.

Similarly, plenty of work has been done on predicting ETA for various transportation modes but few of them focuses on truck. For flight delay prediction, since the weather is one of the major causes for the delay, only considering weather and flight information could generate a good prediction result with machine learning. However, for predicting ETA of the truck, only considering weather and traffic information is not enough to have good prediction power since organization and human factors could frequently cause deviation in arrival time. When a full network state information including human and organization factors is used for predicting ETA of the freight train, a significant improvement of prediction is made compared to the previous prediction model where only traffic and route information is used. Therefore, our work will try to consider organization and human factors into the prediction model for ETA of trucks, since it is unexplored which information could be effective to be used as input features for machine learning models to predict delivery precision of LSP.

2.2 Machine Learning Tool and Terminology

This section is going to introduce machine learning and its relevant terminology such as input and output, algorithm selection, classification and regression models, boosting and bagging, random forest, catboost and gradient boosting, handling class imbalance.

2.2.1 Fundamental Machine Learning Definition

Machine learning is a field covering the main techniques used for data mining which is finding the patterns in the substantial amount of data. The discovered patterns must be insightful which can assist decision making (Witten et al., 2017). There are two extremes about a pattern, from a black box whose mechanisms are incomprehensible to a transparent box whose construction reflects the formation of the pattern. The difference between them is whether the patterns can be explained and interpreted. Both of them could lead to good predictions and knowing the inputs and outputs are way more important than understand the mechanisms in between (Witten et al., 2017).

There are some fundamental machine learning definitions. Input is including concepts, instances and feature. Concept is the thing to be learned. The input to a machine learning model is a set of instances that needs to be classified, associated or clustered. Each instance is an independent example of the concept used for learning or evaluation. There are features which is another set of predefined attributes that are measuring various aspects of the instance (Witten et al., 2017). Dimension of features measures the number of features.

There are typically two types of features for machine learning, namely categorical and continuous one. According to Prokhorenkova et al. (2018), categorical features refer to a discrete set of values that are incomparable to each other in a numerical way. The measurement scale of the categorical features consists of a different set of categories (Agresti, 2018). Categorising the features can be implemented in three different ways. The simplest one is regarded to the situations of having binary features when the values could be categorised in “0” or “1” or “YES” or “NO” segments. Furthermore, the categorical features could be mapped on an ordinal scale. For instance, they could be classified such as: “very late”, “late”, “on time”, “early” and “very early”. These features are also called “ordinal variables”. Nominal features are the final segment according to Agresti (2014). Nominal features have no numeric values and are independent of each other. These features are normally used to identify something (e.g. countries) and have not any kind of natural order. In contrast, continuous features are referred to as the variables that have an infinite number of possible values. Label is the values or categories belonging to instances (Mohri, Rostamizadeh and Talwalkar, 2012).

The input instances are divided into training set and test set. Training set is used to train a machine learning model, while the test set is used to evaluate the performance of the model. The test set is separated with the training set and not available at the training phase. The output of the model is the form of prediction on new instances (Mohri et al., 2012).

2.2.2 Algorithms and Feature Selection

Knowing which algorithm is likely to deliver a good performance for the investigated problem is known as an algorithm selection problem (Rice, 1976). There is no universally best algorithm for solving a vast problem domain (Wolpert and Macready, 1999). Identifying the most suitable machine learning algorithms which can discover the relationship between the output and the relevant features is a challenging issue (Lingitz et al., 2018). It is necessary to well understand the characteristics of the problem in order to choose the suitable algorithms (Smith-Miles, 2009).

There is the ensemble method which can adopt multiple machine learning algorithm to achieve better predictive performance. Based on the different strategy, it is categorized into boosting and bagging. García-Pedrajas et al. (2012) describe the function of boosting by saying that it builds an ensemble in a step-wise manner by making a new classifier and add it to the ensemble. The logic of this process is that the new classifier would be trained towards the biased samples. If any sample has been misclassified during the boosting process they will be assigned by a higher weighted value (García-Pedrajas et al, 2012). Boosting is a general method to use in order to improve learning algorithms since it is capable to reduce the errors of weak learning algorithm (Freund and Schapire, 1996).

In terms of the bagging method, it is a set of predictors based on bootstrapped aggregated samples in order to achieve an aggregated performance (Breiman, 1996). For predicting specific classes, the majority of the votes from multiple predictors for one class would be selected. For the prediction of a numerical output, the average value of the output from the aggregated predictors would be considered.

When adopted machine learning, the first decision is to choose between supervised machine learning which assumes that training examples are labelled, unsupervised machine learning which has focused on the analysis of unclassified examples, or other techniques such as semi-supervised machine learning or reinforcement learning (Lingitz et al., 2018). Semi-supervised learning consider both labeled and unlabelled data which is commonly used when some labeled data are expensive to obtain but unlabeled data could also help achieve better model performance. Reinforcement learning is intermixing the training and test phase, for each move receive immediate rewards to help prediction(Mohri et al., 2012). According to Öztürk et al (2006) supervised learning is considering the relationship between the output and the independent or explanatory features in a model. It aims to predict output based on input features with a prerequisite of a known training set (Pfeiffer et al., 2015).

Feature selection is another key process in machine learning. There are many possible benefits with feature selection: decreasing dimensions for improving prediction performance, providing faster and effective predictors with lost cost, assisting to understand the underlying process of data generation (Guyon and Elisseeff, 2003). According to Dash and Liu (1997), in real word practice, most classification problems require the supervised learning with each instance associated with a class label. Since the relevant features could not be known beforehand, the candidate features are often selected for their representativeness for the domain. Unfortunately, many of these candidate features are often irrelevant or redundant to the output concept and not affecting the output result. However, as soon as the size of features or dataset are up to hundreds to thousands, reducing them could significantly increase the speed of machine learning (Dash and Liu, 1997; Guyon and Elisseeff, 2003)

2.2.3 Classification and Regression models

Classification and regression are two important data mining missions for supervised machine learning. Both of them contribute to building a data-driven model to learn an unknown underlying function that illustrates the relationship between several input features and one target variable as the output of the function (Cortez and

Embrechts, 2013; Lingitz et al, 2018). To compare the regression and classification model, this selection should be based on predictive capability, computational requirements and explanatory power (Cortez & Embrechts, 2012).

The difference between these two types is made by the existence of categorical and continuous features in a model. When the output in a predictive model is set to be categorical variables then the classification techniques would be used. In the case of having a output in the form of a continuous value, the regression techniques would be applied (James et al., 2013).

2.2.4 Random Forest

Random forest has combined two powerful algorithms namely bagging and random feature selection (Breiman,2001; González et al., 2014). According to Breiman (2001), random forest is an ensemble Classification and Regression Trees (CART) classifiers, that each decision tree is created without any pruning and bagging algorithm is applied in order to create a “forest” of classifiers voting for specific labels. Each tree is considered as a predictor. Random forest could be used for both classifications and regression problems. Pfeiffer et al. (2015) adopt the random forest regression to estimate the lead time as a continuous output variable. They argue the random forest model has better performance than the decision tree model and multiple linear regression model. According to González et al (2014), random forest is capable to capture the complex interactions with different data structure and it is also robust to over-fitting problems.

2.2.5 Gradient Boosting and Categorical Boosting

Gradient boosting has been used as an advanced machine learning technique for many years, which can handle complex data sets in an effective way. According to Zhang & Haghani (2015), gradient boosting is a regression tree based algorithm that builds a model in a stage-wise fashion and updates it by minimizing the expected value of certain loss function. Gradient boosting basically applies gradient descent in a functional space to build ensemble predictors. Friedman (2001) describe gradient boosting as an algorithm that is highly robust and explainable for both regression and classification problems.

According to Prokhorenkova et al. (2018), categorical boosting(Catboost) is the execution of gradient boosting that uses binary decision trees as base predictors. In Catboost, the decision trees have the same split criterion along with the entire level of the trees. These trees are less prone to over-fitting and have a higher speed of processing time for the testing data set. Prokhorenkova et al. (2018) claim that Catboost outperformed the other advanced gradient boosting algorithms, XGBoost and LightGBM on plenty of different machine learning tasks. Dorogush, Ershov and Gulin (2018) introduce Catboost as an algorithm that has been successful in dealing with categorical features which are in practice very hard to deal with. The authors also mention that Catboost algorithms can handle the over-fitting problem in a convenient manner.

2.2.6 Handling Class Imbalance

Handling class imbalance distribution is a significant topic happening frequently in practice. Class imbalance arises when classes are represented unequally. Namely, most of instances are labelled as one class, while the rare instances are labelled as the other class which might be of more interest or importance. It is crucial that a classification model should be able to achieve higher identification capability on the rare occurrences in datasets. Many traditional classifiers are not compatible with the learning task with imbalanced classes (Kotsiantis, Kanellopoulos and Pintelas, 2006). According to Ali, Shamsuddin and Ralescu (2015), there are two problems in handling class imbalance. One of the main concerns is that data mining performers could be accuracy driven. The traditional way of examining a model performance focus on accurate performance. Classification algorithms selected for their high accuracy performance are likely to group all the data into the majority class to minimize overall error. This is often at a cost of misclassifying the rare instances. In a class imbalance dataset, classification accuracy tells very little about the minority class and may give a misleading evaluation of classifier performance. Another issue in learning with class imbalance distribution is that standard classification algorithms are based on the assumption of the evenly distributed underlying training set. Failing to consider the skewed distribution of data is most likely to hinder the classification performance (Ali, Shamsuddin and Ralescu, 2015).

The classification performance for imbalanced data is also subjective to the size of the dataset (Kotsiantis, Kanellopoulos and Pintelas, 2006). It may be even worse for an small imbalanced dataset compared to the larger one, due to the insufficient sample size of instances representing minority class for learning. On the contrary, the effects are relatively less severe with larger datasets, as the minority class is better represented by a larger size of examples (Kotsiantis, Kanellopoulos and Pintelas, 2006).

To handle class imbalance classification, sampling techniques and cost-sensitive learning are commonly applied. Sampling techniques are used to either remove a small number of examples from majority class or over-sample minority class or both. By introducing this sampling step, the discrepancy between the two classes is minimized so that traditional classification algorithms can work well. For example, Balanced Random Forest, incorporating under-sampling majority class technique and the ensemble learning, artificially re-balances the class distribution to ensure that classes are equally represented in each tree (Chen and Breiman, 2004).

Cost-sensitive learning approaches, on the other hand, impose an expensive cost on a classifier when a misclassification happens in order to emphasize any correct classification or misclassification regarding the minority class (Kotsiantis, Kanellopoulos and Pintelas, 2006). For instance, in Boosting algorithms, different weights are placed on the training distribution in each iteration. In order to emphasize misclassified examples in the next iteration, boosting increases the weights on the misclassified examples and decreases the weights on the correctly classified examples after each iteration. Since minority classes are more likely to be improperly classified in comparison with majority classes, boosting may improve the classification performance through increasing the weights of the examples from rare classes. Also, as boosting effectively rebalance the distribution of the training data, it can also be considered as an advanced sampling technique (Kotsiantis, Kanellopoulos and Pintelas, 2006).

2.3 Evaluation Metrics for the Prediction Models

Since the overall accuracy could insufficiently or even misleadingly evaluate a classifier performance (Visa, 2006; Japkowicz and Stephen, 2002; Wang and Mendel, 1992), the confusion matrix and its derivations are introduced as a more proper way to summaries the performance results. Feature importance is also introduced as another measurement for the input features.

2.3.1 Confusion Matrix

A confusion matrix shown in Table 2.2. is typical for evaluating the machine learning models' performance with imbalanced classes. Class "C" is regarded as the minority class which is in the focus, while "NC" is a combination of all the other classes. There could be four kinds of results when detecting class "C" (Chawla et al., 2003). The first one is true positives which correctly recognized focused class examples. True negatives are those correctly identified examples that do not belong to the focused class. The third factor, false positives, considers the examples that were incorrectly assigned to the focused class and finally the last one is false negatives which were not successfully recognized as focused class examples. These four factors constitute a confusion matrix (Chawla et al., 2003).

Table 2.2: Confusion matrix defines four possible scenarios when classifying class "C" (Chawla et al., 2003)

	Predicted Class "C"	**Predicted Class "NC"**
Actual class "C"	True Positives (TP)	False Negatives (FN)
Actual class "NC"	False Positives (FP)	True Negatives (TN)

From Table 2.2., recall, precision and F-value are defined as follows:

$$Precision = TP/(TP + FP) \tag{2.1}$$

$$Recall = TP/(TP + FN) \tag{2.2}$$

$$F - value = \frac{(1 + \beta^2) * Recall * Precision}{\beta^2 * Recall + Precision} \tag{2.3}$$

The performance metrics derived from the confusion matrix are including precision, recall, F1 score which comprise of a classification report for the modelling result. Precision measures the exactness, which is the proportion of correctly predicting classes. It shows the ability of a classifier for avoiding misclassifying negative classes into the positive class. Recall measures the completeness, which represents a classifier's ability to learn positive class. It is calculated by the proportion of correct detection of positive example out of all positive example in the data. F-score is a way of balancing the measurement between precision and recall. As the β is commonly set to 1, therefore F1 score is used for classification (Sokolova and Lapalme, 2009).

The common pursue of all learning model is to improve the recall while not to sacrifice the precision. However, there is often the conflicts between them and it may be difficult to improve both of them at the same time. This situation is especially true when one or more classes are rare (Chawla et al., 2003).

2.3.2 Feature Importance

The increasing popularity of machine learning models is largely credited to their capability to handle high-dimension data with large number of predictors and other advantages including relatively good accuracy, robustness, ease of use (Breiman, 2001). However, it is common that not all the features are important and some of input features can be relative irrelevant or redundant in data mining. Identifying the most important features is beneficial because it indicates which features have the highest predictive power for the model and may help the domain users to have a better understanding of the problem. It can also help to develop recommendations for the future, and it may lead to changing the role of the underestimated features more seriously (Petkovic et al, 2016). To identify the features with the most significant impacts on predictions, feature importance is one of the most commonly used measurements, which facilitates feature selection and model interpretation.

The most widely used feature importance measures are the impurity importance and the permutation importance (Breiman, 2001). The impurity importance, also known as Gini importance, is based on the mechanism of mean decrease of impurity. It is the default feature importance measure embedded in some most popular implementation platform such as R and scikit-learn in Python. In the impurity importance, a feature is considered as important if it is effective at diminishing uncertainty for classifiers or variance for regressors. The impurity importance for a feature in random forests, for example, is computed by adding up all impurity decrease measures of all nodes in the forest where a split on this feature has been made, normalized by the number of trees. Another type of importance measure, the permutation importance is also known as mean decrease of accuracy. Under this mechanism, the important features are those positively contributing to reduce the prediction error.

Despite its popularity, for years, the impurity importance is acknowledged to be biased. The impurity importance is likely to inflate the importance of categorical variables with many categories and continuous variables (Breiman et al., 1984; Strobl et al., 2007), also in favor of variables with high category frequencies (Nicodemus, 2011). The permutation importance, on the other hand, is safe from these concerns (Nicodemus et al., 2010; Szymczak et al., 2016; Ziegler and Konig, 2014). However, the permutation importance can be extremely computationally intensive when encountering high dimensional data. Also, Calle and Urrea (2011) argued that feature importance rankings based on the impurity importance can be more robust over those obtained with the permutation importance.

3
Methods

In this chapter, the methods that were used to conduct this project are described. First, the literature review was then conducted and also throughout the entire process of the project. Then the CRISP-DM (Cross Industry Standard Process for Data Mining) methodology and the reasons for choosing it are introduced. The correlation between methods and research questions is also demonstrated. Finally, the reliability and validity issues are described in the end section.

3.1 Literature Review

There are several reasons for conducting a literature review at the beginning of and throughout the project. Bryman and Bell (2015) describe the first thing is to be aware of and understand what has been already discussed in the research area. Secondly, it also gives a way for authors to develop an argument about demonstrating the significance of the project and where it contributes. Beyond that, a literature review with an interpretation from reliable sources in the research field could also increase the credibility of the project. Based on the above reasons, a literature review was conducted with the purpose of providing information for four research questions and assisting the data mining process for realizing the aim of the project.

We searched literature from electronic database including Scopus, Google scholar and Chalmers Library. The keywords used in the search including the combination of lead time deviation, estimated time of arrival (ETA), prediction, delivery precision, machine learning, supplier evaluation, automotive. Peer reviewed articles and books were examined and used in the literature review.The result of the literature review is compiled in the chapter 2.

3.2 General Strategy and Process

The most commonly used process for data mining projects is CRISP-DM (Marban, Mariscal and Segovia, 2009) It is process model being developed by a group of data mining leaders for carrying out data mining projects. The purpose of this process model is to make these projects more reliable and replicable with less money and time spent (Wirth and Hipp, 2000). Wirth and Hipp (2000) discuss that the process can not only be performed by experts, but the novices with less experience and technical skills can benefit in a limited time. This is due to the characteristics of CRISP-DM being both structural and flexible depending on whether it is generic or specialized process. For less experienced people such as master students, we can get guidance and structure of the project, as well as advice for each process.

The processes of CRISP-DM from generic to specific are described as Phases, Generic Tasks, Specialized Tasks and Process Instances (CRISP-DM, 1999). For the top level, the phases of the model include business understanding, data understanding, data preparation, modelling, evaluation and deployment representing the life cycle of a data mining project. The second level is generic tasks with its intention to cover all data mining situations. The third level aims to describe what actions should be taken within the general tasks. The fourth level is a requirement of recording the actions, decisions and results during the process.

We adapted the generic CRISP-DM process based on our data modelling project, and the process is summarized in Figure 3.1. There are six phases in the CRISP-DM process that are described in the following sections.

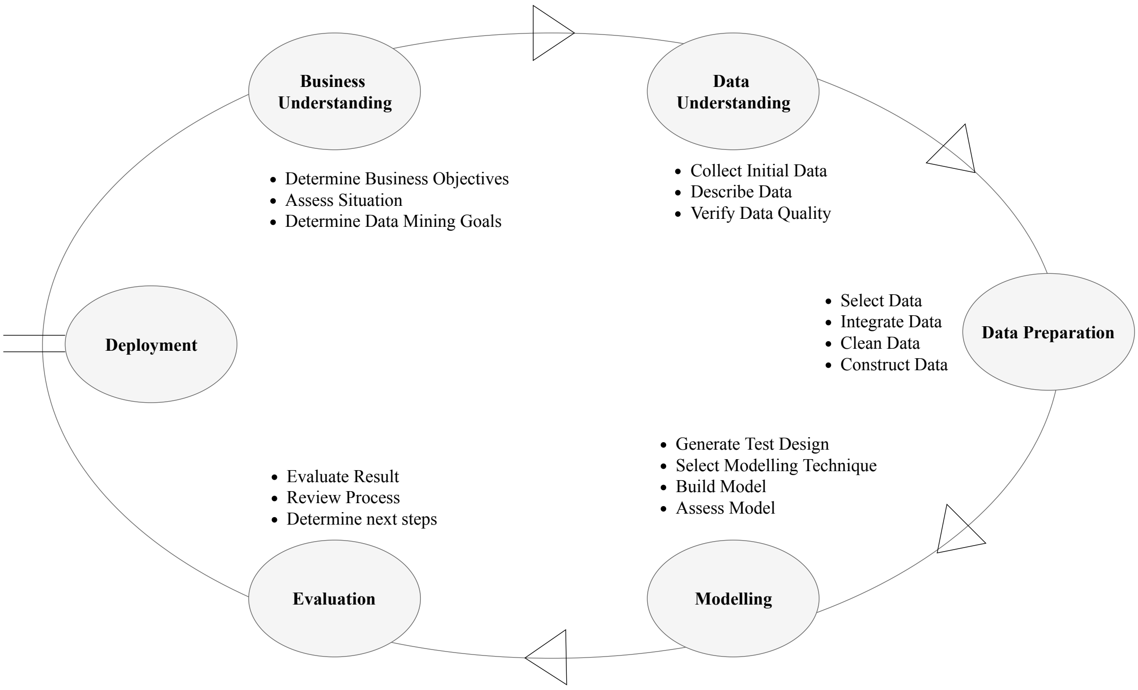

Figure 3.1: Illustration of the data mining process based on CRISP-DM (1999)

3.3 Business Understanding

The first phase is about understanding the business. Business understanding involves figuring out the feasible goals based on the situation and requirements from the business perspective to achieve potential benefits. Therefore, qualitative data about business were collected by means of conducting interviews and examining internal documents in the company in order to set a feasible goal.

3.3.1 Interview

As Yin (2018) says, one of the most important sources of information is interviews since they can help with explanations of key events and providing insights from the interviewees' point of view. The interviewees were selected based on the scope and aim of this project, including representatives such as Continental Material Planners (CMP) and Transport Material Coordinators (TMC). The interviewee list is seen as Table 3.1. Both semi-structured and unstructured interviews were adopted since interviews in the case study are like guided conversations (Yin, 2018). Unstructured interviews were held throughout the project when there was a need to get clarification of any concepts and questions. Semi-structured interviews were adopted to gain an initial understanding of the researched topic. Questions about the performance of lead time and the relevant factors that are associated with lead time deviation were asked. During the semi-structured interviews, audio recordings were collected for the purpose of capturing all the information from the answers, by enabling authors to revisit the answers from interviews. The transcription was generated by one author in the interview with the aid of audio recordings and the interview results were then examined by the other author attending the meeting. The interview templates used in the semi-structured interviews are presented in Appendix A.

Table 3.1: A table for interviewees list

Title	**Interview Topic**	**Date**
Operational Resource Planner	CDC management	2019-02-08
Refill Analyst	Outbound Logistics	2019-02-11
Demand and Inventory Planner	Demand Forecast	2019-02-11
Continental Material Planner	Monitor Material Suppliers performance	2019-02-18
Transport Developer	Transportation Lead time set up	2019-02-26
Supplier Relationship Manager	Material Suppliers evaluation	2019-02-27 2019-03-05
Manager Supplier Management	Evaluation of Logistics Service provider	2019-03-01 2019-03-07
Logistics developer & Business Analyst Material Planning	Data extraction for supplier lead time	2019-03-08
Data Scientist	Modelling	2019-03-15
Transport Material Coordinator	Monitor Logistics service provider	2019-03-25
Project Manager	Lead time strategy	2019-04-02

3.3.2 Internal Documents

For a case study research, the most crucial value of internal documents is to authenticate and argue the evidence from other sources (Yin, 2018). With the internal documents, this project gained up-to-date knowledge about the case company's structures and business processes. The collected information also became evidence to support arguments from interviews. The internal documents used in this project were found in the internal database of the case company, including company presentation, process description system and the team places. These documents could exist in the form of PowerPoints, word documents and other informative data from databases.

After this stage, the goal of the business was defined to respond to the research question 1. To answer the research question 2 about the factors of lead time deviation, a pile of factors were compiled after conducting the literature review, interviews and examining internal documents. A list of preliminary potential features was also identified in this process.

3.4 Data Collection and Understanding

For the data understanding process, one investigated aspect was to collect the historical data of lead time deviation performance, which was used as the output variable for modelling. Another aspect was gathering those available data that could associate with factors of the deviation of lead time identified in the first stage. These quantitative data were extracted from different databases in the case company as archival records, as Table 3.2 shows. Historical lead time performance data of supplier lead time was extracted from the Business Intelligence where the previous two years data (2017 and 2018) were included. The data related to features of the first model were also extracted from business intelligence and the reports generated from supplier management portal VSIB. For the second model, most of the data were extracted from the logistics management portal Atlas. These data were limited to the previous one rolling year as the maximum amount of data the system held at the time the project was conducted. Noted from the transportation delivery precision report, there is up to 30% of delivery where goods were not delivered according to planned deliveries. These missing deliveries were deleted and not considered into the calculation of delivery performance since they are not generated the output of delivery whether they are on time or deviated.

Then data understanding was to get to know the data about its variability and availability, including the quality and quantity of the data. Since the business goal needs to be translated into the goal of data mining, the availability of the data in the company was under consideration. Hence the data mining goal was developed.

Table 3.2: Main data source from the case company's database

Phase	Data	Sources	Content
Supplier Lead Time	Historical lead time performance (Output Variable)	Business Intelligence – Parts-DWH ver1.5 – For Std Report Developer	2017&2018: 402,708 pieces of records
	Features	Parts: Business Intelligence – PartsDWH ver1.5 Suppliers: VSIB – supplier management portal	Segmentation, Sales level spend, delivery precision, ...
Inbound Transportation Lead Time	Historical lead time performance (Output Variable)	LSPs portal Atlas Filter: all volvo truck parts were ended in CDC Ghent	2018.04-2019.03: 49,948 pieces of records
	Features	Parts: Business Intelligence – PartsDWH ver1.5 Suppliers: LSPs portal Atlas Consignors and Volvo logistics scheduling: Atlas	Weight, volume, country ...

3.4.1 Delimitation in the Data Collection

There were a few limitations in the data collection phase. Firstly, when sampling data from the data warehouse, the period was limited to what the data warehouses hold. For the transportation phase, the data are recorded for one rolling year. Therefore the amount of data for training were limited to one year period, which could bring problems of bias and robustness.

The evaluation results of material suppliers were extracted from the supplier management system VISB. The options for evaluation period are from past three months to past one year, the granularity of the evaluation results such as delivery precision is limited by being made as average value for that chosen period.

There were data related to factors that were scattered in lots of separate reports but not integrated into the data warehouses. In this sense, these data were not able to be gathered and used as features for modelling. For example, the logistics audit results of LSP exist in individual excel files for each LSP, then these data were not utilized as a potential feature.

There were factors that relate to deviation but suffering from the data quality in the system and not being used as a feature. For example, the departure time of truck could have effects on deviation since it affects the arrival time of a truck to a warehouse which could cease operation during the night and the late arrival truck need to wait for one night to be processed. However, the departure time is not precisely recorded in the system and therefore not suitable to be used.

There was the data transparency issue that the names of some items in the databases were confusing without further explanation. In order to make sure the right data was used, it also took time for us the data practitioners to find who can explain the data in the company.

Sensitive information such as the relationship between suppliers delivery performance and evaluators in the buying company was also not gathered and examined.

3.5 Data Preparation

The data preparation phrase is including all the actions that creating a final data set which were fed into the modelling from the raw data including selecting data, cleaning data, constructing data, integrating and formatting data.

3.5.1 Transferring Categorical Variable

There were many categorical variables in the feature list, in order to quantify them and feed them into modelling, a function called dummies in the commonly used python package Pandas was used to turn a categorical variable into a series of zeros and one. One example is illustrated below, the feature of categorical variable ‘stackable’ is divided into two columns with ‘1’ represent of the characteristics being true, and ‘0’ for not being true.

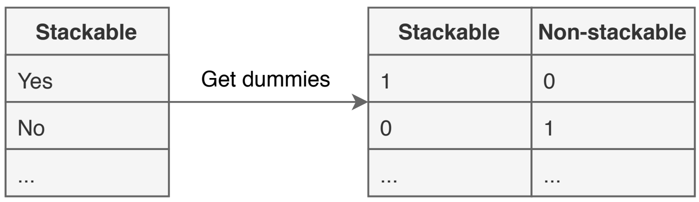

Figure 3.2: Transferring categorical variables into dummy variables

However, some categorical variables have a lot of classes such as 59 kinds of segmentation of spare parts. When directly getting dummies for these variables, the input data will get lots of columns with each one having little weight. Therefore, these categorical variables were reduced into a reasonable amount of columns by reconstructing and combining them based on some criteria. Segmentation of Volvo spare parts is a comprehensive measurement defined in terms of criticality, life cycle, cost and order frequency. For segmentation result, there are five different initial letters from ‘A’ to ‘E’ as main catalogues. From ‘A’ to ‘D’, they represent four kinds of criticality code, and ‘E’ represents non-critical parts. The criticality of a part depends on specific function groups and vital codes. Under each letter, there is the second letter starting from ‘A’ to ‘L’ for the sub catalogues representing the cost, life cycle and order frequency information. Vital code, cost, order frequency are available as independent features, while using function groups directly may result too many categories, and life cycle phase is not directly available. Segmentation was adapted to present information of function groups by keeping the main catalogues, and clustered the second catalogues into ‘fast’ and ‘slow’ to roughly represent the life cycle phase. After the modification, the segmentation was simplified into ‘A-fast’, ‘A-slow’, ‘B-fast’, ‘B-slow’ and so on to roughly reflect the function groups and life cycle phases.

3.5.2 Integrate and Link Data

After the previous phases of understanding, we realized there was the need to build two prediction models for the two phases, since the deviation could happen in each phase and the detection of deviation is necessary to take actions in each phase. For the modelling of supplier lead time deviation, , the information of parts and suppliers were integrated into the records of delivery precision performance Then, for building the models of transportation lead time, to consider the previous delivery precision performance from material suppliers could also be beneficial. However, the data of two phases in the company are independent. They are separated into two systems, managed by different departments and not linked with each other. In consequence, there is no information about which parts are carried in the shipments from the transportation booking. We manually linked the instances from these two phases, using event time (Dispatch week in material suppliers records, Prove of collection date in LSP records) and companies (supplier ID in material suppliers records, consignor ID in LSP records) as linking keys. When these two keys were in line with each other in two instances, these two instances were integrated and regarded as the same ordered flow as Figure 3.3 illustrate. This linkage can help the prediction of TLT to have more potential features including relevant parts and material suppliers information.

Another issue is that one transportation booking could contain several ordered parts, therefore, when left joining parts information into the transportation booking records, several transportation booking instances were duplicated with the only difference of part information between them. Then, in order to integrate these duplicated instances into one independent instance, the information for those parts in the same transportation booking was used their average value in this project.

3.5.3 Delimitation in the Data Preparation

For data preparation in modelling supplier lead time deviation, normally there are existing several orders for a spare part with one supplier in two years duration. Even though the differences between these orders and further integrated features could be only the event time, the deviation could differ from one order to another order. Therefore, all the orders kept for input instance for the benefits of representing the real case, although this might sacrifice the variance of each feature in each instance and affect the model performance and the result of feature importance.

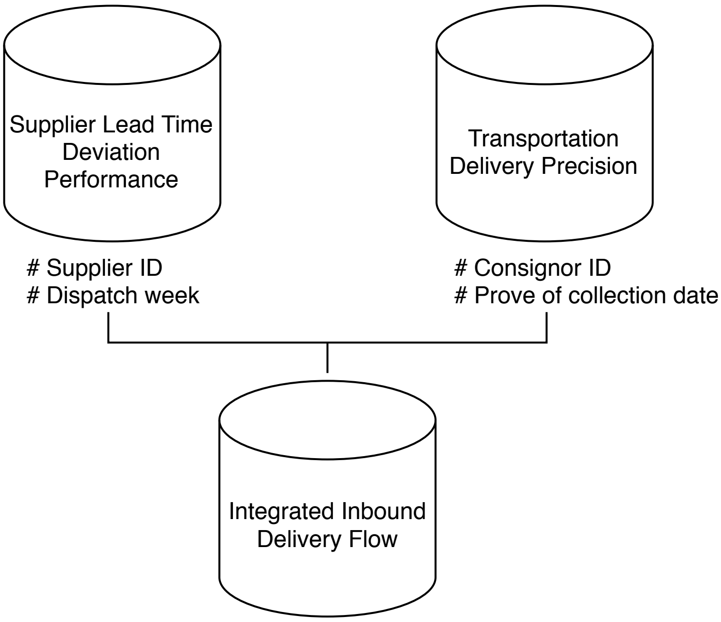

Figure 3.3: Integrate and link data for the phases of material supply and transportation

There are three ways of transportation, namely Door to Door (DDT), Cross-docking and Milk run. For the transportation mode using cross-docking, the transportation booking reservation is separated into two independent transportation booking records. The previous cross-docking is from material supplier to cross-dock point, while the later cross-docking process starts from cross-docking point to CDC. The consignor for the second transportation booking records, therefore, becomes the cross-dock point. In this way, the second phase of cross-docking transportation failed to be linked with previous corresponding records of material suppliers due to the key of supplier ID and consignor was not to be matched. Only the previous leg of cross-docking were linked.

Another limitation happened for the milk run transportation. Even though one milk run generates one transportation booking, with the two keys can be in line with the first material supplier in the milk run, the information of the remaining suppliers and parts information failed to be considered into the input instance for the milk run transportation. As Figure 3.4 shows.

3.5.4 Feature Selection

To represent previous identified factors into candidate features for modelling, there were a few cases occurred in this process. Firstly, there are data which can directly represent the factors such as the demand, value, stackable, hazardous, custom, evaluation results for material suppliers. Secondly, there were data representing the factors at an aggregated level, such as TB weight and volume data for the total weight and volume in one shipment, segmentation data for integrating function groups and life cycles, country for traffic and weather. Thirdly, some factors that were not recorded in the data form, such as the prioritization. Some factors' information is not available in the buying company due to that information is owned by material sup- pliers such as material suppliers' production information. These factors were tried to be indirectly reflected by other available data, such as sales spend level data on suppliers for representing the prioritization, quality and environment certificate for representing the production capacity of suppliers. However, some data currently are not integrated into the database, and we could not either find other suitable data for the indirect representation of their corresponding factors, such as historical delivery precision performance and evaluation results of LSP.

Since the dimension of input features in this project is limited, all potential candidate features were kept as input for the modelling. No further feature selection is needed for the benefits of dimension reduction which is not the case with limited feature dimension.

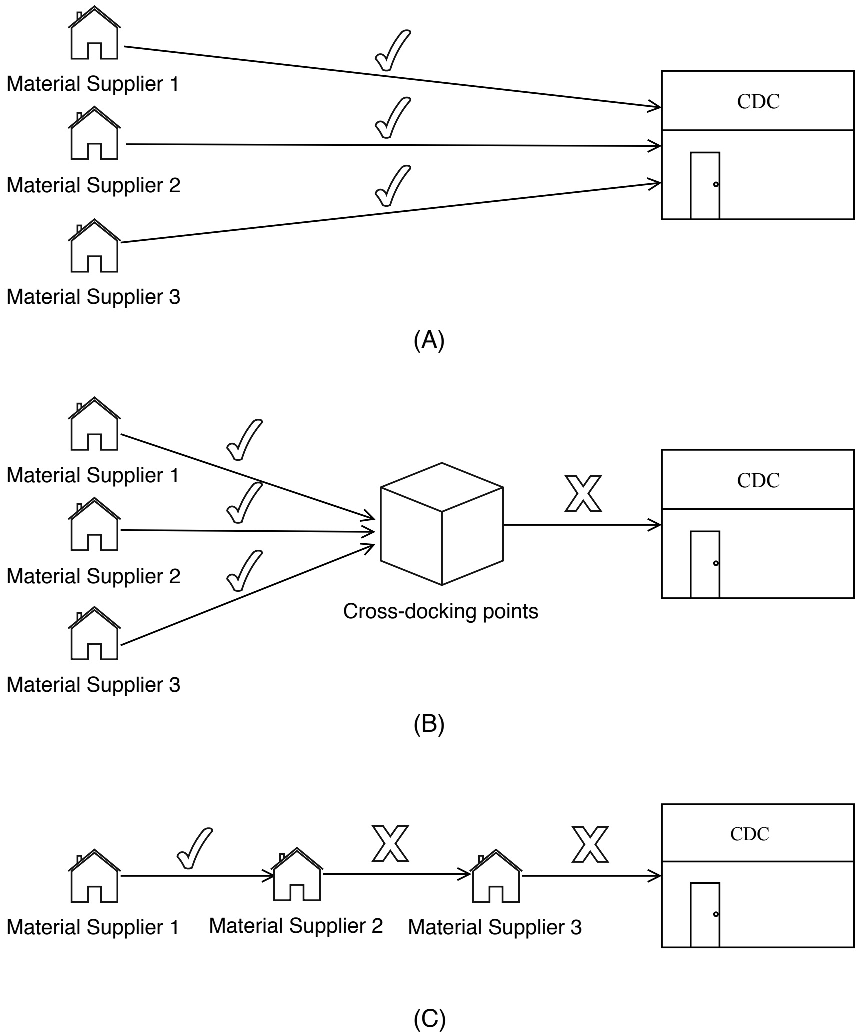

Figure 3.4: The data linkage in different transportation modes (A) Door to door; (B) Cross dock; (C) Milk run

3.5.5 Handling Missing Data

Missing data imputation is a method for filling the missing values with some probable and possible values before the process of learning algorithm begins (Lepping, 2018). Replacing each missing value for a variable by using the average observed values for that variable is a common method that may accurately predict the value of the missing data but, also leads to poor estimation of variances and correlations (Schafer and Graham, 2002). There was a proportion of missing value when we examined the extracting result. For supplier phase, these missing data particularly exist in the evaluation information for suppliers, including the Supplier Evaluation Measurement (SEM) result, logistics audit result and historical delivery precision. There could be several reasons for the missing value. For example, no evaluation has been performed or no more cooperation with those material suppliers. The degree of missing data for supplier phase was presented in Table 3.3. In comparison, for the transportation model in the data preparation stage, only successful linked and integrated records were kept, and therefore there is no missing value. The missing data were filled in with mean value in this project.

Table 3.3: Missing value for supplier lead time phase

Variables	**Number of instances**	**Missing rate (%)**
Dispatch Week	400641	0.00
Part No	400641	0.00
Supplier No	400641	0.00
Lead time deviation	400641	0.00
Parameter reference	388011	3.15
SEM result	288761	27.93
QPM score	399906	0.18
Quality Certificate	329809	17.68
Purchase agreement	400641	0.00
Sales level Spend	399906	0.18
Vital	400641	0.00
Hazardous Code	400641	0.00
Prepacking Type	400641	0.00
Country	400154	0.12
Registration Date	398617	0.51
Stand Price	398617	0.51
Order Hits Roll 13 Period	398617	0.51
Delivery Precision	362068	9.63
Logistics Audit Result	262179	34.56

So far, a list of features has been constructed as input data for modelling and research question two were covered.

3.6 Machine Learning Modelling and Evaluation

Different machine learning modelling and techniques can be chosen and tested in the modelling phase. The parameters are required tuned into the optimal values. Noted the modelling is also closely linked to its previous phase of data preparation since the new problems of data set could not be unveiled until modelling or new ideas are generated for collecting new data.

The first choice in the modelling is to choose from supervised learning, semi-supervised learning and unsupervised learning (Lingitz et al., 2018). Since the purpose of the project work is to predict the lead time deviation as the output with labelled input data from databases, supervised machine learning was used for this situation.

Based on the previous understanding of the business goal and data mining goal (Smith-Miles, 2009), the output variable is made into three classes, namely 'On time', 'Early', 'Late'. This is an imbalanced data set with the majority of the observation falling into the 'On time' class. Balanced Random forest (Chen and Breiman, 2004) and boosting algorithms (Kotsiantis, Kanellopoulos and Pintelas, 2006) could be two approaches to deal with imbalanced data set. In addition, based on the knowledge from the data scientist in the case company, several classification machine learning algorithms were selected to build the models for each phase, including Balanced random forest, Catboost and Gradient boosting. Balanced random forest has been selected as the algorithm is combining the bagging method and under-sampling technique for the majority class (Chen and Breiman, 2004). The reason for selecting the Catboost and Gradient Boosting is that both of them are using the boosting method which can give high penalty to missing classified minority class as a cost-sensitive learning technique (Kotsiantis, Kanellopoulos and Pintelas, 2006).

Finally, an evaluation process was conducted. The performances of the above- constructed models were compared and recorded using confusion matrix. The results were analyzed from a data analysis point of view. Furthermore, the improvement and deployment of the models were examined considering the fulfillment the business goal. The process of the CRISP-DM model was reviewed. Future possible actions were proposed. Until this point, the research question 4 was answered. The relationship between processes and research questions are illustrated below in Figure 3.5.

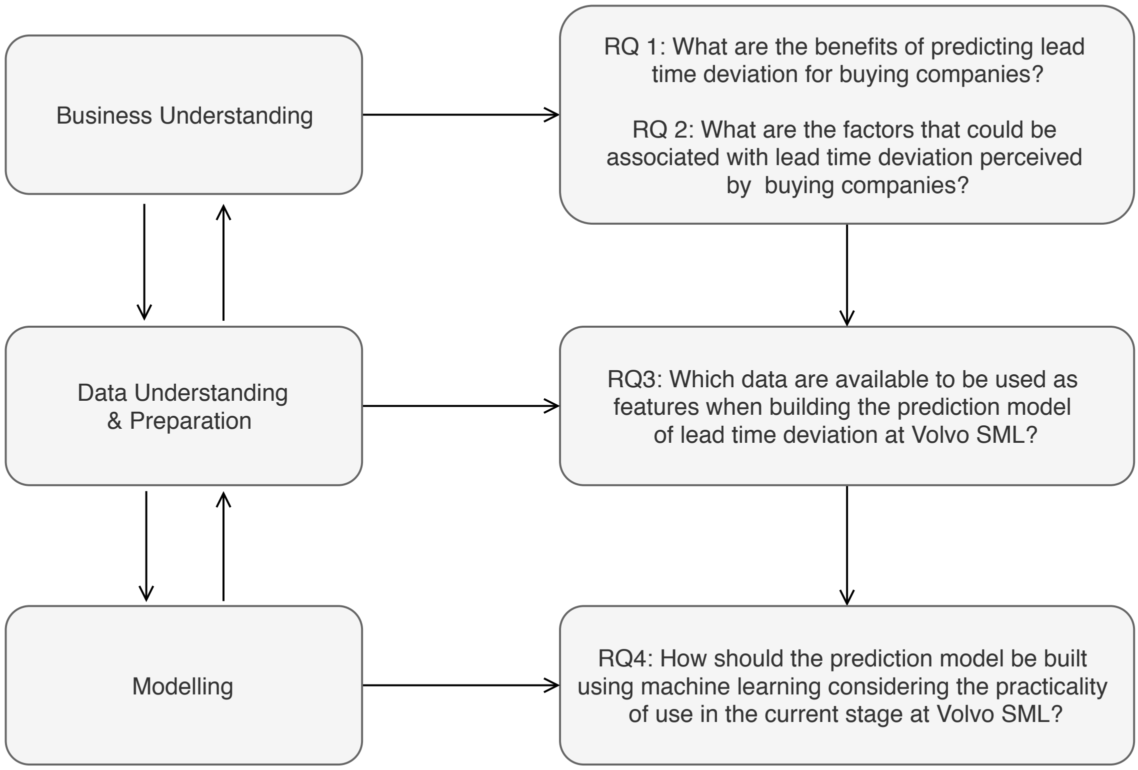

Figure 3.5: The relationship between processes and research questions

3.7 Validity and Reliability

According to Bryman and Bell (2011), there are two important aspects regarding the evaluation of the quality of research, namely reliability and validity. Reliability is about the consistency of measures, whereas validity refers to whether a measure of a concept actually manages to measure it (Bryman and Bell, 2011).

In the qualitative part of this thesis, reliability will be increased by contemplating inter-observer consistency. According to Bryman and Bell (2011), inter-observer consistency is an issue of inconsistent declaration that could happen when there are several observer-constellations judging information subjectively. All the interpretation from interviews were analysed and agreed upon by the presented interviewers. Validity in the qualitative data of research would increase through internal validity, it means that the findings from observations should fit into the theoretical framework developed (Bryman and Bell, 2011). This subject was considered during the thesis process in order to verify the findings from interviews with actual modelling further on.

During the quantitative data of the thesis, face and convergent validity were considered. According to Bryman and Bell (2011), face validity is about the process of evaluation of a model by an outside expert to see if it is reasonable. Based on this factor, a machine learning expert from the department where the thesis project is conducting evaluated the scientific aspect of machine learning algorithms in the context of this project. Convergent validity, according to Bryman and Bell (2011), considers the result of a method and compare it to the outcome of other methods from the same category. Based on this definition, convergent validity was considered for the new prediction model and the generated models were evaluated by considering some measurements such as the goodness of the fit of the model. Reliability of the quantitative data is gained by examining the correlation between different factors of lead time and the lead time deviation by using correlation analysis. In addition, the quality of the data was examined and modified.

4

Results: Business Understanding

The first section in this chapter is going to describe the company's operation around lead time including involved processes and roles. Then, the current performance of lead time deviation and its impact are presented. Furthermore, the business goal for this data mining project is set. Finally, the factors related to lead time deviation are described.

4.1 The Set-up of Lead Time in Volvo

At Volvo SML, most of the lead times are negotiated with material suppliers and LSP. As agreed, these lead times will be set as predefined parameters in the planning systems. The supply process in Volvo SML could be categorized into five processes, as Figure 4.1 shows. Inbound supply phase starts from Continental Material Planner (CMP) placing orders to material suppliers and ends till the orders are received and registered at Central Distribution Center (CDC), including supplier lead time, inbound transportation lead time and internal receiving lead time. Outbound supply phase begins after CDC have received and registered the orders until customers get their requested spare parts including outbound transportation lead time and order lead time. The shipments are carried by LSP.

Since the set up lead time between Volvo and suppliers by negotiation is an estimation of lead time, together with other causes of disruption alongside the delivery process, the deviation in lead time is inevitable. There are also cascading effects along the supply chain. For example, when the material supplier does not dispatch the orders on agreed time, that is going to affect LSP on picking up the orders and further affect later process of transportation. The affected trucks may further arrive at CDC later than the schedule and may need to wait to be unloaded since the capacity of CDC is limited. Most importantly, currently there is no existing process or tool to predict the deviation of lead time in the company.

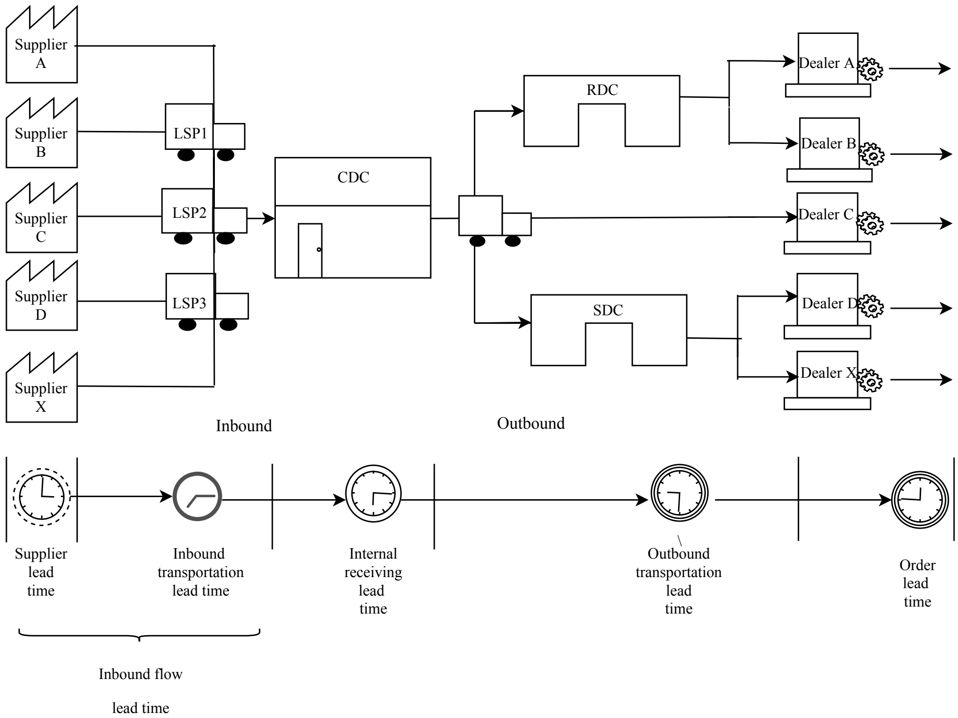

Figure 4.1: The set up of lead time in Volvo SML

4.2 The Process and Roles Involved in Dealing with Lead Time Deviation

The process and roles involved in dealing with lead time prediction are introduced in the below sections. These results lead to the setting of business goal.

4.2.1 Process Overview

The inbound delivery process behind SLT starts from Demand and Inventory Planners (DIP) generate demand forecast for CDC Ghent. The demand forecast contains information about at what time and how much of which spare part is needed in the CDC. These demand forecasts pass through the planning system. Based on the forecast information, CMP place orders to corresponding material suppliers. When material suppliers are ready to dispatch the order, they book the shipments from LSP through Volvo's transportation management portal 'Atlas'. The transportation booking (TB) contains information such as pick up and shipping address, volume, weight of spare parts. LSP will ship the order to the CDC based on transportation booking information scheduled by Atlas. Atlas portal also incorporates the transportation orders from several material suppliers by arranging different ways of delivery including DDT, cross docking and milk run.

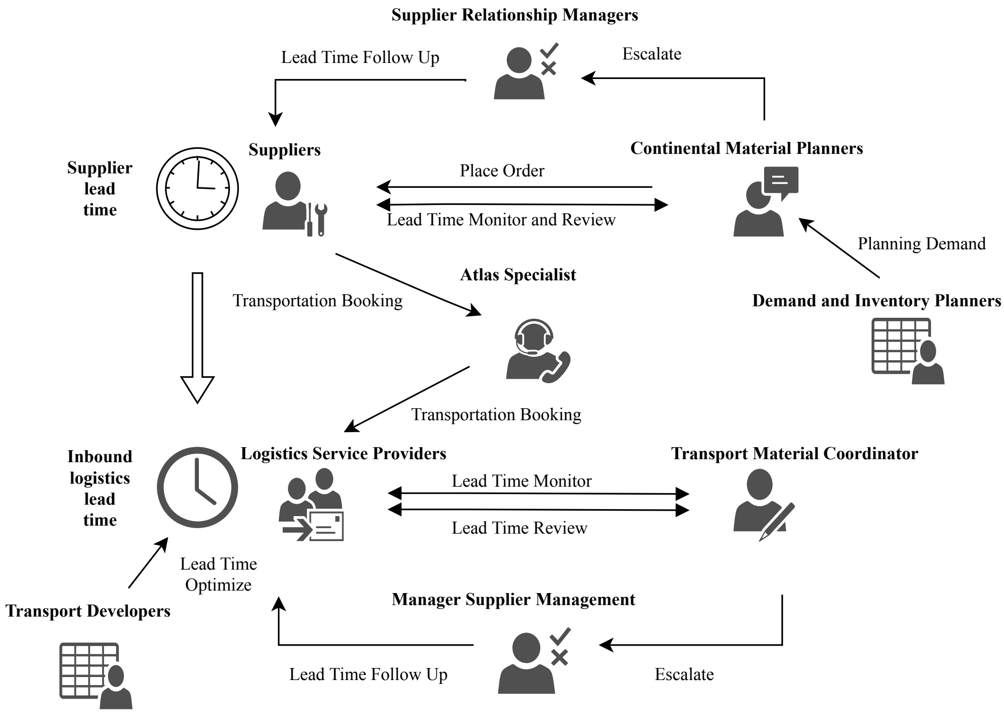

Figure 4.2: The roles involved in dealing with lead time deviation

Noted the role description is in line with current responsibilities, which could be changed from time to time. The following section is going to describe in detail the responsibility of the most relevant roles, that are the monitors and evaluators of lead time deviation, including Continental Material Planner (CMP) , Supplier Relationship Manager (SRM), Supplier Manager (SM). For managing material suppliers, Volvo has CMP for monitoring the individual level of performance on material suppliers and SRM perform a higher integrated level of management. While for transportation, TMC are responsible for managing the individual level of LSP and SM are for a higher level of measurement. Delivery precision measures whether the suppliers dispatch requested order on the scheduled time and this key performance indicator (KPI) directly links to the degree of deviation on SLT. Similarly, there is also delivery precision measuring the transportation lead time deviation from LSP representing the accuracy of ETA. The information about the key roles and KPIs for lead time performance is summarized in Table 4.1.

Table 4.1: The key roles and KPIs for lead time performance

Suppliers	**KPI of lead time performance**	**Key Roles**
Material suppliers	Delivery precision	Continental Material Planner (Monitor) Supplier Relationship Manager (Evaluator)
Logistic service providers	Delivery precision	Transport Material Coordinator (Monitor) Supplier Manager (Evaluator)

4.2.2 Continental Material Planner

CMP are responsible for the inbound material supply process for spare parts. Their mission is to ensure the availability of spare parts at the central warehouse and provide a sharp ETA to the customers. Their first responsibility is to set up SLT with material suppliers when the part is first sourced to them and then to review the lead time after a certain period of time. The guideline is to propose 2 weeks of lead time for high running spare parts which are frequently ordered, 4 weeks for the middle runner, and best possible lead time for low runners. If proposal for SLT is not accepted by material suppliers, then CMP will take what material suppliers answer to them. Lead time review is done once or twice with two material suppliers per year for each CMP. The purpose of lead time review is to shorten lead time and have lead time information alignment with suppliers. SLT is important since it determines the amount of safety stock. Besides, during the period of SLT, CMP cannot change the order from suppliers unless the change is agreed by suppliers.

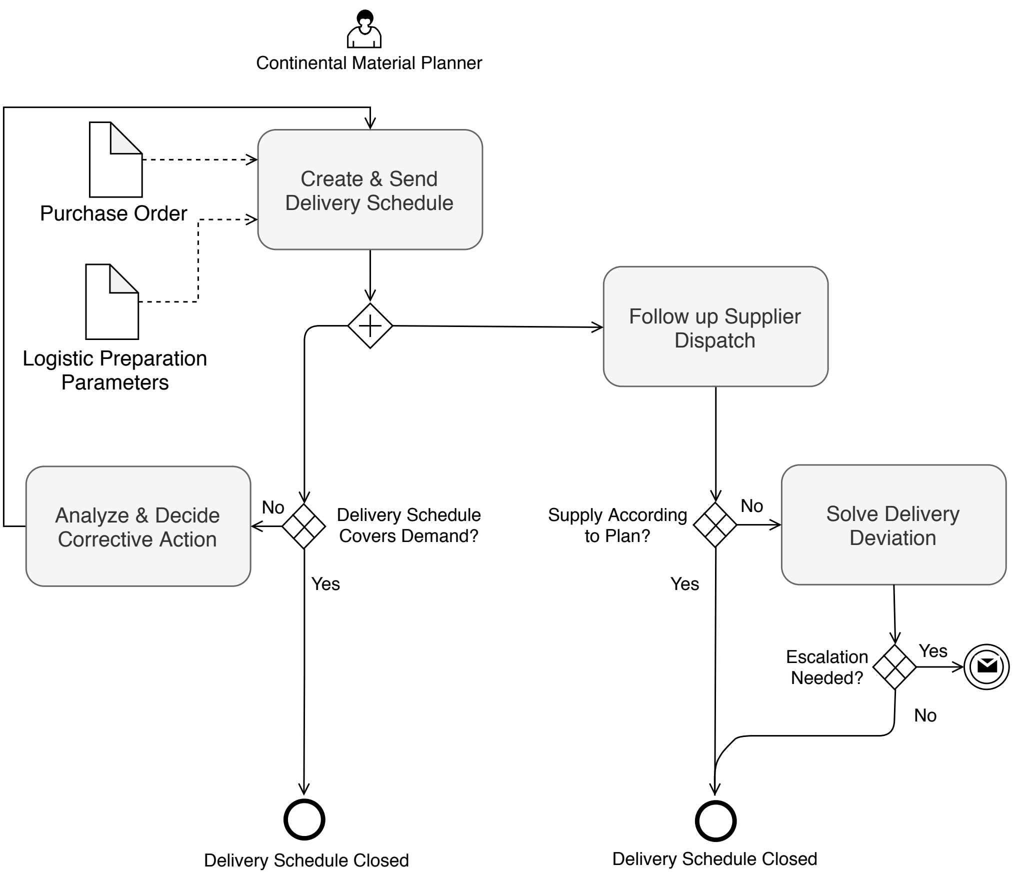

Figure 4.3: The working procedure of CMP

Another important responsibility of CMP is to place the order to material suppliers based on purchase orders from DIP and logistic preparation parameter set in the system. After placing the order, CMP then monitor suppliers' delivery precision by having frequent contacts with them. If suppliers confirm the order information, CMP send the information of ETA to the following process. If there is deviation happened in the material suppliers, CMP are responsible to figure out the reasons for the deviation and take actions for dealing with deviation. For example, if the order is dispatched later than schedules, CMP can arrange extra transport with the rush option in order to ensure the availability of spare parts. Since the rush transport causes high costs, only with critical spare parts and backorder from customers, CMP shall use this option. CMP can also decide to escalate the problematic suppliers to SRM where re-examinations of the suppliers will be performed. In contrast, if one supplier's performance is above a certain percentage for a certain period of time, CMP tend to trust this supplier and may send out the ETA information very soon without confirmation from material suppliers. The process is illustrated above in Figure 4.3.

4.2.3 Supplier Relationship Manager

SRM take responsibility for supporting and developing material supplier in the field of logistics by evaluating supplier delivery performance. SRM are also in charge of conducting Materials Management Operational Guidelines / Logistics Evaluation (MMOG/LE) audit. The purpose of this audit is to evaluate the logistics maturity of material supplier and initiate an action plan for identified gaps. This audit has three levels namely supplier self-assessment, desk verification of a self-assessment and on-site verification. Specifically, in the audit, there is a document of evaluating suppliers performance purely on logistics including lead time agreement, value, material handling, organization, production, communication, planning of all logistics. Material suppliers fill in the report and SRM have a site visit to evaluate these performances when necessary.

SRM are also managing low performing suppliers, if these suppliers performance are not improved for an agreed period of time, SRM should escalate them to supplier purchasing department and these material suppliers may end up losing contract from Volvo. Another task of SRM is prioritizing deliveries between Volvo manufacturing sites and CDC when there is crisis such as lack of capacity in material suppliers. Critical spare parts are among the first priority, and then the manufacturing sites get their capacity, finally, the non-critical spare parts get the rest of capacity.

4.2.4 Transport Material Coordinator

Similar to the responsibility of CMP on material suppliers, TMC is responsible for monitoring the performance of LSP in terms of agreed procedure and targets. For their appointed distribution flow including DDT and milk run, they are following up the performance indicators agreed upon with LSP while cross-docking transports are managed by another specialist.

If deviations happen, TMC also need to analyze the cause of deviations and take corrective actions within their responsibility area or propose corrective actions out of their responsible area. For example, if material suppliers cause the deviation, they should be escalated by TMC. If the deviation is caused by LSP, TMC could take corrective plan or escalate them to SM. This process is demonstrated below as Figure 4.4.

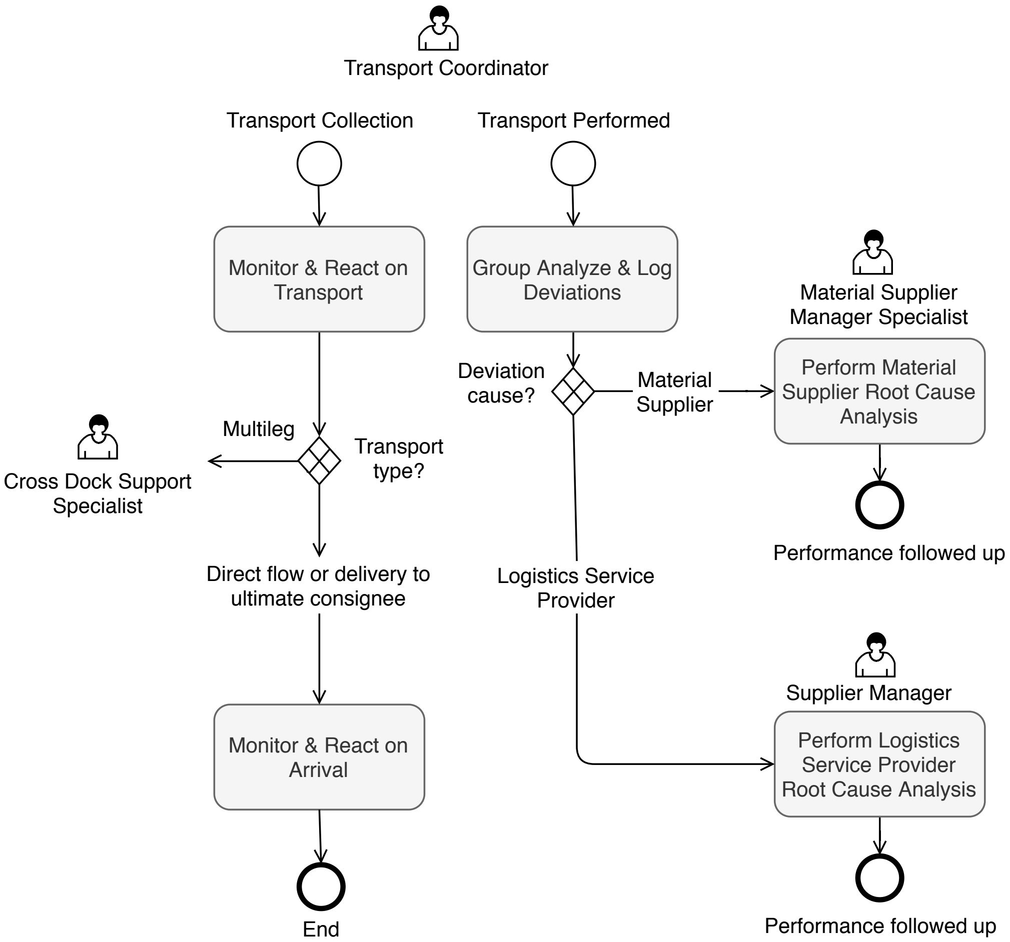

Figure 4.4: The working procedure of TMC

4.2.5 Supplier Manager

One of the responsibilities that SM have is the quality assurance for LSP. This means that SM have to make sure that every appointed LSP will deliver the agreed level of delivery performance based on their contract. There are some predefined targets related to the service levels for the LSP, such as pickup and delivery precision, their communication performance regarding reporting deviation in time. Following up these targets, making improvements and reporting them in terms of different weekly and monthly KPI are SM's tasks. It means that they follow up the performance of LSP in terms of delivery precision.

For those delivery deviations, SM are required to perform root cause analysis and take correction plan accordingly, in order to avoid or limit the consequence of deviation. For example, due to the dynamic character of the business environments, there would be disruptions such as harbour strike, storms, which would affect the planning. Efficient crisis management for them is a must to solve the problem in a short time and be sure that the planning schedule would not be affected too much. One of the solutions SM are using is to arrange meetings with LSP. The objective of these arrangements is to analyse the new situation and agree upon the standards and performance expectations based on new conditions in an open, straightforward and easily understood way to finally reach the target.

4.3 Situation of Deviation

Figure 4.5 shows the average SLT deviation of all spare parts for Volvo truck during the period of 2017 and 2018. The negative value represents the length of early dispatched orders in week (s) while the positive value represents the late ones. As the figure shows, there is one fluctuation in performance happened at the end of 2017, where large deviation occurred. The reason for this fluctuation is because this period corresponds to the Christmas break when the material suppliers cease production and operation. Otherwise, the delivery precision for truck spare parts has no seasonal trend.

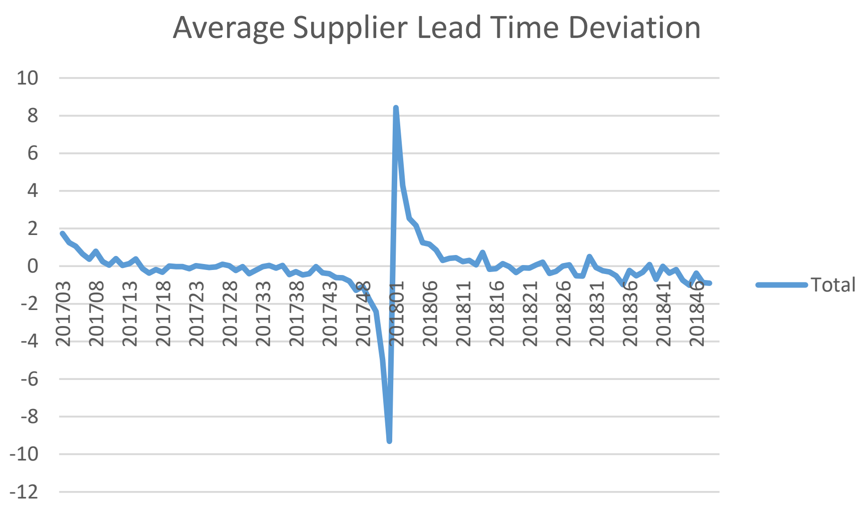

Figure 4.5: Average SLT deviation deviation for 2017-2018

The goal of delivery precision for material suppliers in Volvo is 95%, that contains all the dispatches not being late (including early and on time). Figure 4.6 shows that for the past two years, this actual delivery precision of not being late is 86%. Besides, among this 86%, up to 9% of the order dispatched earlier than scheduled. There is a significant gap between the goal and current deviation of both late and early delivery.

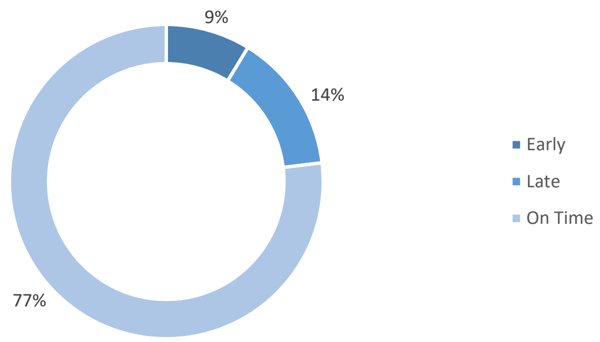

Figure 4.6: Delivery precision of material suppliers for 2017 and 2018

The goal of delivery precision for LSP in Volvo is 97%. However, for the transportation of the spare parts to Ghent CDC for past rolling one year, only 90% of them was not delivered late as Figure 4.7 shows. Further, 27% out of 90% actually delivered earlier than expected. The deviation of transportation is even larger than the previous delivery performance of material suppliers.

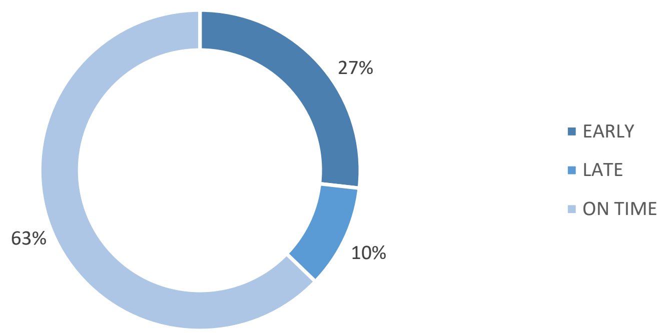

Figure 4.7: Delivery precision of LSP for past one year from 2019

4.4 Impacts of Lead Time Deviation

The deviation of lead time could bring various side effects and deteriorate the company's performance. These potential effects can be closely examined when the deviation occurs in material suppliers and LSP in terms of late and early delivery respectively.

When the spare parts cannot be dispatched on time according to the schedule from material suppliers, the immediate consequence could be the waste of transportation when LSP go to material suppliers based on TB information but end up failing to pick up the requested order. Even if the material suppliers communicate well about the delay information and change the new transport booking, the parts still arrive late at CDC Ghent. This could result in loss of availability when there is a demand for those parts, which means the company will fail to deliver what is requested due to lack of inventory. Likewise, the late delivery of LSP directly affects the stock level in CDC, and could further impact availability of stock possibly. This consequence could also cascade till the rest of the supply chain including the availability in regional distribution center (RDC) and dealers. Finally, it impacts customer satisfaction. In order to maintain the availability of spare parts, the cost is to adopt rush transportation such as air which is bringing in the high cost of transporting freight. The cost of rush air is huge for Volvo SML.

The spare parts could also be dispatched earlier than ordered from material suppliers. This is because on some occasion when they finish producing the orders earlier or have the stocks of the requested order. They could choose to book transportation in Atlas platform and push the parts to Volvo in order to get rid of their stock. In addition, the deviation in TTL could also result in less transportation time than estimated. These early arrived parts could also bring some problems. They could disturb the operation in CDC since the capacity of unloading and storage in a warehouse is limited. These early arrived parts could be waiting to be scheduled capacity and then unloaded at CDC. Furthermore, the inventory cost and tied up capital of keeping the parts for a longer time will also increase.

4.5 Business Goal of the Project

From the investigation, the lead time is a predefined parameter in the planning system, and it is set by the negotiation between the suppliers and Volvo. The lead time is very static in the system which is reviewed and updated relatively infrequently. There is a fairly high proportion of deviation existing in the performance of lead time. There are two roles in Volvo (CMP, TMC) who are directly responsible for monitoring the performance of suppliers regarding lead time deviation and taking action accordingly, which is achieved by close communication with each supplier.

However, to proactively communicate with all suppliers is time-consuming and less effective. To wait for information from suppliers about their deviation situation is not very reliable which depends on suppliers' proactiveness. Therefore, if Volvo can predict the lead time deviation in advance, it could be used as a deviation alert for the monitors. These monitors could selectively pay more attention to the suppliers that are predicted to have deviation. Communication can be more effective between monitors and suppliers to detect the deviation. This could help to trigger the actions to prevent the happening of deviation in advance. For other cases where the deviation is confirmed and unpreventable, the monitors can reschedule the inventory to correct the deviation. To sum up, the business goal for this project is to generate a deviation alert created by predicting lead time deviation of certain suppliers for certain orders. This alert could be used by CMP and TMC to be precautious and proactively contact the suppliers with deviation alert. This could improve the precision of ETA and ensure the availability at a low cost.

4.6 Factors Related to Lead Time Deviation and the Availability of Data

Gathered from the literature study, interviews as well as internal documents, a pile of factors that could be associated with lead time deviation in Volvo are compiled below, the availability of their corresponding data is also examined.

4.6.1 Factors of Material Suppliers' Lead Time Deviation

The sources of deviation can be categorized into two aspects, namely parts and suppliers. The deviation from parts is associated with their characteristics such as the complexity of producing spare parts, demand on the spare parts, and the criticality of spare parts. Most of the factors belong to correlation factors which could have associations with deviation but not directly result in deviation.

Deviation related to parts:

The characteristics of spare parts, including *the function groups, the life cycle position, the demand for spare parts, criticality of spare parts and value of spare parts.*

The *function groups* refer to the parts of the vehicles where the spare parts belong to. For example, the engine, fuel system, brake belong to different functions. This function group could reflect the complexity of the production. The production of the engine is more complicated than producing brake. The risks of suffering from deviation in production for the engine could be higher than those for the brake. The function group could also reflect its components of raw material. Since for automotive manufacturing, material suppliers rely heavily on their suppliers for providing the raw material. The supply situation of raw material could affect the production. For example, crisis of metal happens frequently than the plastics. There are 2882 function groups at Volvo Group.

The life cycle position is a changing statue in the life cycle starting from the introduction, to phase out of a certain truck model. It is determined by the number of years since introduction. The demand for spare parts corresponds to the life cycle phase of their related model. When more trucks sold, more related spare parts will be in need for that model and vice versa. For example, when a truck model is going through phase out, the stock for its spare parts is needed only for serving existing vehicles. There will be a decreasing demand gradually in the future. *The demand for spare parts* directly give the information on the amount of parts that have been ordered. However, the impacts of demand on deviation are uncertain. The higher order amount could bring in the economy of scale for production and draw more attention from production scheduling, therefore, decrease the risk of having deviation, whereas producing a large amount of parts could bring in risks of production disruption.

The criticality of spare parts. Different spare parts could have different criticality regarding their importance to secure up time for the vehicles. The higher criticality of spare parts derives from their higher importance to make the vehicle functions well. For example, a broken part which results in engine failure is more critical than a broken back mirror. Since the produced parts from one material supplier could be used for production sites, service market and powertrain in Volvo. When suppliers do not have the capacity for production from all demand, SRM do the site share based on the priority scale. The spare parts with high criticality are the first priority, and then the production parts get their capacity, finally the rest of the spare parts. Therefore, the non-critical parts could be more likely to suffer from deviations in delivery when there is limited capacity. There is a criterion named Vital code classifying the criticality of spare parts.

The value of spare parts could have a relationship with deviation. Since the expensive parts tend to have higher criticality and could be more important regarding their costs, it could receive more attention and enjoy higher priority for production. However, expensive parts tend to be more complicated regarding manufacture and have higher risk of production disruption. There is *standard price* recorded at Volvo for each part that corresponds to the value of spare parts.

In Volvo, there exists a measurement called segmentation that is classified by all above kinds of characteristics. There are 59 kinds of segmentation currently existing at Volvo.

Deviation related to Material suppliers :

There are several factors related to material suppliers. One factor is the *production disruption* occurred at the suppliers manufacturing site that directly causes deviation. There are also *supplier production capacity*, *supplier prioritization*, *supplier evaluation results* and *supplier historical delivery performance* correlated to the deviation of lead time.

Production disruption refers to the disturbances happened in the process of production that deviate the production process such as machine breakdown and labour shortage. In consequence, production disruption results in deviations of lead time. However, in the company, there is no data available or suitable to represent the production disruption happened in material suppliers. Meanwhile, *supplier production capacity*, that is referred to the maximum production volume that a supplier can handle at one time, could reflect the size and furthermore the ability of suppliers to handle production deviation. Suppliers with large capacity tend to deal with production disturbance smoothly by scheduling resources to bottlenecks and then dispatch orders properly. There is no direct information available in Volvo for *supplier production capacity.* However, it could correlate to *sales level spend, order hits, book off quantity* for material suppliers. Since the more money the company spends on its suppliers and the larger volume the company orders from them, those suppliers are more likely to be larger firms with a larger capacity. *Quality certificates* and *environment certificates* such as certificate ISO 14000 and QS4000 could also reflect different production standards and relate to the ability to handle production disruption.

Supplier prioritization. A material supplier normally supplies to various customers, and therefore the production capacity of material suppliers is going to assign to different buying companies. In case they are overbooked for the production, the order from the buying company may be delayed due to the suppliers' prioritization for other buyers' order when that company is not their prioritized customers. Volvo does not have direct information on how prioritized they are as a buyer. Nevertheless, the size of business between Volvo and the material supplier could closely relate to the prioritization, since suppliers tend to have a closer relationship with buyers with large orders and set these buyers with a higher priority. As Volvo has information about *sales level spend, order hits, book off quantity*, these could be indications of the size of business and thereby the closeness of the business relationship.

Supplier evaluation result. From a buying company's perspective, some performance of suppliers can be perceived and recorded, and this information on suppliers could be evaluated. The knowledge is obtained in order to evaluate and develop supplier performance and make the decision for further cooperation. Since the evaluation information could be closely linked to the suppliers' delivery performance, this evaluation information could be used to predict future deviation. In Volvo, SRM perform supplier evaluation and generate SEM results. This result evaluates the overall performance of a supplier by examining various aspects including company profile, management, environment, quality, logistics, aftermarket, competence, product development, finance, productivity, and sourcing. The SEM results are consolidating all the performance and generating one score with the scale between 0-100 for each material supplier in a certain period. In addition, SRM perform logistic audit specifically. This is an audit of evaluating suppliers' performance purely on logistics including lead time agreement, value, material handling, organization, production, communication, planning information of all logistics aspects. The logistic audit result could also be reflected by another score consolidation the results from these criteria. These evaluation results could be a good indication of suppliers' ability to deliver on time.

Supplier historical delivery precision could be very informative in terms of predicting the future performance of a supplier and lots of traditional prediction methods are purely based on historical information. At Volvo, delivery precision is a key performance indicator for material suppliers. It is the percentage of the number of parts dispatched on time divided by the total number of dispatched parts. The result indicates the percentage of a material supplier fulfilling orders with the right quality at the right time with the right paperwork attached.

Table 4.2: Factors related to lead time deviation in Volvo

Phase	Sources	Factors
Factors of Material Suppliers	Parts	Characteristics (Function groups)
		Life cycle phase
		Demand
		Criticality
		Value
	Material suppliers	Production disruption
		Supplier prioritization
		Supplier evaluation results
		Supplier historical delivery precision
Factors of Inbound Transportation Lead Time	Parts	Weight, Volume
		Stackable
		Hazards
		Custom
		Demand
		Value
	LSP	Evaluation results
		LSP historical delivery precision
	Supply chain	Material supplier delivery precision
		Transport scheduling
		Warehouse scheduling
		Country
		Traffic and weather

4.6.2 Factors of Inbound Transportation Lead Time

For the deviation in inbound TLT, the factors can also come from three aspects. The first one is the parts considering that transportation is sensitive to its carried freight. It could include the logistics characteristics of the parts. The demand for parts could affect transportation scheduling and further the risk of delay. These factors all belong to correlation factors.

Deviation related to parts

Logistics characteristics of parts refer to the characteristics of parts that could influence the transportation including the weight, volume, stackable, hazards, the requirement of custom. The weight, volume and the stackable could affect the scheduling of transportation. For example, non-stackable parts and high weight or volume parts may need to book more than one truck to carry all ordered parts from material suppliers, and therefore these separated shipments increase the risks of deviation for all of them arriving on time. The hazardous parts or parts required to clear custom result in more process during transportation, and these extra processes may endure the fluctuation of processing time along with the delivery.

Demand for parts refers to the number of parts that requires to be shipped by LSP from material suppliers. Different level of demand directly determines the way of transportation. The high demand of a part could achieve the full truck load (FTL) from one supplier, and simply adopt DDT transportation. That is different from when there is less demand for a part. It brings in the order from a material supplier is limited truck load (LTL) which has to go through cross-docking or milk run in order to fully utilize the capacity and realize the cost benefits of shipments. Waiting time at the cross-docking point or material suppliers throughout the milk run is a high-risk factor of lead time deviation. This information of transportation solution is available in Volvo. Besides, the receiving quantity of one spare part and the accumulated number of all parts in one shipment are available.

The *value* for one shipment could associate with the on-time delivery positively. Since the higher value for one shipment, the more attention and priority it receives. This attention and priority could help the shipment being processed earlier and get rid of the extra waiting time. This record of value for one shipment is available at Volvo.

Deviation related to LSPs

The second aspect belongs to LSP. *LSP evaluation* and *historical delivery performance* could be associated with the deviation of TLT.

Similar to material supplier evaluation, *LSP evaluation result* is obtained in order to evaluate the performance of LSP and this information can be used to predict the delivery performance of LSP. At Volvo, SM perform LSP evaluation and generate a final score for each LSP. The content of evaluation is including pickup and delivery precision, administration, deviation reporting in real time and communication. However, not like SEM and logistics audit results for material suppliers which are logged into the database, the scores of LSP are scattered in each evaluation report of LSP and not integrated into databases. Therefore, this information is not likely to be considered into the prediction model. The same situation exists in the LSP historical delivery performance records. Lack of information also brings the difficulty of estimating the ability of LSP handling uncontrollable disruptions of environment and society such as labour shortage and storms.

Deviation related to supply chain

The third aspect is the supply chain including partners' performance and the traffic and weather information alongside the route.

Delivery precision of material suppliers could relate to the deviation of lead time. Material supplier could order transportation booking earlier than the schedule to get rid of the finished stock and cause the early arrival of orders. On the contrary, waiting at material suppliers happens when LSP arrive at material suppliers but suppliers are not ready to dispatch the orders. This situation could be prevented if the material supplier communicate and update the delay information proactively. If not, the delay at material suppliers is a key factor causing a delay in transportation. The transportation in the company currently does not link with the corresponding delivery precision from previous material supplier. Extra time is also generated from missing documents from material suppliers such as proof of collection. The departure time also determines the arrival time at CDC, and the arrival time beyond the operational hours at CDC will need to wait overnight to be unloaded. However, the departure time is not accurately recorded in the company's system.

Likewise, *waiting at warehouse* results from lack of capacity in the warehouse to receive and unload the freight. This waiting time all relate to the scheduling issues since other actors in the supply chain share the resource of the warehouse. The capacity of the warehouse, however, is not integrated into the database either and link with previous transportation at Volvo.

The *country* of material suppliers could determine the condition of transportation by having different roads quality, geology. Besides, the political and economic situation differs from country to country. *Traffic and weather information* all the way to the destination of CDC also directly affect the transportation lead time. The country information is available, while the weather and traffic information are not existing at Volvo.

5

Results: Data Understanding and Preparation

In this chapter, a data mining goal is first generated based on previous business goal. Then the result of features selection is presented based on the result of relevant factors of deviation.

5.1 Data Mining Goal

To realize this business goal, we need to set up the data mining goal accordingly. To choose between regression and classification model, the regression generate continuous value for lead time deviation, while the classification could give the outcome of three classes, namely early, on time, and late. In terms of predictive capability, the classification model predicts whether there will be deviation while the regression model can give more information on deviation including how much deviation there will be. However, it could be more difficult to have a reliable result to be the numerical values, considering the distribution of lead time deviation with the majority of the case being on time which corresponds to the deviation to be 0. This high portion of 0 could distort the result of regression, since it is difficult to learn from not enough instances with different distribution of days of the deviation. In comparison, classifying the output could accumulate a lot of instances to learn for each class. The goal of data mining is thereby to generate two machine learning models for predicting deviation in material supplier lead time and truck arrival time respectively by testing various classification machine learning algorithms and evaluate their performance.

5.2 Features Selection

To represent each factor related to lead time deviation, a list of relevant and available data is collected and examined for the selection of features. There are three cases occurred in this process. Firstly, there are data that can directly represent the factors such as using order hits to represent demand. Secondly, some factors do not have data to directly represent them or the corresponding data are not available in the company, but there are some data may represent these factors indirectly. For example, the material supplier's prioritization for Volvo could be represented by Volvo's sales level spend on that supplier. However, there are some factors that could not either find suitable data for indirect representation, such as historical delivery precision performance and evaluation results of LSP. They are currently scattered in different excel sheet for each LSP and not logged into the database. The relationship between the factors and features is illustrated at Table 5.1. After manually linking these two phases of transportation and material supply, previous deviation of material suppliers is available, and the delivery performance from material suppliers could affect the success of pick up for LSP.

The description and characteristics of features for SLT and TLT are presented in Table 5.2 and Table 5.3 separately. Among data type, the number within parentheses represents the dimension of each categorical variable.

Table 5.1: Features selection

Phase	Sources	Factors	Features
Factors of Material Suppliers	Parts	Function groups	Segmentation (adapted))
		Life cycle phase	Segmentation (adapted)
		Demand	Order hits, Book off quantity
		Criticality	Vital code
		Value	Standard Price
	Material suppliers	Production capacity	Sales level spend, Quality and environment certificates, Regions
		Production disruption	\
		Supplier prioritization	Sales level spend
		Supplier evaluation results	SEM results, Logistics audit results
		Supplier historical delivery precision performance	Delivery precision
Factors of Inbound Transportation Lead Time	Parts	Weight, Volume	TB Actual Weight, TB Actual Volume
		Stackable	Stackable
		Hazards	Hazardous code
		Custom	Custom
		Demand	Receiving quantity, TB units
		Value	TB Value
	LSP	Evaluation results (Company performance & logistics maturity)	\
		LSP historical delivery precision performance	\
	Supply chain	Material supplier dispatch performance	Previous deviation
		Transport scheduling	Delivery method, Transport load, POC required, POD required
		Warehouse scheduling	\
		Traffic and weather	\
		Country	Country

Table 5.2: Current available features for SLT model

Source	Features	Description	Data type
Spare Part	Segmentation (adapted)	Parts segmentation of parts in terms of function groups and life cycle, ‘slow A’, ‘fast A’, ‘slow B’, ‘fast B’......	Categorical variable (11)
	Vital Code	The importance of the parts for 1- Prioritized parts 2- Service Parts 3- Consumption Parts 4- Non-vital parts	Categorical variable (4)
	Prepack	The requirement whether a part needs to be prepacked before transporting or not	Binary variable
	Standard price	Standard price for spare parts	Continuous variable
	Book off Quantity	The amount of part ordered	Continuous variable
	Order hits	The historical order frequency for spare parts	Continuous variable
Material Supplier	SEM result	Supplier evaluation model score, measurement including company profile, management, environment, quality, logistics, aftermarket, competence, product development, finance, productivity, sourcing	Continuous variable
	Logistics audit Results	Evaluating suppliers’ performance purely on logistics including lead time agreement, production, planning information and etc. of all logistics aspects.	Continuous variable
	Sales level Spend	The amount of money from Volvo spends on supplier	Continuous variable
	Delivery precision	Supplier historical performance regarding the percentage of parts delivered on time	Continuous variable
	Regions	The region where suppliers are located, including ‘EMEA’, ’APAC’, ’Americas’	Categorical variable (3)
	Purchase agreement	Whether there is an agreement with material suppliers including confidentiality agreement, development agreement, price agreement, warranty charter etc.	Binary variable
	PPM	Defective parts per million	Continuous variable
	QPM	A percentage calculated from PPM for production quality	Continuous variable
	Environment Certificate	Production standard measurement, including ‘IATF16949’, ’ISO17025’, ’ISO9000’, ’ISO9001’, ’ISO9002’	Categorical variable (5)
	Quality Certificate	Quality standard measurement, ’QS9000’, ’VDA6’	Binary variable

Table 5.3: Current available features of truck arrival time model

Source	Features	Description	Data type
Spare Part	TB Actual Units	The unit of parts in a transport booking	Continuous variable
	TB Actual Weight	Transport handling weight in a transport booking	Continuous variable
	TB Actual Volume	Transport handling volume in a transport booking	Continuous variable
	Value	Value for parts in a transport booking	Continuous variable
	Received Quantity	Quantity received at CDC for one spare part	Continuous variable
	Stackable	Whether the part is stackable or not	Binary variable
	Custom	Whether the part needs to clear custom or not	Binary variable
	Hazardous code	The category of a part being hazardous	Categorical variable (3)
Logistics Service Provider	Proof of delivery required	Document required for delivery proof for parcel deliveries	Binary variable
Supply chain	Delivery method	The way of delivery including DDT, Milk run, and Cross-docking	Categorical variable (3)
	Truck load	Truck load information including Full truck load and Limited truck load, dynamic planing information of milk run	Categorical variable (4)
	Previous deviation	delivery deviation of material suppliers, 'Early', 'Late', 'On Time'	Categorical variable (3)
	Country	The country where the material suppliers located	Categorical variable (11)

6

Results: Models and Evaluation

In this chapter, based on the previous result of available features in the company that are associated with the lead time, two prediction models of lead time deviations are generated using various machine learning techniques. The results are presented by classification report which include precision, recall, f1 score. The feature importance for the most relevant features is also introduced

6.1 Classification Report

The confusion matrix result of deviation prediction model in supplier lead time and truck arrival time are represented below in table 6.1 and table 6.2 respectively. The precision score represents the accuracy of prediction. For example, with random forest model for SLT, it predicts 25,826 (=13626+1824+10376) observation to be 'Late' while 10376 of them actually arrived late. The precision will be 0.4 out of 1 (=10376/25826). The recall score represents the missing of capturing the occurrence of a class. For example, with catboost model for truck arrival time, it correctly predicts 111 observation to be 'early' while there is 176 (=62+111+3) cases in fact being early delivery. Therefore, the recall score is calculated as 0.63 out of 1 (=111/176), which means 37% of early delivery is not predicted by the model to be 'early'. The higher the recall score, the lower the number of missing capture.

Table 6.1: Confusion matrix for SLT models (columns being predicted classes and rows being actual classes)

	catboost			Gradient Boosting			Random Forest		
	On Time	Early	Late	On Time	Early	Late	On Time	Early	Late
On Time	58617	16813	17006	90100	531	1805	67560	11250	13626
Early	2202	6377	1947	8012	2010	504	3012	5690	1824
Late	3333	2904	10994	13002	527	3702	4541	2314	10376

Table 6.2: Confusion matrix for truck arrival time models (columns being predicted classes and rows being actual classes)

	Catboost			Gradient Boosting			Random Forest		
	On Time	Early	Late	On Time	Early	Late	On Time	Early	Late
On Time	1408	115	135	2391	50	37	1302	182	174
Early	62	111	3	165	105	2	39	130	7
Late	131	13	109	256	4	120	108	20	125

As the classification report shows in table 6.3 and table 6.4 respectively, for the SLT model, catboost outperforms the other two methods in terms of recall scores for deviation prediction (0.61 for ‘early’ and 0.64 for ‘late’). The score is slightly higher compared to random forest. While random forest overpasses catboost in F1 score where both precision and recall are taken into consideration. Focusing on the deviation class, the F1 scores in early and late are less than 0.5, far from deploy level which should be better at least above 0.8.

For the truck arrival time model, random forest has the highest score of recall (0.74, for ‘early’ and 0.49 for ‘late’) while performing not that well in precision compared to catboost. Similar to the first model, the average scores of each algorithm for the deviation class are less than 0.5.

Table 6.3: Classification report for SLT models

	Catboost			Gradient Boosting			Random Forest		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
On Time	0.91	0.63	0.75	0.81	0.97	0.89	0.90	0.73	0.81
Early	0.24	0.61	0.35	0.66	0.19	0.30	0.30	0.54	0.38
Late	0.37	0.64	0.47	0.62	0.21	0.32	0.40	0.60	0.48
Total	0.78	0.63	0.67	0.77	0.8	0.75	0.78	0.70	0.72

Table 6.4: Classification report for truck arrival time models

	catboost			Gradient Boosting			Random Forest		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
On Time	0.88	0.85	0.86	0.85	0.96	0.90	0.90	0.79	0.84
Early	0.46	0.63	0.53	0.66	0.39	0.49	0.39	0.74	0.51
Late	0.44	0.43	0.44	0.75	0.32	0.45	0.41	0.49	0.45
Total	0.79	0.78	0.78	0.82	0.84	0.81	0.80	0.75	0.76

6.2 Feature Importance

Feature importance gives insightful information on the relevance and the relative importance of features for the models. The most important features feature importance calculated by the two better performed algorithms in each phase is presented below in Table 6.5.

Table 6.5: Feature importance in two models in each phase

	Random Forest	**Catboost**
Supplier lead time deviation	1. Standard price	1. Delivery precision
	2. Delivery precision	2. Standard Price
	3. Order hits	3. Order Hits
	4. SEM result	4. Sales Level Spend
	5. Sales level spend	5. SEM Result
	6. Logistics result	6. Logistics Audit Result
	7. QPM score	7. QPM Score
Transportation lead time deviation	1. France	1. Received Quantity
	2. TB Actual Weight	2. TB Actual Weight
	3. LTL_DP	3. Value
	4. Received Quantity	4. TB Actual Volume
	5. TB Actual Volume	5. TB Actual Unitt
	6. Sweden	6. FTL
	7. Germany	7. Sweden

7
Discussion

In this chapter, the results of this thesis are interpreted and discussed. Firstly, the implication of results on literature is discussed by comparing with current literature. Secondly, the implication of the result on the case company is discussed. Finally, the underlying assumption and limitation are presented.

7.1 Implication on literature

This thesis project works on two areas in the theory. The first one is predicting material suppliers' delivery precision performance. The second one is predicting the deviation of truck arrival time.

7.1.1 Predicting Material Supplier Delivery Precision

Although a company's performance is much affected by its suppliers' performance including delivery precision (Krause et al., 2007), very few literature has investigated the evaluation and prediction of supplier performance in operational level during the period of cooperation. This may due to the complex relationship between the suppliers' performance and several criteria of suppliers (Rezaei et al., 2014). However, with powerful analysis tools such as machine learning and a large number of instances, this complex relationship could be examined. For example, Khaldi et al. (2017) and Jiang et al. (2013) implement machine learning algorithms to evaluate and predict suppliers' overall performance based on their historical performance data in several aspects such as delivery, costs, quality.

This thesis project specifically focuses on predicting supplier delivery precision performance using machine learning. The input features that have been used to train the prediction model are including not only the supplier historical performance and evaluation information but also the information of parts ordered from that supplier. Since different parts have different characteristics which could relate to the difficulties of production, and therefore these characteristics further relate to the deviation in production.

When we examine the result of modelling for SLT deviation, however, the prediction power is not enough with poor precision and recall score of models on ‘Late’ and ‘Early’ classes where the deviations locate. That means with current features of parts and supplier information, machine learning models still cannot well capture the relationship between the occurrence of deviation and these features. That may due to in majority case, the deviation could result from the production disruption such as machine break down, labour shortage, and waiting time in the production line. This information currently is only owned by material suppliers themselves and not able to be utilized by a buyer company.

7.1.2 Predicting Deviation of Truck Arrival Time

The deviation of transportation is most likely to be subjected to the weather and traffic situation. The successful implementation of machine learning on predicting the deviation of the airplane could purely based on the weather information at departure and arrival airport (Belcastro et al., 2016). Whereas, predicting the truck arrival time based on only weather and traffic information does not achieve a good result (van der Spoel et al., 2015). When considering the full network state including physical characteristics of the train and train crew information for predicting the arrival time of freight train, machine learning brings large improvement in prediction (Barbour et al., 2015). To interpret these differences, we can consider and compare the causes of deviation in each transportation mode. For air transportation, one of the major causes of the deviation comes from the weather (Belcastro et al., 2016), while the freight train is less prone to be impacted by the weather but more likely to be affected by the scheduling of train network. Therefore, the common thing in these two successful implementations is that they manage to make the causes of deviation into features for the prediction model. In contrast, for the arrival time of the truck, it is affected by not only weather and traffic but also could be largely affected by factors such as the scheduling information from the consignor and consignee of truck transportation. Only providing weather and traffic information is not going to make machine learning model capture the pattern of deviation.

This thesis considers that lead time deviation of the truck is related to the scheduling of transportation and prone to the situation from both consignor and consignee, that is, for example, the deviation of delivery from the consignors could affect the pickup precision for LSP and further impact the delivery precision to the consignees. However, the deviation from consignor used as a feature in the model does not get a high feature importance score as expected, this may party due to that the deviation from is only structured as a classification feature with three dimension of ‘Late’, ‘Early’ and ‘On time’ instead of the exact number of deviation.

This thesis also innovatively considers the logistics characteristics of the transported cargo and further the size of shipment including weight, volume, units. It turns out some of them highly contribute to the performance of the prediction models. The scheduling method of transportation such as the truckload and delivery method also contributes to the prediction model.

Although we consider some features of organization and cargo into prediction, and some of them are much likely to be associated with deviation as feature importance indicates. This proves that the organizational and cargo features are relevant. However, the model performance is still limited due to the unavailability of some important features in the company. A successful prediction model of trucks' ETA may not stand alone without considering traffic and weather information.

7.2 Implication on the Case Company

Volvo strives to maintain a high delivery precision from its suppliers. In terms of the KPI delivery precision, the early dispatches and deliveries should not be regarded as fulfiling the delivery precision since they are also harmful according to our investigation.

This thesis mainly investigates predicting the deviation precision from their material suppliers and LSP. For these two phases, Volvo has more power in inbound transportation than material supply. Since Volvo has a platform for scheduling and coordinating all the transportation which could make all the transportation information available, while Volvo currently does not have production information of their material suppliers. Therefore, while in this project the two phases share the same goal which is to aid the monitoring process by generating deviation alerts, the results from these two phases will not have the equal implication on the case company. In this section, the implication of two prediction models on the company's monitoring is discussed separately.

7.2.1 Monitoring on Material Suppliers

While in this phase, the limitation on the modelling is related to the feature selection for production information in material suppliers. Considering improving the prediction model performance, information sharing with material suppliers regarding their production disruption is necessary in the future.

The models still achieve some prediction power for deviation of SLT. Therefore, the feature importance generated for the models may deserve an examination for their close relationship with deviation (Trevor et al., 2009). For example, for selected two models' most important features (Catboost and Random Forest), they share all top seven important features with a slightly different ranking. It is most likely that the features of evaluation results from suppliers including delivery precision, SEM result, logistics audit result, sales level spend are negatively related to the deviation of lead time. That is the lower the performance of these indicators, the more the deviation there tends to be.

To examine the influence of characteristics of parts on deviation, for example, the standard price of parts, as the below Figure 7.1 shows, 44% of orders with the spare parts valued more than 10,000 has deviation while this figure is only 21% for spare parts with price lower than 1,000. This indicates the higher standard price of the spare parts, the more likely the deviation of SLT could happen. That may due to the higher the price, the more complicated of the parts to produce which bring in the risks of deviation. Therefore, the CMP can pay more attention on the spare part order with expensive price.

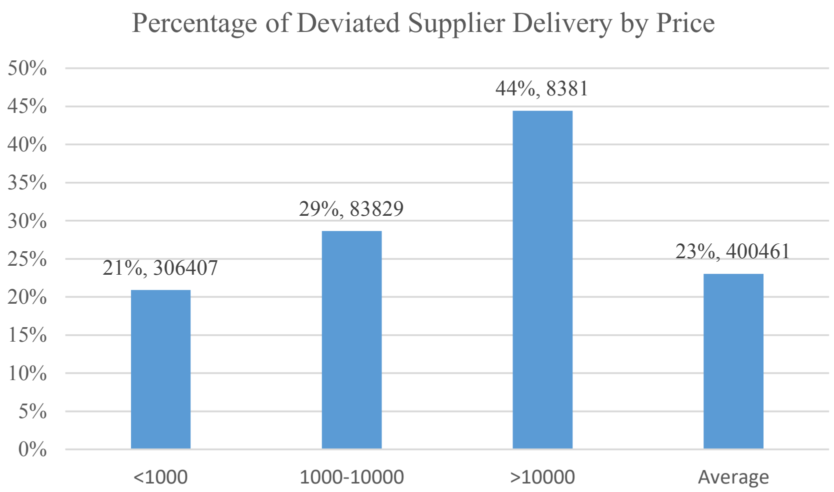

Figure 7.1: Percentage of deviated supplier delivery by price

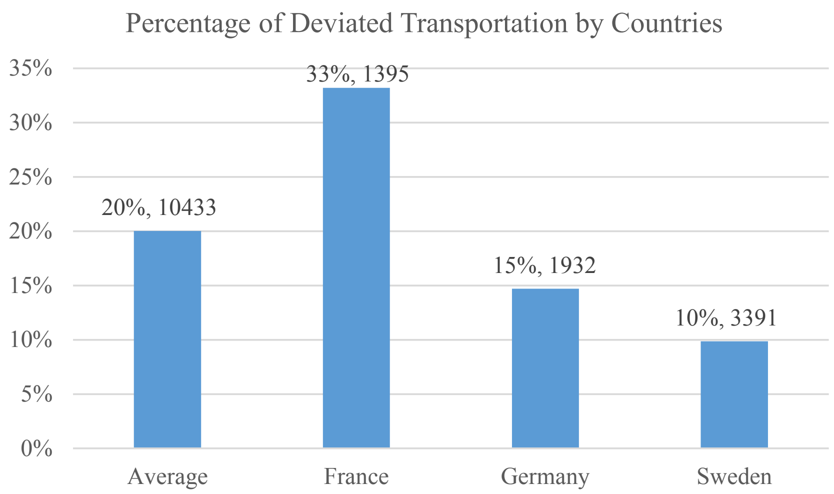

Figure 7.2: Percentage of deviated transportation by countries

7.2.2 Monitoring Process on LSP

Similar to the monitoring on material suppliers, in the short term, LSP could examine the relationship between the most important features and lead time deviation. For example, we take a close look at the deviation by countries indicated by the high feature importance of 'France', 'Sweden', 'Germany'. As Figure 7.2 shows, compared to average deviation case of being 20%, 33% of transportation from 'France' are deviated. Therefore, LSP can give special attention on transportation from France to

CDC. On the contrary, the transportation from Germany and Sweden to CDC are performing way better, especially for Sweden, only 10% of truck transport has the deviation which is 10% less than the average.

In the long term, a dynamic deviation alert could be embedded in the system to assist the TMC to monitor the performance of LSP. This alert can help TMC preventively reach LSP to figure out whether there will be a deviation according to the prediction. This alert could require LSP to examine their operation statues. TMC could also examine the delivery precision statues from their consignors and the capacity situation at the consignee warehouse. If there could be a deviation, TMC can help to take actions to prevent the deviation. If the deviation is irresistible, such as the extreme weather, some corrective actions could be scheduled to alleviate the influence brought by the deviation.

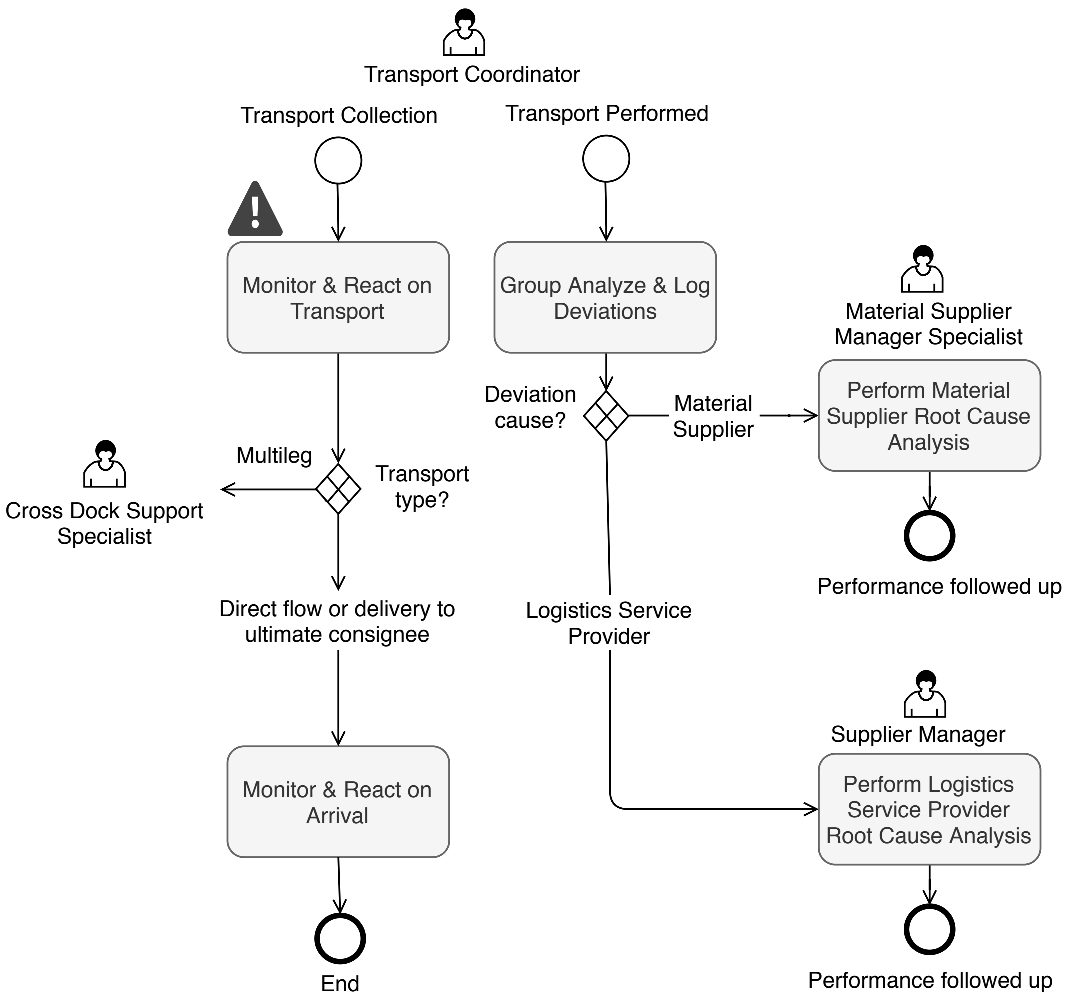

Figure 7.3: Generating a deviation alert in the process of monitoring LSP

In order to construct a reliable deviation alert, some improvement needs to be achieved. Since the alert is generated by the machine learning prediction models, the performance of model on 'Late' and 'Early' classes is important to improve. The precision score affect the accuracy of the model and therefore affect the reliability of the alert generated, while the recall score reflects the degree of miss capture of these two classes which means the miss capture of the deviation (Sokolova and Lapalme, 2009). The recall score could be even more important considering Volvo's availability performance. In order to improve the performance of the prediction model, the linkage of the database between the material supplier phase, inbound transportation phase and internal warehouse phase should be constructed. It means transportation booking information should contain the ordered part information. The optimal result of constructing the database is that when we input one transportation booking ID, all the part number in this shipment will present with all needed feature information regarding the parts linked, such as their standard price and order hits. Besides, all the features needed in the supply chain including the material suppliers (consignor), LSP, warehouse (consignee) should also be linked in one click. Furthermore, in order to generate a dynamic and reliable alert, the open source weather and traffic forecast information should also be added in the database. This result of this new dataset considering all the linkage is shown in Figure 7.4.

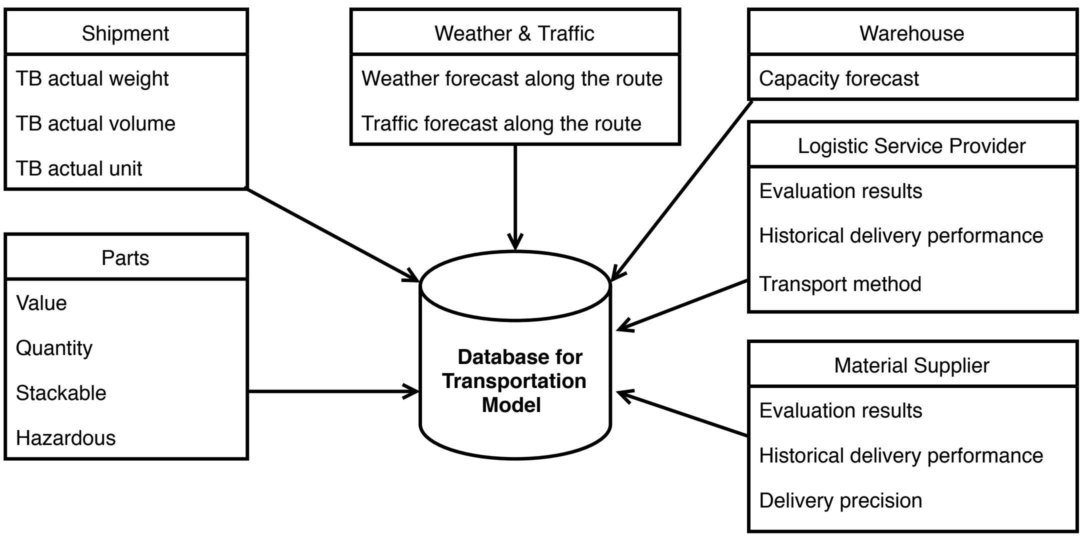

Figure 7.4: Recommendation for linkage of prediction model for truck arrival time deviation

7.3 Underlying Assumptions

The data from the company database we used are assumed to be accurate and reliable. Regarding SLT analysis, we use the same source of data with the logistics analysts in the company. For material suppliers' information, we extract data from the supplier management portal as the SRM are using the same portal. Regarding truck arrival time, the logistics service portal of Volvo is the place where transportation data are compiled so it is regarded as a reliable source of accurate data. The false of this assumption is inaccurate data, which could affect the performance of the prediction model. However, the effect could be subtle since the missing features could be the main bottleneck to improve model performance.

Another assumption is that the sampled data are representative. For SLT deviation performance, we extract data of two years data from 2017 and 2018. For truck arrival time, the past rolling one year data is extracted. These data are regarded to be representative for the whole population. It is assumed the lead time data from each year is homogeneous and the result can be generalized to the future period. That means the patterns between the selected features and lead time deviation remain stable for each year. Even though this relationship could be evolved as years pass, those features with high importance score are most likely to be important, only the degree of importance could be reexamined each time when building the model.

This thesis also assumes that for SLT, those causal factors of production disruption which directly result in deviation cannot be assessed by a buying company. Furthermore, we assume the data the company have can indirectly represent those causal factors to some degree. However, according to the result of the models generated, the poor prediction performance indicates that this assumption of the above representativeness could be false. The causal factors of production could be not well represented by the substitute information available on the buying company. This could lead to a conclusion that, when there is limited access to production information, the deviation of SLT could not be well predicted with the machine learning algorithms from the buying company's perspective.

7.4 Limitation

The first limitation comes from the scope of this thesis, with only focus on one business area and one geographical region, this could limit the size of data and the complexity of features. The performance of the prediction model is most likely to be affected by the amount of data and dimensions of features. Besides, since the characteristics of outbound logistics could be very different from the inbound logistics, with only lead times in inbound transportation investigated, this thesis project cannot be directly generalized to the outbound phase without adjustment. This thesis also only use the database from the case company, that result in the traffic and weather information is roughly represented by the origin country. This information could be accessed from an open source database and considered in the prediction model for the future.

The feature selection and data preprocessing also have some limitation. There is a small portion of 'outliers' in historical deviation performance such as late or early up to 4 years. The reason for these strange numbers is not examined. However, since the proportion of these outliers and the percentage of deviation classes are both small, and the outputs of these instances are very likely to be late or early. Therefore, they are kept in the dataset belonging to either 'Late' or 'Early' classes to contribute to the machine learning of minority classes. The influence of this way of handling potential outliers could be small. Furthermore, for handling missing values in the features, this mainly exist in the evaluation results of material suppliers. Only average values are used to replace them, which may result in a poor estimation of variances and correlations for the feature (Schafer and Graham, 2002). Considering the pattern of the missing value is unclear to us for now, therefore advanced imputation methods remain to investigate. However, as we discuss before, the missing of key features would be the main reason for the poor performance of deviation prediction model. The exact representation of missing value in those evaluation features would not improve the prediction model to a large extent.

There is limitation existing in the modelling process. This thesis is only implementing classification models rather than regression models. Considering even though the regression model could generate the prediction of the exact time of deviation, this exactness also increases the difficulty of regression models to achieve a better and reliable result compared to classification models at this very first stage. However, the construction of classification models could also be improved by further increasing the granularity of the response variables into more classes such as 'Very Late', 'Very Early' to further increase the informativeness of classification models. This improvement is not tested due to the time limit of this data mining project.

Another limitation in modelling is the lack of examination of the feasibility of algorithms. Since the results of modeling with tested algorithms are not good enough and closed to each other, the performance is believed to highly relate to the features. The algorithms used in modelling are regarded as feasible and optimal from the knowledge of literature review and expertise of data scientists based on the characteristics of input features and output. For future model improvement, however, algorithms should be re-evaluated for their feasibility and optimisation for new modelling after the improvement of features.

The SLT modelling also does not take time series into account, since the evaluation information from the buying company towards their material suppliers last for a long period. For the two years period, even though considering the event time for each order, the variance of features could be very low for each instance of the same spare parts. Therefore, the result is not likely to be influenced by the time factors.

8 Conclusion

In this chapter, the research finding is summarized by answering four research questions, followed by the practical and theoretical contribution of this thesis project. Finally, the recommendations and future researches for the case company and academia are presented in the end.

8.1 Research Finding

The research questions are covered through analysis and discussion in this thesis work, therefore these answers are stated below as a summary of the research finding.

RQ 1: What are the benefits of predicting lead time deviation for buying companies?

Certain roles in buying companies are responsible for monitoring the delivery precision of suppliers and taking actions to deal with deviations. For example at Volvo SML, continental material planners have the responsibility for monitoring the lead times for the suppliers. Likewise, transport material planners monitor the lead times performance of LSP. The prediction of lead time deviation can help create deviation alerts that assist those monitors for monitoring suppliers' delivery performance, and the benefits of the alerts are to reduce the deviation and decrease the impact of deviation by taking preventive and corrective actions.

RQ 2: What are the factors that could be associated with lead time deviation perceived by buying companies?

The factors related to the deviation of lead time from the buying company's perspectives can be categorized into three levels. The first one is the part level regarding the characteristics of parts in demand, criticality, value, life cycles and function groups for material supplier lead time and the part level regarding the logistic characteristics including volume, weight, hazardous, custom, stackable, demand and value. The second level regarding the supplier level, it represents the evaluation results and historical for material suppliers and LSP for supplier lead time and truck arrival time respectively. These evaluation performances are covered in multi-criteria aspects of the suppliers with special focus on delivery performance. This level may also include the priority of the buying company within the material suppliers. The third level of factors could be exclusively existing for transportation lead time which is related to the actors in the supply chain level including their consignors, consignees and buying company. The deviation of truck arrival time is affected by the delivery performance of material suppliers which affects pick-up precision of LSP. The route of shipments scheduled by the buying company could affect lead time. Besides, unloading shipment is also subjected to the capacity of warehouses. The country and environment including weather and traffic are most like to be relevant.

RQ 3: Which data are available to be used as features when building the prediction model of lead time deviation at Volvo SML?

To turn the above factors into features for modelling, there are a few cases occurred in this thesis project. There are data which can directly represent the factors such as the demand, value, stackable, hazardous, custom, evaluation results for material suppliers. There are data representing the factors at an aggregated level, such as TB weight and volume data for the total weight and volume in one shipment, segmentation data for integrating function groups and life cycles, country for traffic and weather. Some factors that are not recorded in the data form, such as the prioritization, some factors are not available in the buying company due to that information is owned by material suppliers such as material suppliers or not integrated into the databases. These factors are tried to be indirectly reflected by other available data, such as sales spend level data on suppliers for the prioritization, quality and environment certificate for the production capacity of suppliers. However, some factors could not either find suitable data for indirect representation, such as historical delivery precision performance and evaluation results of LSP.

RQ 4: How should the prediction model be built using machine learning considering the practicality of use in the current stage at Volvo SML?

Through the business analysis and data analysis, the goal for a prediction model at current stage could be generating a deviation alert for monitors in the buying companies. A classification model with the output of three classes ‘On Time’, ‘Late’, ‘Early’ could achieve this goal. However, modelling with currently available features for both two phases do not deliver deployable results. In order to improve the results of modelling, more representative features should be added for capturing the pattern of deviation. For predicting truck arrival time at CDC Ghent, since Volvo has more power in this phase, most of the key features can be filled in and improved with Volvo’s efforts. The prediction model for truck arrival could be improved and put into use in the future when databases are constructed as expected as figure 7.4 shows. However, for the SLT model, the key missing features could be the production disruption information. Predicting material supplier lead time with machine learning from the buying company perspective is, therefore, regarded as not practical until production information could be shared with the buying company.

8.2 Practical Contribution and Future Work for the Case Company

This thesis for the first time demonstrates the possibility for predicting the deviation of lead time for Volvo truck spare parts including supplier lead time and truck arrival time at CDC Ghent with machine learning. The prediction model on the deviation of truck arrival time shows its potential deployment after the future improvement of data linkage for input features, while the model of supplier lead time deviation is believed to be improved only when some key features including the production information are accessible from the material suppliers.

In the short run, the feature importance generated by machine learning models already gives insights into the relationships between the deviation and some of the most relevant features. Through the examination of these relationships, some characteristics of orders demonstrate much more possibility of having deviation. This could help monitors in Volvo such as CMP and TMC selectively pay more attention to orders with these characteristics, and preventively react on the deviation and take precautious actions.

Regarding the future work for the Volvo, the key moves are related to data management. Firstly, some important information has not been logged into the database. For example, logistics evaluation results of LSP is not integrated into the database and therefore not able to contribute to the prediction model. The second move is related to linking the data from different phases and construct them into one common database for the benefits of data preparation for modelling. For example, since there is no part number information in the transportation booking information, we have to manually link the order from material suppliers to LSP. The requirement for the data linkage is demonstrated in Figure 7.4. The mapping of currently available data in the company for the relevant features is demonstrated in Chapter 5. External data sources such as weather and traffic information are required to generate a dynamic and real-time prediction for ETA truck arrival time. Besides, the data should be stored and managed for a longer period. For the transportation phase, current platform merely stores data last for at maximum one rolling year, which could not be enough for machine learning modelling.

When the construction of the database is done and the performance of the prediction model is improved to a deployment level as expected, the prediction model could be deployed as deviation alert embedded in the program. The informativeness of the deviation alert could be improved by increasing the granularity of output classes. The regression machine learning models could also be tested with all the desired features available. This also requires the transportation booking records to generate the information of the length of deviation in days rather than the current three classes of being early, on time and late, since the regression models require continuous values as the response variable.

Another improvement that could be made to pave the road for data mining is to add a detailed description and the responsible role for each item in the database. We discover during the project that the clarity of data in the company's current database is not satisfying. With no further description for each item especially for those items with similar names, it increases the difficulty for data practitioners to select the correct data.

Future improvements in predicting supplier lead time should be under the condition of information sharing between suppliers and buying companies. Volvo could constantly search for opportunities for production information sharing with its suppliers or assist them to predict the lead deviation from their perspective and utilize the prediction result.

8.3 Theoretical Contribution and Future Research within Academia

The first theoretical contribution of the thesis is that it is the first trial in the theory for the buying company to predict material supplier's delivery precision performance by machine learning. It shows that constructing the characteristics of ordered parts and material's supplier evaluation results as well as historical delivery performance into the input features only deliver a weak prediction power. It indicates production information of material suppliers such as the production disruption of the orders is necessary to fully capture the deviation of supplier lead time. The second contribution comes from the second prediction model of truck arrival time deviation. The factors of truck arrival time deviation are investigated and sorted. The deviation could be associated with the logistic characteristics of the cargo, the delivery performance of consignors and the capacity of consignees, the macro-environment including countries of consignors, as well as traffic and weather condition. It turns out logistic characteristics of cargo are important features. The consideration of organizational factors is much under the constraints of available data currently in the company.

For the future research of predicting supplier lead time in academia, the idea of predicting supplier performance with machine learning could be expanded to other industries where buyers have more information regarding suppliers. The performance is also not limited to the delivery precision and inbound logistics phase since machine learning has the ability to identify the complex relationship between the performance and multi-criteria of suppliers. For future research of predicting the truck arrival time, comprehensive factors should be considered as Figure 7.4 shows. The regression model should be considered when important features are ready since the outcome of a regression model is more powerful. The output being the exact number of deviation could help evaluate the default lead time in the system and reschedule the safety stock.

Bibliography

Agresti, A. (2018). An introduction to categorical data analysis. Wiley.

Ali, A., Shamsuddin, S. M., & Ralescu, A. L. (2015). Classification with class imbalance problem: a review. Int. J. Advance Soft Compu. Appl, 7(3), 176-204.

Barbour, W., Samal, C., Kuppa, S., Dubey, A., & Work, D. B. (2018, November). On the Data-Driven Prediction of Arrival Times for Freight Trains on US Railroads. In 2018 21st International Conference on Intelligent Transportation Systems (ITSC) (pp. 2289-2296). IEEE.

Belcastro, L., Marozzo, F., Talia, D., & Trunfio, P. (2016). Using scalable data mining for predicting flight delays. ACM Transactions on Intelligent Systems and Technology (TIST), 8(1), 5.

Breiman, L. (1996a). Bias, variance, and arcing classifiers.

Breiman, L. (1996b). Out-of-bag estimation.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Bryman, A., & Bell, E. (2014). Research methodology: Business and management contexts. Oxford University Press Southern Africa.

Chawla, N. V., Lazarevic, A., Hall, L. O., & Bowyer, K. W. (2003, September). SMOTEBoost: Improving prediction of the minority class in boosting. In European conference on principles of data mining and knowledge discovery (pp. 107-119). Springer, Berlin, Heidelberg.

Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. European Journal of Operational Research, 184(3), 1140-1154.

Chen, C., Liaw, A., & Breiman, L. (2004). Using random forest to learn imbalanced data. University of California, Berkeley, 110, 1-12.

Cohen, M. A., Agrawal, N., & Agrawal, V. (2006). Winning in the aftermarket. Harvard business review, 84(5), 129.

Cortez, P., & Embrechts, M. J. (2013). Using sensitivity analysis and visualization techniques to open black box data mining models. Information Sciences, 225, 1-17.

Dawson, C. (2001). Machete Time. Business Week, (3727), 42-42.

Dekker, R., Pinçe, Ç., Zuidwijk, R., & Jalil, M. N. (2013). On the use of installed base information for spare parts logistics: A review of ideas and industry practice. International Journal of Production Economics, 143(2), 536-545.

Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363.

Freund, Y., & Schapire, R. E. (1996, July). Experiments with a new boosting algorithm. In icml (Vol. 96, pp. 148-156).

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189-1232.

García-Pedrajas, N., Maudes-Raedo, J., García-Osorio, C., & Rodríguez-Díez, J. J. (2012). Supervised subspace projections for constructing ensembles of classifiers. Information Sciences, 193, 1-21.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer.

González-Recio, O., Rosa, G. J., & Gianola, D. (2014). Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. Livestock Science, 166, 217-231.

Heydari, J., Kazemzadeh, R. B., & Chaharsooghi, S. K. (2009). A study of lead time variation impact on supply chain performance. The International Journal of Advanced Manufacturing Technology, 40(11-12), 1206-1215.

Ho, W., Xu, X., & Dey, P. K. (2010). Multi-criteria decision making approaches for supplier evaluation and selection: A literature review. European Journal of operational research, 202(1), 16-24.

Huiskonen, J. (2001). Maintenance spare parts logistics: Special characteristics and strategic choices. International journal of production economics, 71(1-3), 125-133.

Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. Intelligent data analysis, 6(5), 429-449.

Ioannou, G., & Dimitriou, S. (2012). Lead time estimation in MRP/ERP for make-to-order manufacturing systems. International Journal of Production Economics, 139(2), 551-563.

Jiang, B., Chen, W., Zhang, H., & Pan, W. (2013). Supplier's Efficiency and Performance Evaluation using DEA-SVM Approach. JSW, 8(1), 25-30.

Jonsson, P. (2008), Logistics and supply chain management, McGraw Hill Education, London.

Liu, Y., & Zhao, H. (2017). Variable importance-weighted random forests. Quantitative Biology, 5(4), 338-351.

Mobley, R. K. (2002). An introduction to predictive maintenance. Elsevier.

Kennedy, W. J., Patterson, J. W., & Fredendall, L. D. (2002). An overview of recent literature on spare parts inventories. International Journal of production economics, 76(2), 201-215.

Khaldi, R., Chiheb, R., El Afia, A., Akaaboune, A., & Faizi, R. (2017, March). Prediction of Supplier Performance: A Novel DEA-ANFIS Based Approach. In Proceedings of the 2nd international Conference on Big Data, Cloud and Applications(p. 60). ACM.

Kotsiantis, S., Kanellopoulos, D., & Pintelas, P. (2006). Handling imbalanced datasets: A review. GESTS International Transactions on Computer Science and Engineering, 30(1), 25-36.

Krause, D. R., Handfield, R. B., & Tyler, B. B. (2007). The relationships between supplier development, commitment, social capital accumulation and performance improvement. Journal of operations management, 25(2), 528-545.

Lingitz, L., Gallina, V., Ansari, F., Gyulai, D., Pfeiffer, A., & Monostori, L. (2018). Lead time prediction using machine learning algorithms: A case study by a semiconductor manufacturer. PROCEDIA CIRP, 72, 1051-1056.

Öztürk, A., Kayalıgil, S., & Özdemirel, N. E. (2006). Manufacturing lead time estimation using data mining. European Journal of Operational Research, 173(2), 683-700.

Petkovic, D., Sosnick-Pérez, M., Okada, K., Todtenhoefer, R., Huang, S., Miglani, N., & Vigil, A. (2016, October). Using the random forest classifier to assess and predict student learning of software engineering teamwork. In 2016

IEEE Frontiers in Education Conference (FIE) (pp. 1-7). IEEE.

Pfeiffer, A., Gyulai, D., Kádár, B., & Monostori, L. (2016). Manufacturing lead time estimation with the combination of simulation and statistical learning methods. Procedia CIRP, 41, 75-80.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. In Advances in Neural Information Processing Systems (pp. 6638-6648).

Rezaei, J., Fahim, P. B., & Tavasszy, L. (2014). Supplier selection in the airline retail industry using a funnel methodology: Conjunctive screening method and fuzzy AHP. Expert systems with applications, 41(18), 8165-8179.

Rice, J. R. (1976). The algorithm selection problem. In Advances in computers (Vol. 15, pp. 65-118). Elsevier.

Sarmiento, R., Byrne, M., Rene Contreras, L., & Rich, N. (2007). Delivery reliability, manufacturing capabilities and new models of manufacturing efficiency. Journal of Manufacturing Technology Management, 18(4), 367-386.

Schmidt, B., & Wang, L. (2018). Cloud-enhanced predictive maintenance. The International Journal of Advanced Manufacturing Technology, 99(1-4), 5-13.

Smith-Miles, K. A. (2009). Cross-disciplinary perspectives on meta-learning for algorithm selection. ACM Computing Surveys (CSUR), 41(1), 6.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information Processing & Management, 45(4), 427-437.

Takacs, G. (2014, October). Predicting flight arrival times with a multistage model. In 2014 IEEE International Conference on Big Data (Big Data) (pp. 78-84). IEEE.

The CRISP-DM process model. (1999). Retrieved from http://www.crisp-dm.org/

Trevor, H., Robert, T., & JH, F. (2009). The elements of statistical learning: data mining, inference, and prediction.

van der Spoel, S., Amrit, C., & van Hillegersberg, J. (2017). Predictive analytics for truck arrival time estimation: a field study at a European distribution centre. International journal of production research, 55(17), 5062-5078.

Visa, S. (2007). Fuzzy classifiers for imbalanced data sets(Doctoral dissertation, University of Cincinnati).

Wang, L. X., & Mendel, J. M. (1992). Generating fuzzy rules by learning from examples. IEEE Transactions on systems, man, and cybernetics, 22(6), 1414-1427.

Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining (pp. 29-39). Citeseer.

Wolpert, D. H., & Macready, W. G. (1999). An efficient method to estimate bagging's generalization error. Machine Learning, 35(1), 41-55.

Yin, R. K., & Campbell, D. T. (2018). Case study research and applications: Design and methods. Thousand Oaks, CA: SAGE Publications.

Zanjani, M. K., & Nourelfath, M. (2014). Integrated spare parts logistics and operations planning for maintenance service providers. International Journal of Production Economics, 158, 44-53.

Zhang, Y., & Haghani, A. (2015). A gradient boosting method to improve travel time prediction. Transportation Research Part C: Emerging Technologies, 58, 308-324.

A
Appendix: Interview Questions

Interview Questions Template

Introduction:

- We describe the purpose and scope of this thesis work.

- We describe the purpose/expected output of this interview

Background Question:

- What is your role in the company?

- What is your responsibility and daily tasks?

- Are you responsible for a certain part segments, or part of supply chain or part life-cycle?

Lead time in inbound process:

- What processes are the inbound delivery process including? How long does each process take (your lead time)?

- How is this lead time generated in your system? Do you or your department set up the lead time? If yes, how? How often do you do the planning? Which data do you use?

- How much is the deviation between your theoretical lead time and real lead time? How often do they delay or ahead of time?

- How does these deviation affect your work and the company in your mind? Who will take your lead time prediction into consideration when they do their job?

- Is there any certain type of spare part (criticality/frequency/price) or any carriers/forwarder with the largest lead time deviation?

- Regarding the later process about LSP performance, how do you perceive the pick-up precision of LSP? (For material supply)

- Regarding the previous process about supplier performance, how do you perceive the actual time of spare parts ready to be shipped? (For transportation)

- What are the factors do you think that are influencing your lead time? Which factors do you have available data to measure?
    - Internal factors
    - External factors

- Do you measure the performance of your lead time prediction? If yes, how?

- Is there anyone following up your lead time deviation? Is there anyone responsible for updating your lead time prediction if necessary? And how do they do that?

Other questions:

- Do you recommend any other people who have the relevant knowledge about lead time or lead time prediction?

- Is there anything you want to add that did not mention before?

- Would you mind being contacted further if any information is needed to be checked or addition information are needed from you?