MASTER

Improving transport planning at a logistic service provider using machine learning

Wijnhoven, T.H.M.

Award date:
2020

EINDHOVEN UNIVERSITY OF TECHNOLOGY

Department of Industrial Engineering and Innovation Sciences

Operations Planning Accounting & Control Group

Improving transport planning at a logistic service provider using machine learning

T.H.M. Wijnhoven

Supervisors:

dr. ir. P.M. Singh

dr. L.P. Veelenturf

H. Schut, DHL

Key words: Machine Learning, Business rules, Consolidation, Transport, Planning, Interoperability

EXECUTIVE SUMMARY

This research is done for the Lead Logistics Partner (LLP) group within DHL Global Forwarding. LLP is responsible for bringing continuous improvements and cost reductions; introducing lean logistics processes and optimizing logistics networks.

PROBLEM STATEMENT

The LLP core operations are guided by Transport Management Systems (TMS). The 'heart' of the LLP TMS is the Business Rule – concept. Based upon business rules shipments get assigned to a carrier. These rules have been the same for years, and are still applied every day. Therefore, these rules could not be representative anymore, which leads to more costly solutions. Current planning solutions for Company A yield 37% higher costs in comparison to the optimal solutions, which implies that the current business rules should be updated. With the current approach it takes a lot of effort to update these rules. Therefore a new approach is investigated, that can improve the business rules to make better planning decisions using machine learning algorithms.

RESEARCH GOAL

The goal of this research is to investigate if machine learning can be used to improve the current rules for the assignment of shipments to carriers, and to extent the current rules with the option for consolidation of shipments. Consolidation is the process of combining different shipments into a single vehicle load. The economies of scales thus achieved have many benefits including lower cost, lower carbon emission, greater flexibility and supply chain resilience. In short, the following parts will be present in this thesis:

- An extensive data analysis of the client dataset.
- The development of machine learning models that are able to make planning decisions for shipments.
- An improved set of business rules to support transport planning decisions, derived from machine learning models.
- A final set of recommendations on how to use the new set of rules obtained.

DATA ANALYSIS AND PREPARATION

First, the historical dataset provided by the client was analysed to gain insights, and to detect possible outliers and errors in the data. Some outliers and errors were identified, that were taken care of. Then, another analysis was done to verify if there was enough activity on transportation lanes for consolidation. In combination with the decreasing cost per kg structure for the operating carriers, it was concluded that consolidation should be feasible.

Next, the deterministic models were developed to obtain optimal planning decisions for both single shipments and consolidation of shipments. At forehand, it was decided to consider two different consolidation scenarios: *same day consolidation* and *multiple day consolidation.* The optimal planning decisions obtained with the deterministic models were then labelled to all shipments in the historical dataset.

Machine Learning Models

With the optimal planning decisions being labelled to the datasets, *classification tree (CT)* and *neural network (NN)* models were trained to learn planning decisions. Separate models were developed to for the planning of single shipments and consolidation of shipments. Shipments are consolidated per transport lane following a same day or multiple day consolidation plan. Both plans were supplied to the models. For single shipments alone savings of 33.5% (CT) and 34.8% (NN) can be obtained. For consolidation; same day consolidation yielded more cost savings than consolidation over multiple days for both algorithms used.

Final Results and Recommendations

From the CT models IF THEN rules could be generated. Since separate rules exist for single shipments and consolidation per transport lane; it is recommended to combine the rules for single shipments with the different rules for consolidation on transport lanes. This resulted in around 1300 new business rules that can directly be integrated into the TMS. With these new business rules, cost savings of around 33.5% can be expected for the planning of single shipments. On the 5 transport lanes with a same day consolidation program additional cost savings between 7% and 32% can be expected, dependent of the transport lane.

q1: could you may be explain how you combined the rules for single shipment and for consolidation of shipments?

CONTENTS

1 INTRODUCTION

Machine learning is gaining popularity fast around the globe, both to drive performance improvements and to gain new insights (Papadopoulos, Gunasekaran, Dubey, & Fosso Wamba, 2017). With machine learning it is possible to discover patterns in data, relying on algorithms that quickly determine relationships within data. Various research studies have indicated the benefits machine learning can yield by optimizing supply chain design, like reducing costs and mitigating risks. (Fosso Wamba, Gunasekaran, Papadopoulos, & Ngai, 2018). However, still a lot of companies rely on knowledge and expertise within the company to make decisions. This knowledge is put into business rules that form the basis of decision making (Appendix A). Mainly due to the lack of top management support, IT infrastructure/capabilities, and financial readiness, business are not able/ready to adapt machine learning yet in their current decision making (Lai, Sun, & Ren, 2018).

In this research, it will be investigated if machine learning can be used to improve decision making in a supply chain/logistics environment. A case study is done for a client of DHL (referred to as *Company A),* to investigate how the current decision making can be improved to reduce costs. Due to current limitations in the IT infrastructure, machine learning models cannot directly put to use. Therefore, this research goes beyond the model itself. We investigate the learnt relations within the machine learning models to generate business rules. These business rules can then be used to support and improve transportation planning, without having to implement a machine learning model into the management systems. This often requires the use of other management systems or extended versions of the current management systems.

A novel approach will be presented to improve transportation planning via machine learning in the supply chain/logistics sector. An approach like this could be used to smoothen the transition from logical decision making, to machine learning models that support the decision making via business rules.

1.1 COMPANY INTRODUCTION

DHL is one of the biggest logistics companies in the world; DHL’s total revenue over 2018 added up to more than 60 billion EUR. DHL's business is organized into four different divisions: Express, Supply Chain, Global Forwarding and Deutsche Post/Parcel. The Global Forwarding division, which is involved in global package transportation, accounted for almost 24% of the total revenue, or about EUR 14 billion. In contrast to DHL's other divisions, it has a very asset-light business model, which is based upon the brokerage of shipment services between clients and freight carriers. As a result, the division accounts for less than 10% of DHL's 459,000 employees.

This research is specifically done for the Lead Logistics Partner (LLP) group within DHL Global Forwarding. LLP is responsible for investigating and managing change across the entire supply chain of a client. This is done by bringing continuous improvement and cost reduction, introducing lean logistics processes and optimizing logistics networks. LLP offers all kind of services, like: office management services, customs services, inventory management, freight bill and supply chain design. Furthermore, LLP organizes and operates these services for day to day operations in a control tower setting.

1.2 PROBLEM STATEMENT

The LLP core operations are guided by Transport Management Systems (TMS). A TMS is a software system that deals with the controlling, planning, execution and optimization of the transport operations within a company. The 'heart' of the LLP TMS is the Business Rule – concept. These business rules are used to assign shipments to carriers. These come down to rules such as "Any package made of type X under 70 kg has to be transported via carrier Y", and are based upon logic and knowledge.

Due to the ever changing sector, changing prices of transport services and changing customer demands, the current business rules get outdated. Such changes have impact on the quality of the decisions supported by the current business rules. However, still the same rules from years ago are applied every day. This leads to more costly solutions (*Appendix B*). Current solutions for Company A lead to 37% higher costs in comparison to the optimal solutions (*Appendix C*). To give an idea of the size of Company A, 10,000 shipments every six months with a total weight and a total volume of around 340,000 kilogram (kg) and 3,100 m^3 respectively (*Appendix D*). Figure 1 gives an overview of the current decision making.

Figure 1: Current planning of shipments

To provide the most cost effective planning solutions (optimal solution) to its clients, the current decision making needs to be optimized. Due to limitations in the current IT infrastructure, planning algorithms/models cannot be implemented within the current TMS. This implies that the business rules should be updated. With the current approach it takes a lot of effort to update these rules, therefore a new and smarter approach is required.

1.3 RESEARCH GOAL

A novel approach is investigated to generate improved business rules via *machine learning*. Various research studies have indicated that machine learning can be beneficial for optimization within the supply chain/transportation sector (Fosso Wamba et al., 2018). However, most machine learning algorithms give rise to what has been called the "*the black-box problem*" (Castelvecchi, 2016). In recent years already a lot of research has been done to better understand the learnt relations within machine learning models to generate understandable rules from them (Townsend, Chaton, & Monteiro, 2019). By using a machine learning approach to generate business rules, we want to investigate:

- the application of machine learning in the field of transportation planning, and
- the interpretability of machine learning algorithms by deriving business rules from the learnt relations within the models.

Currently the (outdated) business rules support planning decisions of individual shipments on an operational level, e.g. the assignment of shipments to carriers. This implies shipments are not combined and transported with other shipments. The economies of scales achieved when combining shipments have many benefits including lesser costs, lower carbon emission, greater flexibility and supply chain resilience. The process of combining different items (shipments) into single vehicle loads, is called consolidation (Ülkü, 2012). Therefore, it will be investigated how to improve the current business rules for the assignment of shipments to carriers, and how to extend the business rules with the option for consolidation of shipments. These new business rules should provide the optimal planning solution to the customer; the most cost effective planning solution. To come up with these new business rules, it is investigated if machine learning models can be trained to make optimal decisions regarding the planning of individual shipments and consolidations. To be able to train optimal planning decisions, these optimal decisions first need to be obtained. A deterministic model will be developed to determine the optimal planning of shipments and consolidated shipments for all shipments contained in a historical dataset. Then it is investigated if it possible to derive business rules from the trained machine learning models. Concluding, it is investigated if the current planning situation (illustrated in Figure 1) can be transformed into the situation sketched in Figure 2.

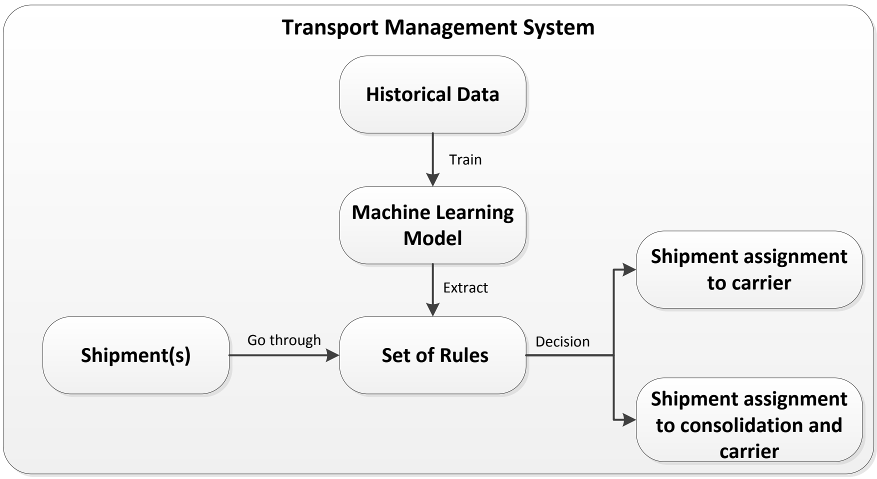

Figure 2: To-be planning of shipments

The following research question is formulated: *How can business rules be developed with machine learning, to select the optimal planning solution for shipments that benefit DHL LLP customers?*

1.3.1 Sub-questions

The following sub questions need to be answered in order to fulfil the research goal:

1. Which machine learning algorithms can be used for transportation planning?
2. What is the current planning situation of shipments for LLP?
3. How can the transportation network of Company A be described?
4. How can the freight consolidation problem be defined for LLP?
5. Which consolidation methods are used in prior research?
6. Could a consolidation program be feasible for LLP, given the historical data of Company A?

1.3.2 Contributions of research

The contribution of this research based upon the research questions is the following:

- An extensive data analysis of the client dataset.
- An overview of the current planning of shipments.
- An overview of consolidation methods used in prior research.
- Implementation and evaluation of the machine learning models developed.
- An improved set of business rules, derived from machine learning models.
- A final set of recommendations on how to integrate the new sets of business rules obtained.

q2: If you remove the last three words and replace by 3pl, what is your contribution then, I think your research offers somethings on the generic scope, what is it?

2 BACKGROUND STUDY: MACHINE LEARNING

A background study about machine learning is done, to search for algorithms that could be applicable for usage in the transportation planning domain. Starting with a brief explanation of machine learning, then an overview of applicable algorithms is provided. Finally, literature is explored on the application of these algorithms in the transport/logistics domain.

Machine learning refers to a class of data science models that can learn from data to improve their performance over time. Machine learning extracts knowledge from data, which can then be used for predicting and generating new information. Machine learning is particularly useful in dealing with tasks that cannot be explicitly instructed by an analytic solution, such as image and voice processing, pattern recognition, or complex classification tasks (Ghoddusi, Creamer, & Rafizadeh, 2019). Machine learning problems can be divided in two different learning tasks (Provost & Fawcett, 2013):

1. *Un-supervised learning*; when there is no clear target to predict.
2. *Supervised learning*; when there is a clear target to predict

Each of these learning tasks have their own subset of problems. The goal of this master thesis is to learn optimal planning decisions (target) with machine learning. This requires a supervised learning approach. Relevant supervised learning problems are:

- *Classification problem:* is the problem of training an algorithm to predict several predefined categorical classes. The algorithm learns a function that maps (classifies) a data item (set of variables) into one of the several predefined categorical classes (Choudhary, Harding, & Tiwari, 2009).
- *Regression problem:* maps a data item to a real-valued prediction variable (Mitra, Pal, & Mitra, 2002).

It was decided that the problem at hand is a classification problem. To solve classification problems numerous algorithms can be deployed. A brief overview of relevant algorithms that can be used for supervised learning problems will be discussed now.

- *Decision tree/Classification tree:* is an algorithm based upon the structure of trees, that classify instances by sorting them based on feature values. Each node in a decision tree represents a feature in an instance to be classified, and each branch represents a value that the node can assume. Instances are classified starting at the root node and sorted based on their feature values. The problem of constructing optimal decision trees is a NP-complete problem (Kotsiantis, Zaharakis, & Pintelas, 2006). Efficient heuristics have been developed for constructing near-optimal trees e.g. C4.5 (Salzberg, 1994). For each path from the root to a leaf, rules can easily be extracted (Kotsiantis et al., 2006). The design of an example classification tree model is illustrated in Figure 3.

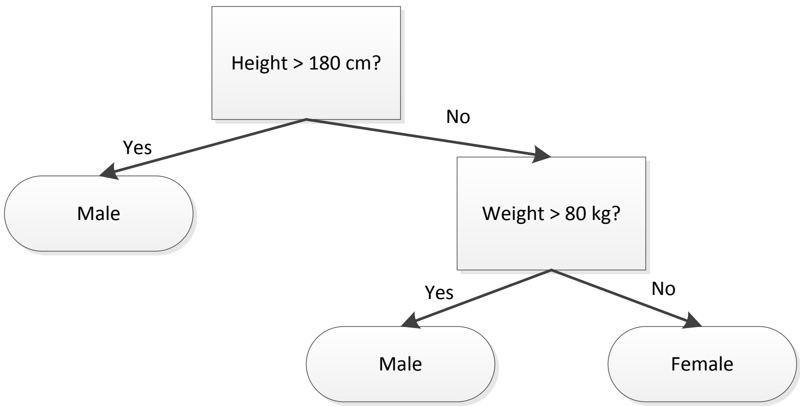

Figure 3: Example classification tree that can classify a person as either a man or a woman

- *Neural Network (NN):* was initially introduced as an algorithm that simulates how the brain works based on the connection between neurons (Ghoddusi et al., 2019). General, a neural network model includes the input layer, neural network layer and output layer. Each layer consists of a number of neurons, the neurons from different layers are connected with each other. The input layer is the beginning of the neural network, which brings the input information into the neural network for further processing by the following layers. In the neural network layers (hidden layers), different neural networks such as artificial neural networks, convolutional neural networks (CNNs) and recurrent neural networks (RNNs) can be employed to learn the desired output. The output layer is generally used to implement the desired output based on the latent feature representation (Wang, Zhang, Guo, & Yi, 2018a). An example design of a neural network is illustrated in Figure 4.

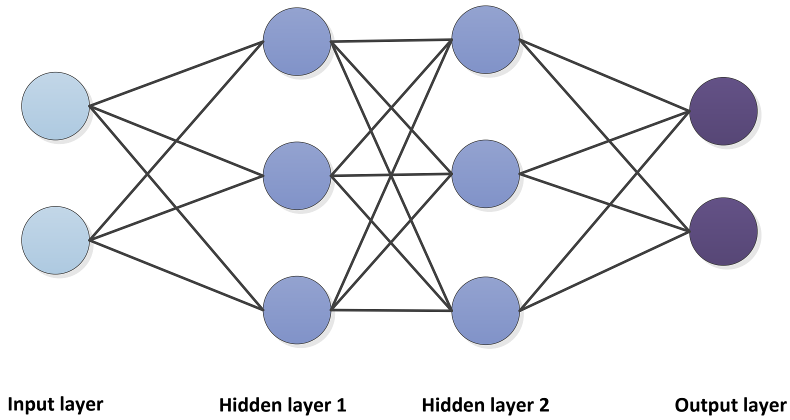

Figure 4: Example design of a neural network, where each dot represents a neuron

- *Bayesian Network:* is a graphical model for probability relationships among a set of variables (Kotsiantis et al., 2006).
- *K-nearest neighbours (KNN):* is an algorithm that classifies data points according to their similarity. To determine similarity, several distances metric can be used within the algorithm, e.g. *Manhattan distance* and *Euclidean distance* (Bhavsar & Ganatra, 2012).
- *Support Vector Machine (SVM):* is a classification method that classifies data through a hyperplane that maximizes the distance between observations that belong to each category (Ghoddusi et al., 2019). An example is visualized in Figure 5.

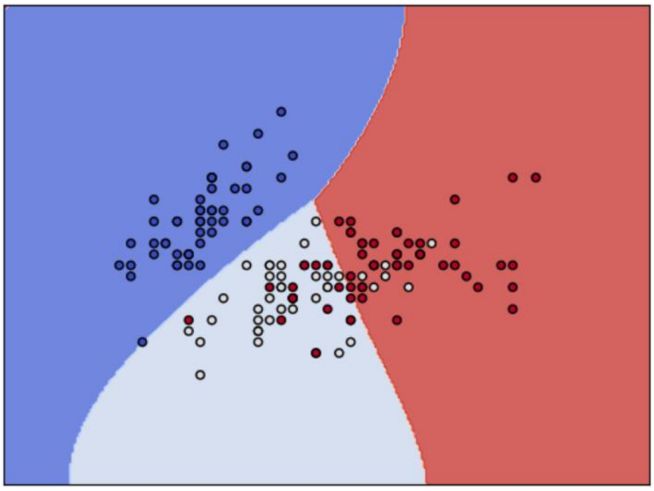

Figure 5: Example of hyperplanes classifying dots of three different colours

- *Hybrid and Ensemble Models:* are multiple machine learning models (based on different algorithms) combined to improve the forecasting accuracy and robustness compared to individual models (Kourentzes, Barrow, & Crone, 2014).

Literature was explored on the application of these classification algorithms in the field of transportation planning. It was discovered that there exists a literature gap on this topic. However, it was found that the application of neural networks has been studied often in the last 10 years within the transportation/supply chain sector. *NN* approaches are characterized as the foundational technological drivers for the optimization of transportation and logistics (Cheng, Yang, Gen, Jang, & Liang, 2020). It must be noted that also the other algorithms just mentioned, are developed to solve problems within the sector. For example (Hughes, Moreno, Yushimito, & Huerta-Cánepa, 2019) investigate different algorithms (e.g. *decision tree*, *SVM* and *NN*) to optimize last mile logistics with a better approximation of delivery times.

q4: do you still think you choose the right algorithm, what is the research gap actually?

Could you point out one article from the reference..may be?

3 METHODOLOGY

In this research, machine learning models will be used to fulfil the research goal. Therefore, the research design is based on the cross-industry process for data mining (CRISP-DM) methodology (Wirth, 2000), since the CRISP-DM methodology provides a framework for carrying out machine learning projects. As can be seen in Figure 6, the model consists of 6 phases. It was decided, that this research will be done in two cycles, since both cycles have different objectives:

- *Research cycle 1: Single shipments,* investigate if business rules can be developed using machine learning that are able to allocate non- consolidated shipments to the optimal carrier.
- *Research cycle 2*: *Consolidation of shipments,* investigate if business rules can be developed using machine learning for the *consolidation of shipments*.

The CRISP-DM methodology is slightly adapted to suit the needs of this research. The main difference is, that in the business understanding phase of research cycle 2 also a literature review will be conducted. The deployment phase is not done due to time and money constraints, and is left for future research. Furthermore, a new phase will be introduced which will include the development of the deterministic models. This phase will take place between the data understanding and data preparation phase. Each phase of the research design is elaborated briefly. For each phase, the two research cycles are elaborated. For the exact research design see Figure 6, where the output of each phase is specified per research cycle.

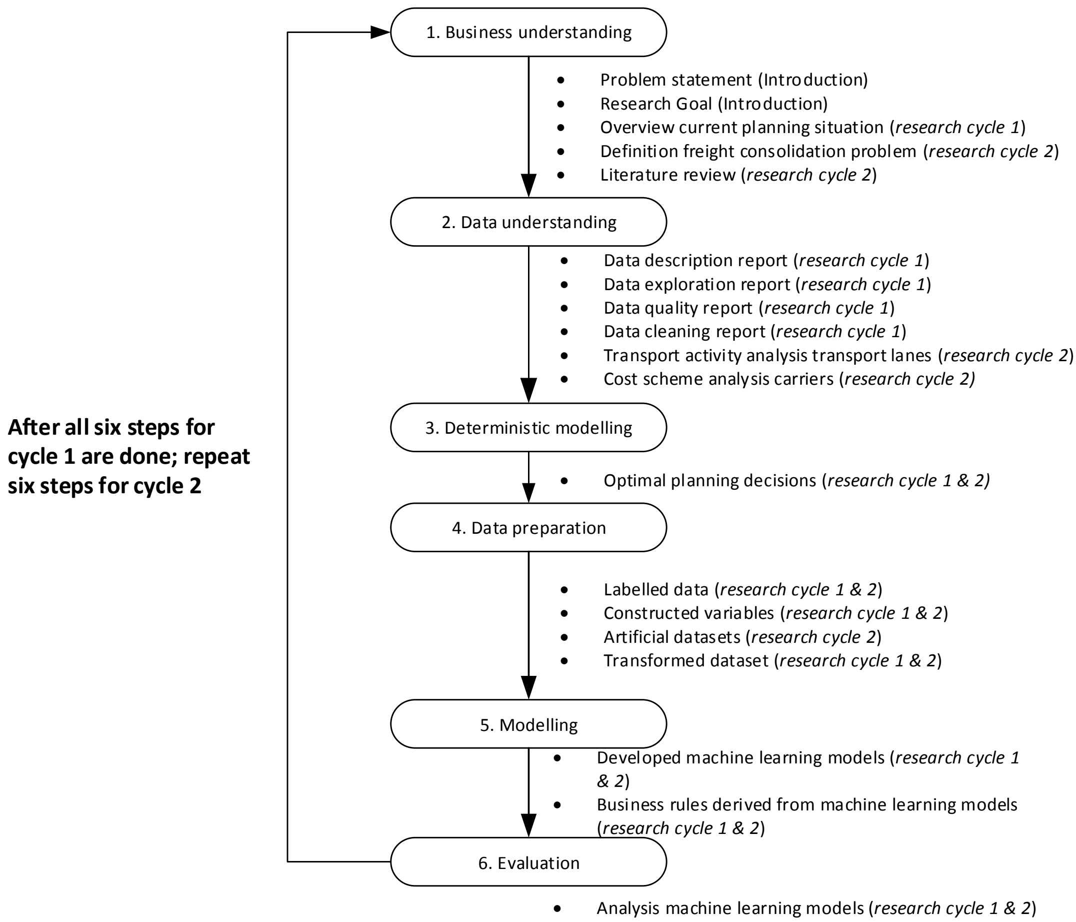

Figure 6: Research design

3.1 BUSINESS UNDERSTANDING

The first step is to understand the current planning of shipments at LLP and to understand how they could influence future planning decisions. More specific, in this phase an answer will be provided to sub-question 2, 3, 4 and 5:

2. *Research cycle 1*: What is the current situation for LLP regarding the planning of shipments?
3. *Research cycle 1*: How can the transportation network of Company A be described?
4. *Research cycle 2*: How can the freight consolidation problem be described for LLP?
5. *Research cycle 2:* Which consolidation methods are used in prior research?

In research cycle 1, semi-structured interviews with employees of LLP will be done to understand the current planning of shipments and the transportation network of Company A. In research cycle 2, a literature review is done to determine which consolidation methods are used in prior research and the consolidation problem is defined to fit the current business processes. The literature review will be a systematic literature review (Kitchenham & Charters, 2007), and the consolidation problem will be defined based upon semi-structured interviews.

3.2 DATA UNDERSTANDING

The goal for both research cycle 1 and 2 in the data understanding phase is the same: to describe the data, to understand the strengths and the limitations of the data, and to explore the data to get insight in the problem. In essence historical data is often collected for purposes unrelated to the current business problem (Provost & Fawcett, 2013). Therefore, checking the quality of the data is important. Descriptive statistics and visualizations will be used in both steps, to get more understanding of the data, and to detect possible outliers and limitations of the data. For the data understanding in research cycle 1, the main goal is to identify outliers and limitations of the data, and to get an idea of the mix of shipments in the historical dataset. Quality issues identified will be solved, since data needs to be cleaned before the deterministic modelling phase. The data understanding in research cycle 2 will focus more on determining whether a consolidation program would be feasible, given the historical data provided and the cost schemes of the carriers in scope. This will answer sub-question 6.

3.3 DETERMINISTIC MODELLING

In the third phase, the deterministic models for both research cycle 1 and 2 will be developed. For research cycle 1, a model will be developed that is able to assign shipments to the optimal carrier. Remember that the optimal carrier is defined as the most cost effective carrier, while maintaining service constraints. To be able to do so, prices need to be calculated. Information on how to calculate the price of a shipment for a carrier, will be gained by semi structured interviews with employees of LLP. For research cycle 2, the optimal consolidation scheme will be created by formulating and solving an integer optimization problem (Tyan, Wang, & Du, 2003b). Inspiration will come from the literature review done in the business understanding phase (section 5.1.2).

3.4 DATA PREPARATION

Data preparation in this research project differs from the usual data preparation described in the CRISP-DM model. Since the dependent variable for research cycle 1 and 2 needs to be constructed first to complete the dataset. Based upon the optimal planning decisions obtained with the deterministic models, the historical dataset will be labelled to create the dependent variable. This is an important step in the data preparation process for both cycles. Then for both research cycles, the data needs to be prepared for modelling, this implies: selecting data, transforming categorical values, transforming numerical data and creating a test and training set. The deliverable for this phase is a prepared dataset for research cycle 1 and 2.

3.5 MODELLING

In the fifth phase, the machine learning models will be developed to learn planning decisions for research cycle 1 and 2. In Chapter 2, it was already stated that the machine learning approach investigated, will be classification learning approach. From the algorithms described in our background study, it was decided to develop machine learning models based upon two different algorithms: *classification trees* (*CT*) and *neural networks* (*NN*). For the reason that rules can directly be derived from *CT* models without loss of accuracy (Kotsiantis et al., 2006). We also choose to develop a *NN*, since in the past 10 years the application of *NN* in the transport/supply chain domain is often being studied and literature shows understandable rules can be derived of *NN* models (Townsend et al., 2019). However, rule extraction techniques for *NN* will be not studied in this research, and will be left for future research.

From the background study in Chapter 2, it became known that there are three main types of *NN* algorithms that can be used for classification: *Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN)* and *Recurrent Neural Networks (RNN). CNN* are in general used for image classification (Al Rahhal, Bazi, Al Zuair, Othman, & BenJdira, 2018). Whereas *RNN* are used for classifications problems involving (time) sequences, because of their memory mechanisms (Wang, Zhang, Guo, & Yi, 2018b). *ANN* have proved to be successful in different classification applications (Haglin, Jimenez, & Eltorai, 2019). The problem at hand does not involve image data or sequence data, therefore it was chosen to develop an *ANN* model.

In research cycle 1, first a brief explanation of both algorithms will be given which includes an overview of relevant parameters of both algorithms. Then for both research cycles the whole modelling process is described, development and parameter tuning. The deliverables after this phase are: *CT* and *ANN* models for both research cycles.

3.6 EVALUATION

The sixth phase is the evaluation phase, in this phase the performance of the *CT* model and *ANN* model are evaluated. This is done by simulating the current decision making against the improved decision making by the machine learning models. Note that rules can directly be derived from the *CT* models without loss of accuracy, which implies that performance obtained with the *CT* model and the rules derived from the *CT* model will be the same. The purpose of the evaluation stage is to assess the results rigorously and to gain confidence that they are valid and reliable before moving on (Provost & Fawcett, 2013). To do so, it has been decided that five different performance indicators can be used to evaluate performance:

1. *Performance based upon accuracy and precision* (Hossin & Sulaiman, 2015); *accuracy* can be calculated as the percentage of all labels classified correct by the machine learning model:

$$\frac{Number\ of\ right\ predicted\ labels}{Number\ of\ right\ predicted\ labels + Number\ of\ wrong\ predicted\ labels}$$

Whereas the *precision* states how accurately a specific label can be classified by the machine learning model. *Precision* can be calculated with a formula based upon the confusion matrix (see Figure 7) for a specific label:

$$\frac{True\ Postives}{True\ Postives + False\ Positives}$$

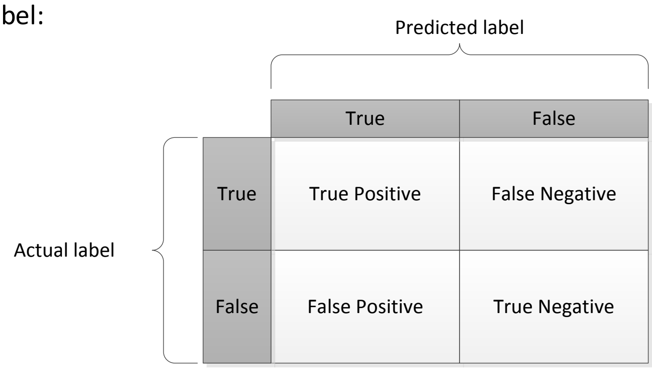

Figure 7: Confusion matrix

2. *Relative savings*: relative savings obtained with machine learning models with respect to the current decision making.
3. *Optimality gap*: gap obtained between the costs of the deterministic model (optimal solution) and the machine learning model.
4. *Feasibility of the carrier allocations done by the machine learning models*. Non-feasible allocations occur when one of the following constraints are violated (from most important constraint to least important constraint):
    - *Hazardous constraints*: hazardous transport should only be transported with carriers that are allowed to do so.
    - *Lead time constraints*: transports should be at their destination before or at the requested delivery date.
    - *Capacity/dimensions constraints*: capacity and dimensions of shipments should comply with the chosen carrier.
5. *Additional savings (research cycle 2 only)*: Additional savings obtained with consolidation over the planning of single shipments.

3.7 OVERVIEW

Table 1 gives an overview of the research sub-questions and with which method they are solved in which phase of the research design.

Table 1: Overview research sub-questions

Research question	Phase	Method	Research cycle
1. Which machine learning algorithms can be used for transportation planning?	Background study (see Chapter 2)	Literature review	n.a
2. What is the current planning situation of shipments for LLP?	Business understanding	Semi-structured interviews with stakeholders	1
3. How can the transportation network of Company A be described?	Business understanding	Semi-structured interviews with stakeholders	1
4. How can the freight consolidation problem be defined for LLP?	Business understanding	Semi-structured interviews with stakeholders	2

5. Which consolidation methods are used in prior research?	Business understanding	Systematic literature review	2
6. Could a consolidation program be feasible for LLP, given the historical data of Company A?	Data understanding	Data analysis & data visualizations	2

3.8 Scope

This research is performed specifically for Company A, a company that operates in spare parts for machinery. The business rules proposed will be exclusively valid for this company. However, the approach adopted can be used to generate business rules in the whole supply chain/logistics domain. This research focusses on generating business rules from machine learning models that can make optimal planning decisions.

For each shipment, two different planning decisions are considered:

- Assign non-consolidated shipments (single shipment) to carrier
- Consolidate shipments on pre-determined origin-destination pairs and assign consolidation to a carrier.

When consolidating shipments, shipments are hold in inventory for a certain time. Usually some sort of inventory costs occur during this time. However, these costs are not relevant for this research project.

3.9 Report Outline

This Master thesis will be structured as follows; Chapter 4 describes all phases of research cycle 1: single shipments. Then, in Chapter 5 all phases of research cycle 2 are discussed: consolidation of shipments. Next, a discussion on the results is provided in Chapter 6. Based upon the results obtained in both research cycles, recommendations, limitations and future research will be discussed. Finally, a conclusion will be provided in Chapter 7.

4 RESEARCH CYCLE 1: SINGLE SHIPMENTS

In this section, research cycle 1 regarding single shipments is discussed, in which will be investigated how the current rules for single shipments can be improved using machine learning. In section 4.1, the transportation network and the current planning of shipments is discussed. Then, in section 4.2 the dataset is analysed for more understanding of the demands required by Company A. Next, in section 4.3 the deterministic model is presented to obtain optimal planning decision for single shipments. The data preparation done before modelling is discussed in section 4.4. Then, the development of the machine learning models is discussed in section 4.5, which includes a brief explanation on the algorithms and its parameters, libraries used, variable selection and parameter settings. Finally, the results obtained with the models developed, are presented and discussed in section 4.6.

4.1 BUSINESS UNDERSTANDING

As stated in the introduction, the business model of DHL Global Forwarding is based upon the brokerages of shipment service between clients and freight carriers. One of the most important tasks of LLP involves the planning of shipments requested by their customers. A shipment is defined as the collective of all packages (ship units) within the same transport request. A transport requested is initiated by a customer to transport a shipment (containing of ship units) from A to B. LLP can plan shipments all over the world having several carriers to choose from.

4.1.1 Transportation network

To illustrate the transportation network of Company A, an example is given in Figure 8.

Figure 8: Transportation network client LLP (hub-spoke formation)

The blue dot represents the location of Company A, and the green dots represents customers of Company A. Every shipment goes either from Company A to its customers (outbound shipments), or vice versa (inbound shipments). This implies that the transportation network of Company A has a *hub-spoke* formation, where each spoke is called a *transport lane*.

4.1.2 Current Planning of Shipments

The planning of shipments is done via the TMS. Inside the TMS, business rules determine which shipment to assign to a carrier. The business rules are simple rules that have been developed a couple of years ago to help the TMS make decent decisions fast. These rules distinguish between the characteristic of a shipments to determine which carrier to choose. The business rules are mutually exclusive collectively exhaustive, which implies that every shipment can only be assigned to one carrier at the time. Each shipment has certain characteristics (e.g., weight, volume etc.). These different characteristics of a shipment, used to make decisions, will be described in the following section.

Each shipment has an origin (where to pick up the shipment) and a destination (where to deliver the shipment). Since not all carriers can provide transportation to all regions, this information can influence which carrier is chosen. Furthermore every shipment has a weight and a volume. Based upon the volume, the volumetric weight can be calculated by multiplying the volume with a constant factor dependent of the carrier. The greatest of the weight and volumetric weight is called the chargeable weight, and determines for a great extent the price of a shipment. The volume of a shipment is determined as the total volume of all ship units contained in a shipment. Where the volume of a ship unit is calculated based upon the dimensions of a ship unit; length, width and height. Given all ship units within a shipment; the maximum weight, length, width and height found among those ship units are relevant for planning decisions of a single shipment. Since carriers can have restrictions on the maximum length, width and height of ship units contained in a shipment. The same holds for the maximum weight of ship units contained in shipments. For every shipment, the client requests a pickup date and a delivery date. The difference in time between these two is the requested lead time of the shipment. Since not all providers can handle the same lead time, it is important to know the lead time of a shipment. Besides a requested lead time, every shipment also has a service level. The service level of a shipment states how urgent a shipment is for the customer. Different service levels can be distinguished:

- *Aircraft on ground (AOG):* Outbound shipments with high priority (lead time of 0-2 days)
- *Critical (CRT):* Outbound shipments with medium priority (lead time of 2-5 days)
- *Routine (RTN):* Inbound/Outbound shipments with low priority (lead time > 5 days)
- *Non-routine (NRTN):* Inbound shipments with high to medium priority (lead time of 0-5 days)

4.1.3 Carriers in scope

There are 5 different carriers in scope; these carriers will be referred to as: Carrier A, Carrier B, Carrier C, Carrier D and Carrier E. Each carrier can provide different service for different prices, these services can differ in service conditions (e.g. how fast a shipment can be delivered), capacity restrictions and area restrictions. For Carrier A, Carrier B, Carrier C and Carrier E, these services differ in terms of capacity restrictions, while having the same service conditions. For Carrier D services differ in terms of delivery speed: service level 1 and service level 2. It should also be mentioned that Carrier D has divisions in several countries, each country can charge different prices for the same service level. This implies that for each shipment assigned to carrier D; Carrier D can operate from either the division of the origin country or the destination country.

4.2 DATA UNDERSTANDING

In this section, first is described which data was collected for this project in section 4.2.1. Then, the data is described in section 4.2.2, followed up by an exploration report of the data in section 4.2.3. Finally, the quality of the data is assessed in section 4.2.4. Quality issues found are then solved in section 4.2.5.

4.2.1 Data collection

For this project data of the year 2019 was provided by Company A, which contains shipments from January 2019 till June 2019 (149 days). This period of 149 will be referred to as the *planning period* in the latter stages of this report. The data consisted of two separate files for inbound shipments and outbound shipments.

4.2.2 Data description

A brief overview of the data contained in the inbound shipments file and outbound shipments file are described in Table 2.

Table 2: Brief overview datasets

File	Columns	Rows	Start date	End date
Inbound shipments	49	8055	01-01-2019	31-05-2019
Outbound shipments	49	19957	02-01-2019	31-05-2019

It was verified that both files have the same columns available in the same order. However, it must be noted that service levels are stated in different columns, due differences in terminology (see section 4.1.1). All variables contained in both files, and the difference obtained are described in *Appendix E.* The data is structured in a way that every shipment has a unique *"Transport code"* Each shipment can consist of multiple ship units which each have the same "*Transport code*", but an unique *"Transport ship unit code"*. Every shipment consists of a "*dummy*" row (which summarizes all ship unis contained in a shipment) and one or multiple rows that contain information specific for the ship units in the shipment (e.g., dimensions, weight etc.). The amount of unique shipments per dataset can be determined based upon on every unique *"Transport code"*; this resulted in 3603 inbound shipments and 10249 outbound shipments.

4.2.3 Data exploration and visualization

In this section, the data is explored to find outliers, errors and patterns in the data. For this purpose the inbound shipments and outbound shipments were merged into one dataset. First the data exploration will be done for numerical variables, and then for categorical variables.

Descriptive statistics of the gross weight and gross volume of shipments are given in Table 3.

Table 3: Descriptive statistics of the gross weight and gross volume of shipments

	Mean	Std.	Min.	25%	50%	75%	Max.
Gross weight (kg)	281.45	1983.93	0	0.6	1.8	8	26000
Gross volume (m^3)	0.54	18.74	0	0.01	0.02	0.09	1911.36

As can be seen in Table 3, the mean gross weight of shipments is 281.45 kg with a high standard deviation of 1,983.83 kg relative to the mean. 75 percent of shipments are 8 kg or less which indicates most shipments are relatively small. Same holds for the gross volume, the standard deviation of 18.74

m^3 seems high compared to a mean of 0.54 m^3. This is confirmed when looking at the fact that 75% of shipments have a volume of 0.09 m^3 or less. As can be seen in Table 3, volumes can be as large 1911.36 m^3. Every shipment with a gross volume higher than 100 m^3 is excluded from the dataset, since these shipments cannot be transported with one of the carriers in scope. It can also be observed that there are shipments that have either 0 kg gross weight or 0 m^3 gross volume. To learn more about the distributions of gross volume and gross weight boxplots were made, see Figure 9 and 10.

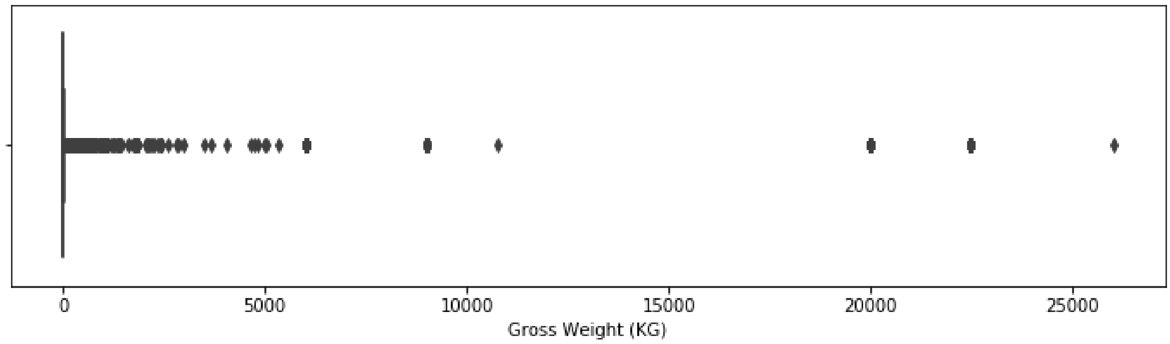

Figure 9: Boxplot gross weight showing positively skewed distribution

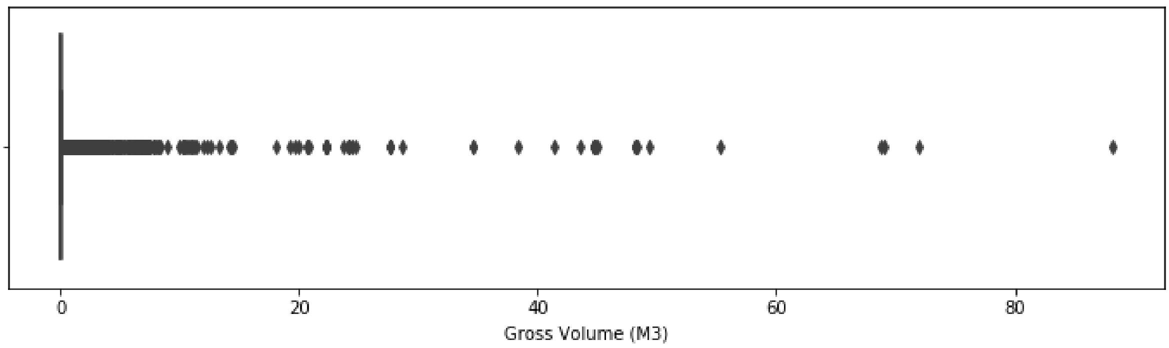

Figure 10: Boxplot gross volume showing positively skewed distribution

Figure 9 and 10 show that both the gross volume and the gross weight of shipments are positively skewed, which confirms the fact that the mean is higher than the median for both volume and weight. Furthermore, it stands out that the great majority of shipments have a low gross weight and gross volume. This expected considering the core business of the client in scope (spare parts). There are shipments with a gross weight of more than 20,000 kg. A gross weight of above 20,000 kg is not uncommon in the transport sector, but it could indicate possible outliers in this dataset. These shipments will be further investigated in the next section. The relation between gross weight and gross volume is checked to identify possible patterns and outliers in the data. This is visualized with a scatterplot, as can be seen in Figure 11.

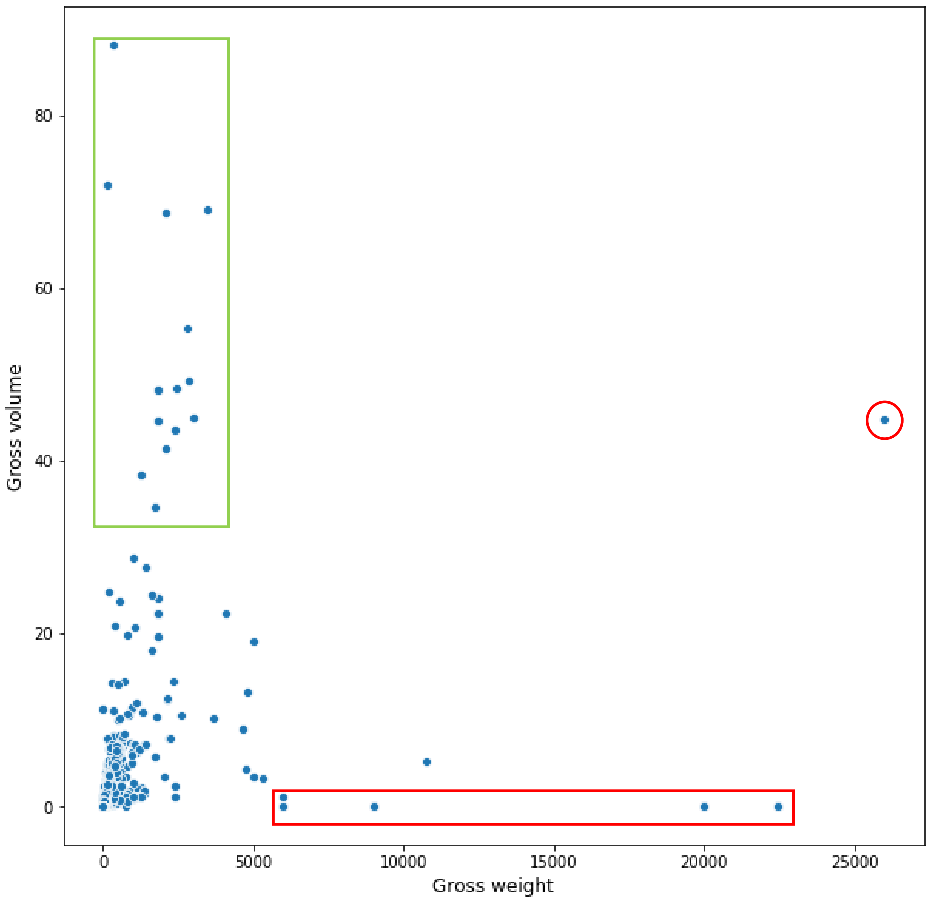

Figure 11: Scatterplot shows no relation between gross weight and gross volume

From the scatterplot in Figure 11 becomes clear that there is no real relation between gross weight and gross volume. Whereas most shipment have low gross volume in combination with low gross weight, there are shipments with high gross volume and low gross weight and vice versa. Furthermore, it can be seen in the red box that there are a few shipment that have a gross weight of more than 5,000 kg and 0 m^3 gross volume. This was checked, and eventually it was concluded that these shipments are so called "*Template shipments*". Template shipments are automatically generated non-real shipments by the TMS to reserve capacity for peak demands, and should therefore be excluded from the dataset. In section 4.2.5, these shipments will be deleted from the dataset. The shipment circled in red has a weight of 25,000 kg. None of the carriers in scope can transport a weight like that, therefore this shipment should be deleted from the dataset. All the shipments in the green boxes stand out because of their large gross volume in combination with relative small gross weight. They were verified manually and are considered valid shipments.

Then, data was explored on the maximum dimensions of ship units found within shipments. Remember that carriers can have restrictions on these, and therefore they are important for the planning of shipments. Table 4 provides an overview with descriptive statistics on the maximum dimensions for shipments found in the dataset.

Table 4: Maximum weight and dimensions found in shipments given all ship units

	Mean	Std.	Min.	25%	50%	75%	Max.
Max. length of ship units in a shipment (cm)	66.12	134.15	0	25.4	35.56	60	6000
Max. width of ship units in a shipment (cm)	37.25	34.86	0	20	27	40	750
Max. height of ship units in a shipment (cm)	31.27	41.13	0	12	21	35	1500

A few things stand out from Table 4. First of all, 75 percent of shipments are within either 60 cm of length, 40 cm of width or 35 cm of height. Second, for all dimensions the standard deviation is relatively high compared to the mean, which indicates high variation. This is confirmed by the maximum values observed; these maximum values can be 50 times higher than the mean. This requires further investigation, therefore the distribution of the maximum dimensions found in shipments are presented in Figure 12, 13 and 14 with a boxplot. Furthermore, the minimum values of dimensions of ship unit can be 0 cm, this implies length, width and height are not specified. This is due to the template shipments discussed earlier.

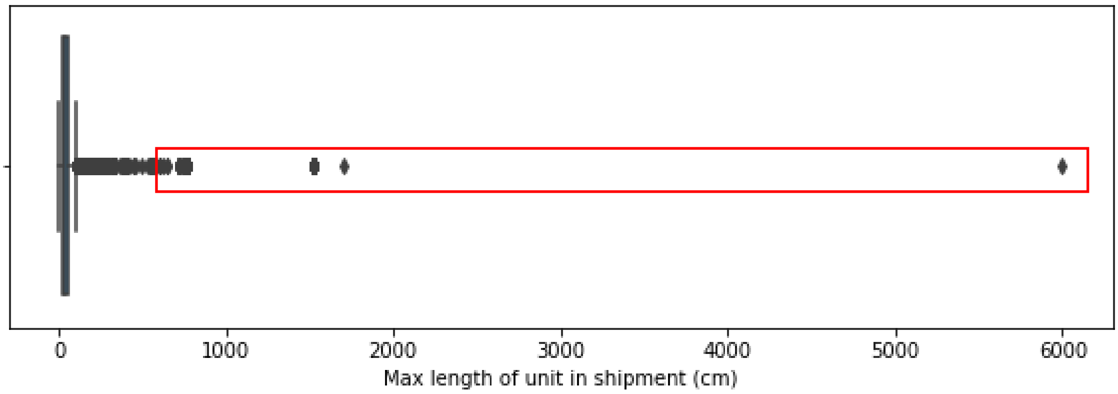

Figure 12: Boxplot maximum length ship units showing positively skewed distribution

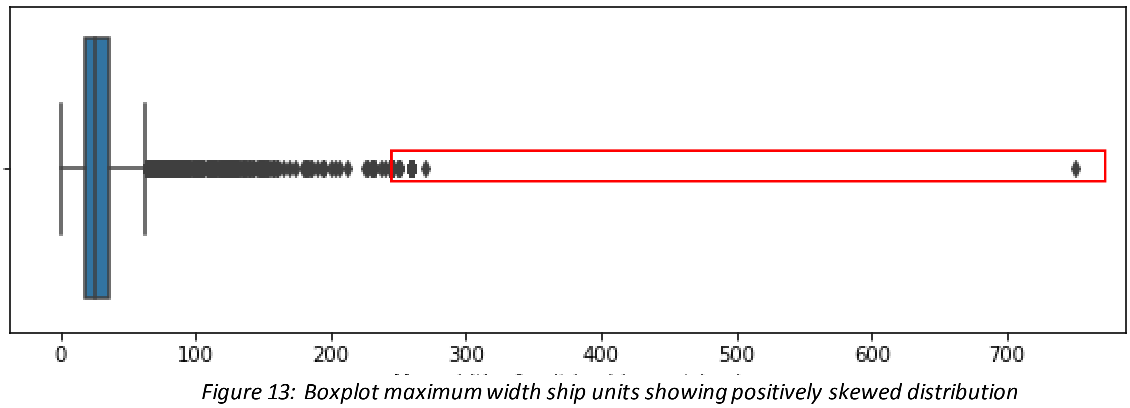

Figure 13: Boxplot maximum width ship units showing positively skewed distribution

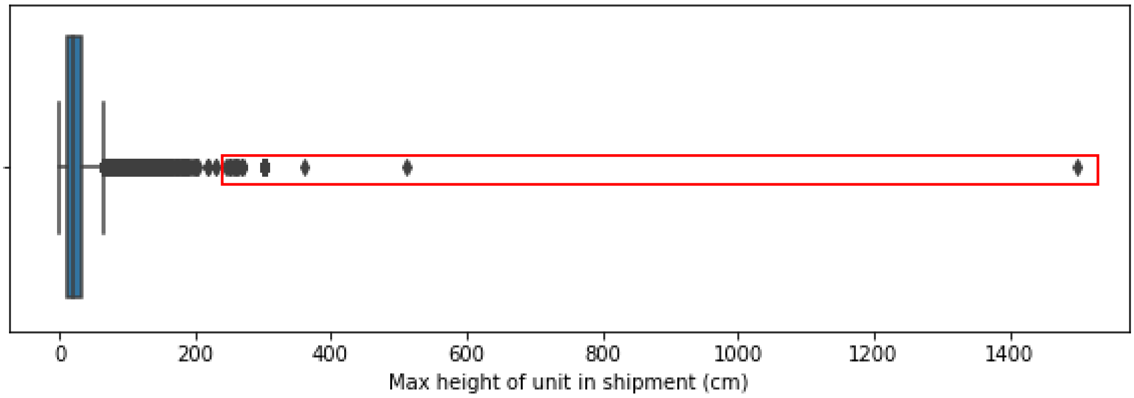

Figure 14: Boxplot maximum height ship units showing positively skewed distribution

In Figure 12, 13, and 14, can be seen that all dimensions are positively skewed, which is expected since the mean is higher than the median for all dimensions. Furthermore can be observed that most packages have small dimensions, which is expected looking at the descriptive statistics in Table 4. All shipments in the red boxes are deleted for one of the following reasons:

- *Outliers*: dimensions that do not comply with any of the carriers in scope.
- *Typing error*: typing error when entering shipments in TMS, e.g. dimensions that are not corresponding with volume of ship unit.
- Template shipments: non-real shipments.

Next, the categorical variables found in the dataset are explored. The categorical variables are visualized with bar plots, to get an idea of common categories for each variable. Starting with a bar plot of the variable "Hazardous goods", which states if a shipment contains hazardous goods, see Figure 15. It stands out that the big majority of the shipments do not contain hazardous goods.

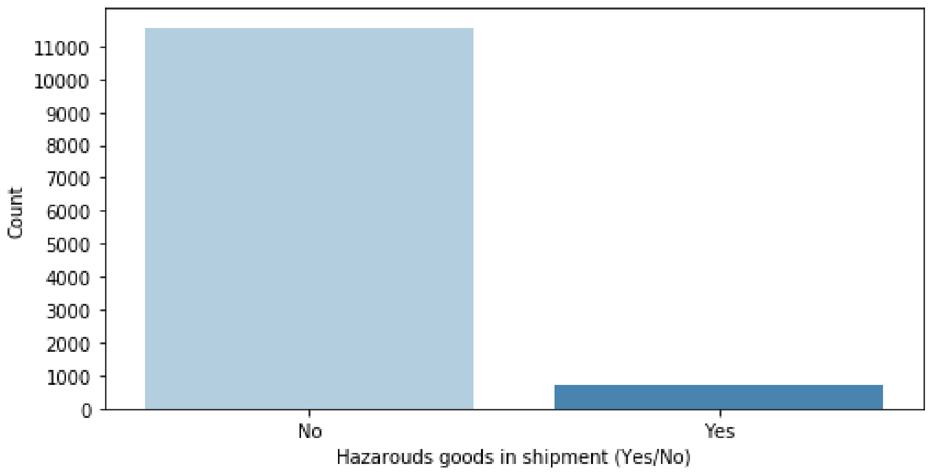

Figure 15: Bar plot hazardous shipments

Then, the 5 most common origin and destination countries found in the dataset are visualized via a bar plot in Figure 16 and Figure 17. A distinction between inbound shipments and outbound shipments has to be made, since the dataset contains both inbound (IB) and outbound (OB) shipments.

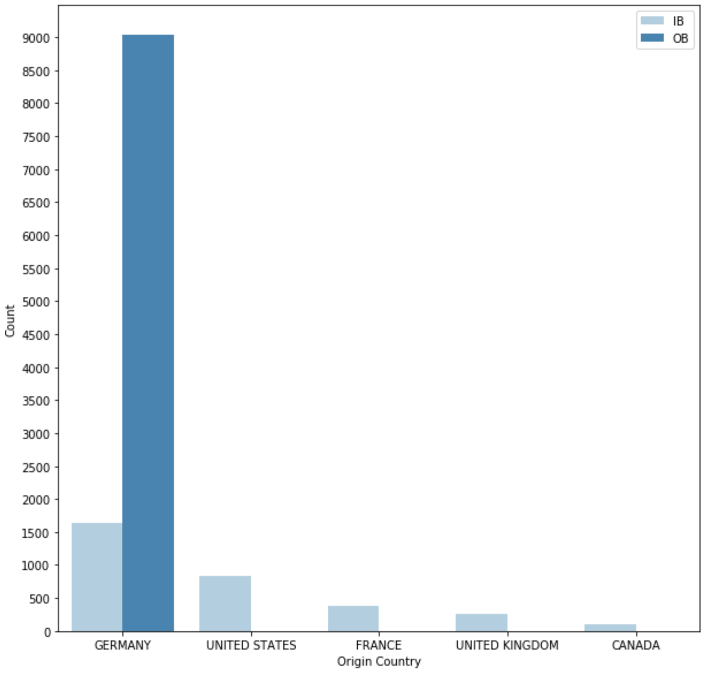

Figure 16: Bar plot origin countries

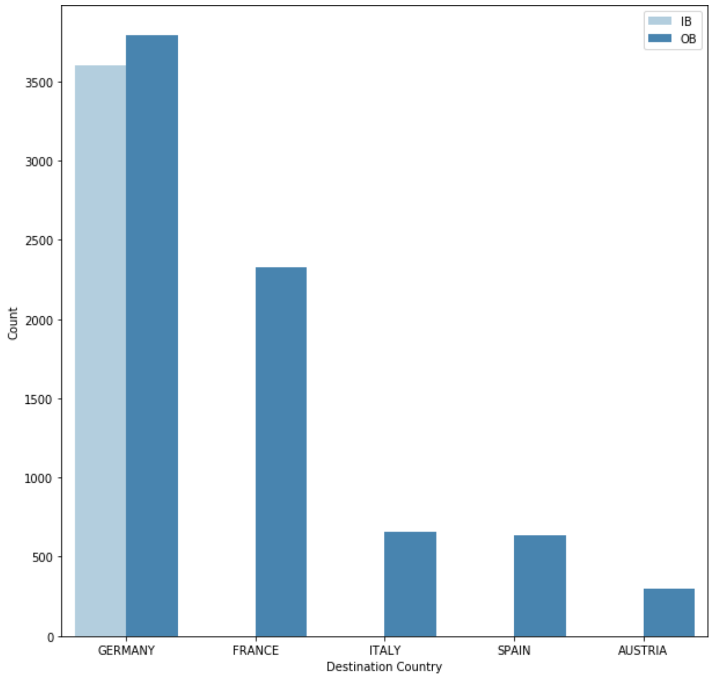

Figure 17: Bar plot destination countries

From Figure 16 and 17, can be concluded that most shipments are within Germany. Furthermore most of inbound shipments come from Germany, United States, France, United Kingdom or Canada as can be seen in Figure 16. For the outbound shipments, most shipments are going to Germany, France, Italy, Spain or Austria as can be seen in Figure 17.

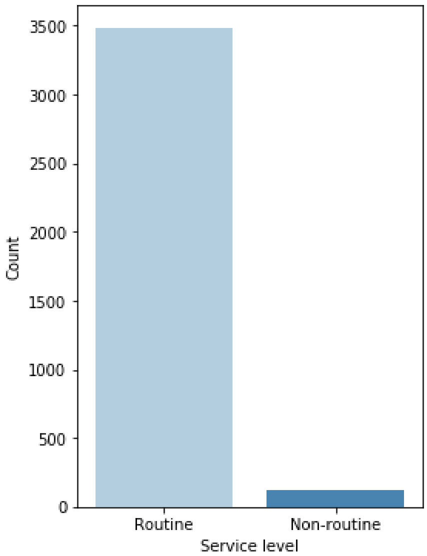

Figure 18: Bar plot showing service levels of inbound shipments

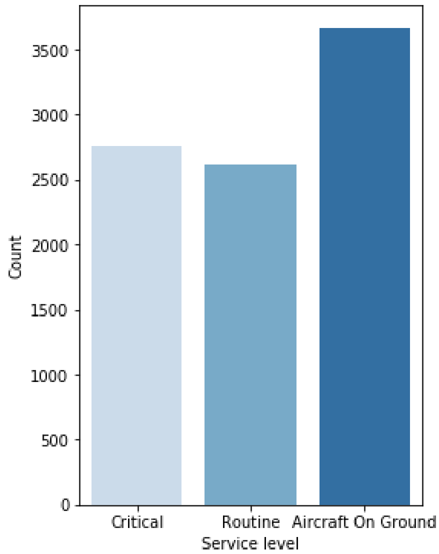

Figure 19:Bar plot showing service levels of outbound shipments

To get more insight in the service level of the shipments within the dataset a bar plot is made. Remember that different terminology is used to distinguish between service levels for inbound shipments and outbound shipments. Therefore two separate figures have been made, see Figure 18 and Figure 19. For inbound shipments, most shipments are *Routine* and only few are *Non-routine* as can be seen in Figure 18. For the outbound shipments, most shipments have an *Aircraft On Ground* service level, as can be seen in Figure 19. The rest of the shipments are balanced between *Critical* and

Routine. This implies outbound shipments are on average more urgent than inbound shipments. Remember that these service levels were defined at the end of section 4.1.2.

4.2.4 Data quality

In this section the data quality of the dataset is addressed. In section 4.2.3, already some outliers and error were identified. In this section, the dataset is checked on missing values and spelling inconsistencies. For a complete overview of the number of missing values per variables, *see Appendix F*. In total there are 23 variables with missing values. Relevant variables with missing values are discussed now.

- For "*Transport name*" and "*Transport type*"; respectively 3603 and 10249 missing values can be observed. "*Transport name*" states the service level for all outbound shipments and has missing values for all inbound shipments (3603). The same holds for "*Transport type*", where all missing values are outbound shipments (10249). In section 4.4, both variables will be combined into one variables that provides information on the service level of both inbound shipments and outbound shipments.
- For the variable "*Hazardous goods (Yes/No)*", which states if shipments contains hazardous goods, a total of 332 missing values were observed. These missing values will be handled in section 4.2.5.
- The variables: "Maximum gross weight of ship units in shipments (kg)", "Maximum gross length of ship units in shipments (kg)", "Maximum gross width of ship units in shipments (kg)" and "Maximum gross height of ship units in shipments (kg)", contain missing values. It was checked that some of these shipments could be linked to the shipments with zero gross weight. Other missing values are due to the "Template Shipments" in the dataset.

For city names and country names stated in the dataset, it was determined whether there were spelling inconsistencies. This was done using the *Levenshtein distance*, this string metric measures the similarity for two strings expressed in percentages (Okuda, Tanaka, & Kasai, 1976). Every city was checked with every other city in the dataset on similarity. Pairs of cities with a similarity of more than 70 percent were manually checked for spelling inconsistencies. In total 19 different spelling inconsistencies were identified, and will be handled in section 4.2.5. The same was done for countries in the dataset. For countries in the dataset no spelling inconsistencies were found. To summarize, the following data quality issues were identified in this section:

- Non- real shipments: "Template shipments"
- Shipments with zero gross weight.
- Outliers regarding gross volume
- Outliers regarding gross weight
- Outliers regarding maximum length of ship units contained in shipments
- Outliers regarding maximum width of ship units contained in shipments
- Outliers regarding maximum length of ship units contained in shipments
- Missing values for the variable hazardous goods (yes/no)
- Spelling inconsistencies for the names of cities in the dataset

Note: Some of these outliers are linked to each other.

4.2.5 Data Cleaning

Quality issues identified in the dataset are solved in this section. Solving these identified quality issues, is crucial to obtain reliable results with the deterministic models and to train machine learning algorithms successfully. Outliers and errors in de data can lead to irrelevant relations learned by machine learning algorithms, such learnt relations can lead to decreased performance. Therefore errors and outliers found, need to be deleted from the dataset. Deleting data samples yields less samples to train the algorithm, but increases the quality of the data. Furthermore, missing values need to be handled to ensure completeness of all variables presented to the algorithm.

In section 4.2.3 some outliers and errors were identified. Based upon this, it was decided to delete another 9 shipments. The different reasons for deletion of these shipments are summarized in Table 5.

Table 5: Overview outliers and errors found in dataset

No. shipments deleted	Reason
1	Outlier regarding gross volume of shipment
1	Outlier regarding gross weight
2	Outliers regarding height of shipments: height of these shipments do not comply with any of the carriers in scope
5	Errors in data: shipments with no gross weight and gross volume

It was discovered that the variable "*Hazardous goods (Yes/No)*" had 332 missing values. It is expected that hazard goods in a shipment are always specified, because of the potential danger and different procedures necessary to transport these shipments. Therefore these missing values were handled by imputing "No" for each missing value.

The existence of the template shipments in the dataset was discovered, these shipments need to be deleted. This resulted in a total of 278 shipments being deleted from the dataset. Furthermore, it was discovered that on some transport lanes shipments are already consolidated on a daily or weekly basis. It was decided, that all of these shipments are no longer in scope of this project. This resulted in an additional 1901 shipments being deleted from the dataset.

Finally, it was discovered in the data understanding phase that there are spelling inconsistencies for the city names in the dataset. For each city name, the most common way of spelling was used to replace all the other variants of the city name in the dataset.

4.3 Deterministic Model

In this section the deterministic model (DT) will be presented. With the DT model; shipments can be assigned to carriers according to two different scenarios:

1. Base case: carrier allocation based upon the current set of business rules
2. Optimal case: to-be situation regarding single shipments (optimal shipment allocation)

For both the base case and optimal case will be explained how shipments were allocated to carriers and how prices were calculated based upon these allocations in section 4.3.1 and 4.3.2 respectively. Prices for shipments are calculated based upon cost-rate files provided. Finally, the results obtained for both scenarios will be presented in section 4.3.3

4.3.1 Base Case

The base case is defined as the scenario where decisions are made based upon the current business rules. First, will be elaborated how the base case was constructed and then how the costs are calculated. The business rules are used to select for each shipment in the dataset a carrier. Business rules only specify the carrier that needs to be used, but not the exact service. Therefore, some assumptions about the exact carrier allocation via business rules have been made.

- Due to limited available pricing schemes for the carriers in scope, it could happen that prices for shipments with a certain combination of characteristics cannot be calculated for the base case carrier. These shipments will then be excluded from both the dataset used in the base case and optimal case.
- Remember that Carrier D can operate from several countries. Since the business rules only specify the carrier and not the exact service, it is assumed that every shipment is transported with Carrier D operating from Germany. This implies that shipments are either imported (inbound shipments) or exported (outbound shipments) with carrier D operating from Germany.
- Furthermore, carrier D can provide two different services. Based upon the lead time of shipment, a shipment will be assigned to either service level 1 or 2.
- For all other carriers in scope (except Carrier D), given the carrier, the cheapest service within this carrier was chosen as the base case carrier. This assumption was made, since the exact decision (service specific) is not specified by the business rules.

See Figure 20, for an overview of the steps set to determine the base case carrier, and to then calculate the price. Note that these steps need to be performed for all shipments available in the dataset.

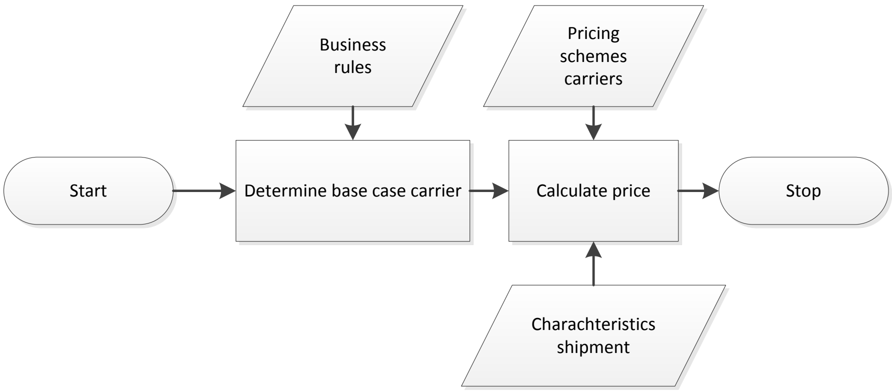

Figure 20: Steps to obtain the base carrier and corresponding price of a shipment

4.3.2 Optimal Case

The optimal case is defined as the scenario where all planning decisions made, are economically optimal, while satisfying service and capacity constraints. Constructing the optimal carrier for each shipment is an important step in this research, since this variable will be the target value used later for the machine learning models. In order to determine the optimal carrier, a DT model is developed that can choose for each shipment the cheapest option based upon the cost rates of all carriers. Due to limited available pricing schemes for the carriers in scope, it could happen that prices for shipments with a certain combination of characteristics cannot be calculated. These shipments will then be excluded from both the dataset used in both the base case and optimal case. The steps of the proposed DT model to determine the optimal carrier for a shipment are illustrated in Figure 21. Note that these steps need to be performed for all shipments available in the dataset.

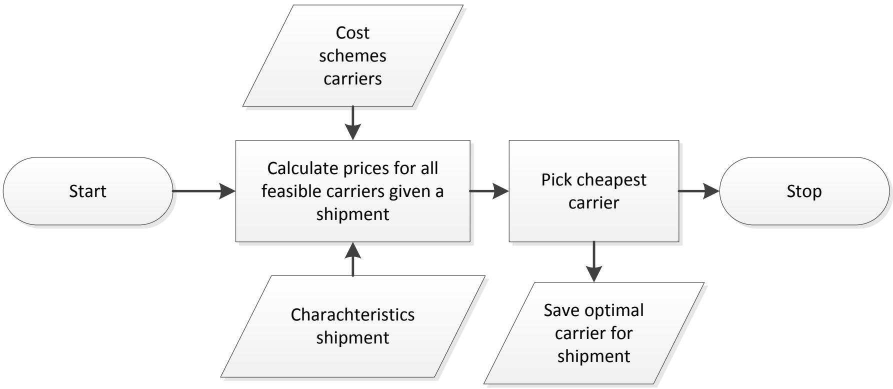

Figure 21: Steps to obtain to optimal carrier for a shipment

4.3.3 Results

Based upon the DT models presented in section 4.3.1 and section 4.3.2, both the base case and optimal case were constructed. In this section results obtained for both scenarios with the DT model will be presented. The results are based upon a total of 10107 shipments. For the base case and therefore also for the optimal case, a total of 59 shipments was excluded from the dataset due to the fact that the prices of these shipments for both scenarios could not be calculated. As mentioned before, carriers can provide different services to their customers. To keep results clear, carriers are classified into 9 labels, see Table 6.

Table 6: Overview carrier names with a description

Carrier	Description
Carrier A	Includes all shipments done with Carrier A, independent of service used
Carrier B	Includes all shipments done with Carrier B, independent of service used
Carrier C	Includes all shipments done with Carrier C, independent of service used
Carrier D1	Includes all shipments done with Carrier D inside Germany (domestic shipments)
Carrier D2	Includes all shipments done with Carrier D operating from the destination country, with service level 1
Carrier D3	Includes all shipments done with Carrier D operating from the destination country, with service level 2
Carrier D4	Includes all shipments done with Carrier D operating from the origin country, with service level 1
Carrier D5	Includes all shipments done with Carrier D operating from the origin country, with service level 2
Carrier E	Includes all shipments done with Carrier E, independent of service used

The differences in carriers used for the transportation in the base and optimal case are visualized in Figure 22.

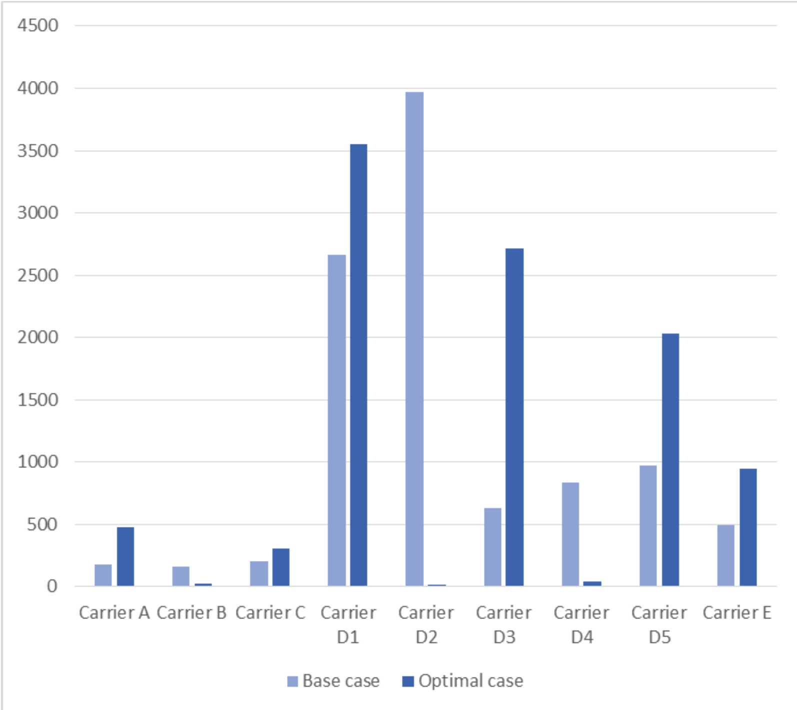

Figure 22: Difference in carrier distribution (Base case vs. Optimal case)

When looking at the differences between the carrier distribution for the base case and the optimal case, an increase in the usage of Carrier A, Carrier D1, Carrier D3, Carrier D5 and Carrier E can be observed for the optimal case. While a major decrease in the usage of Carrier D2 and Carrier D4 can be observed.

Table 7 gives an overview of the costs obtained for both scenarios. With optimal planning decisions (optimal case); 38.72% of costs can be saved over the current decision making (base case).

Table 7: Results Base Case vs. Optimal Case showing relative savings and absolute savings

Base Case	618,018 EUR
Optimal Case	387,711 EUR
Absolute savings	232,716 EUR
Relative savings	38.72%

To give an idea of the daily savings obtained with the optimal case compared to the base case for single shipments, the costs per day are visualized in Figure 23.

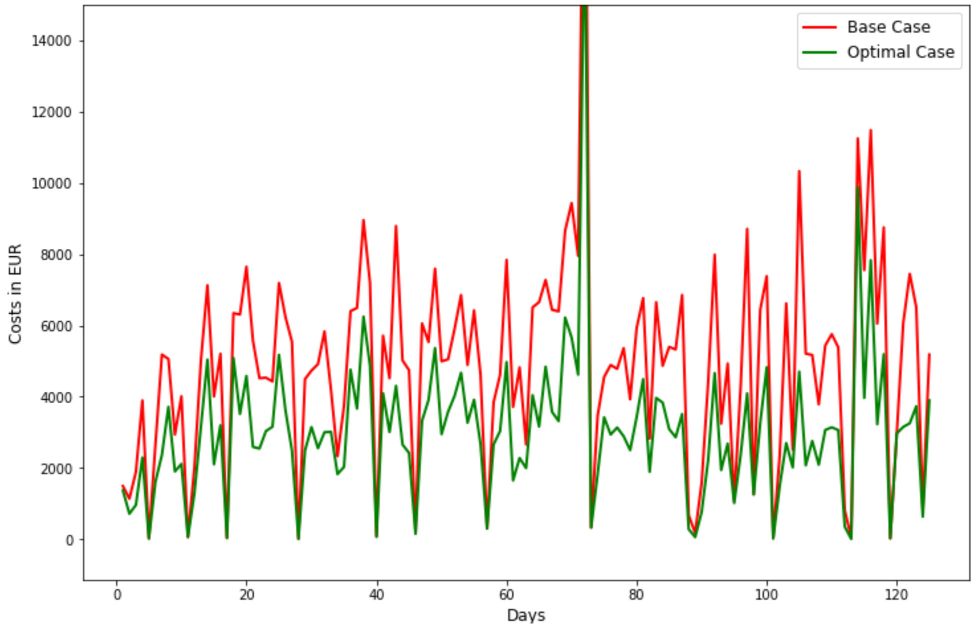

Figure 23: Costs per day (Base Case vs. Optimal Case)

4.4 Data Preparation

An overview of the data preparation steps set, are visualized in Figure 24. The whole data preparation process is now explained. Remember that in section 4.2.5 the data was already cleaned, since this was necessary to obtain valid results with the deterministic model. Next steps are explained in this section.

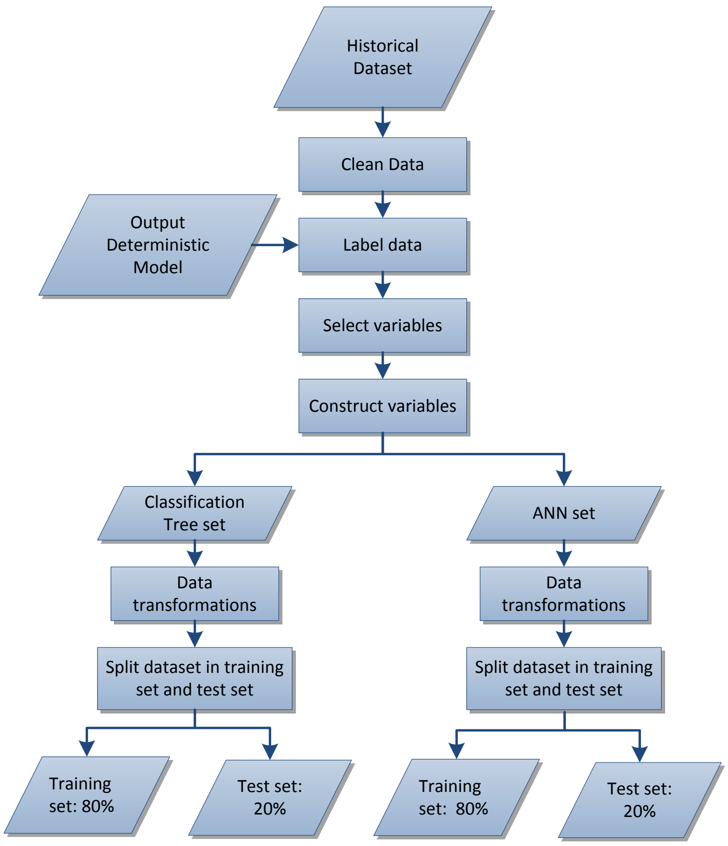

Figure 24: Overview data preparation for research cycle 1

After the data cleaning phase (section 4.2.5), it is time for the next steps in the data preparation process. The first step thereafter, is to label the historical dataset with the optimal carrier allocation obtained by the DT model. For this labelling, the same carrier labels are used as presented in Table 6. This is a crucial step, since these labels are used to train the machine learning algorithms on. The next step is to select relevant variables to train the model on. Remember that the goal of the machine learning models, is to choose the optimal carrier for each shipment. There are different variables that can influence the price of a shipment given a carrier. These variables have the strongest relation to the target variable (optimal carrier), and are therefore chosen to train the machine learning algorithms on. Based upon this, the following variables are included for modelling: *origin country, origin city, destination country, destination city, hazardous shipment* (Yes/No), *gross weight* in kg*, gross volume* in m^3*, maximum length* in cm*, maximum width* in cm*, maximum height* in cm and *maximum weight* in cm.
However, there are some variables missing in the dataset that are used to calculate prices for certain carriers. These variables are necessary to train the machine learning algorithms, and need to be constructed:

- *Lead time of transports in days*; constructed by taking the delta of the requested delivery date and the requested pickup date.
- *Transport inside Europe, Schengen area (Yes/No)*; if both destination and origin are within Europe then the transport is inside Europe, else the transport is not in Europe. This is relevant, since some carrier can only operate within Europa, while other carriers have no area restrictions.

The next step is to take care of the categorical variables in the dataset. The categorical variables need to be transformed into numbers in order for machine learning models to understand them. This can be done using various methods, like: label encoding, one-hot encoding and binary encoding (Pai & Potdar, 2017). For the *CT* model, label encoding yielded the best results in terms of *accuracy* and *precision*. Then, the *CT* set was split in 80 percent training set and 20 percent test set. The training set and test set are generated at random. For the *ANN* model, the same categorical encoding techniques were tested as just mentioned. Eventually, it was decided to use the one-hot-encoding technique for the *NN* model. Since the best results were obtained with this technique in terms of *accuracy* and *precision*.

To ensure that larger values do not overwhelm smaller values for certain variables, transforming numerical data, prior to the training process, is crucial for obtaining good results with A*NN* models (Shanker, Hu, & Hung, 1996). Different data transformations can be done like: normalization and standardization (Brownlee, 2019). With standardization the best results were obtained in terms of *accuracy* and *precision,* and therefore this technique was used to transform the numerical data. After these steps, the *NN* set was split in 80 percent training set and 20 percent test set at random.

4.5 MODELLING

In this section, first the chosen algorithms are briefly explained, then the development of the machine learning models is discussed.

4.5.1 Development Classification Tree

In our background study in Chapter 2, a brief explanation of the *CT* algorithm was already given. Remember that each node in a classification tree represents a feature in an instance to be classified, and each branch represents a value that the node can assume. Instances are classified starting at the root node and sorted based on their feature values. For the development of the *CT* model, the Scikit-learn library was used within Python ("sklearn.tree.DecisionTreeClassifier — scikit-learn 0.22.2 documentation," n.d.). The Scikit-learn library provides an optimised version of the *CART CT* algorithm. The most important step in development phase of the *CT* model, is to tune the hyper parameters of the algorithm. Within the Scikit-learn library several parameters can be adjusted to create the right classification tree for the problem at hand. The following parameters are considered in this research:

- *Criterion:* function that measures the quality of a split.
- *Splitter:* the strategy used to choose the split at each node.
- *Maximum depth:* the maximum depth of the tree.
- *Minimum samples split:* the minimum number of samples required to split an internal node.
- *Minimum samples leaf:* the minimum number of samples required to be at a leaf node.

A grid search, with 10 fold cross validation, was done to find the optimal combination of parameters for the model, see *Appendix G* for the used parameter grid. The *CT* model was trained on the variables defined in section 4.4. No significant improvements were found in performance over the default settings. Therefore, it was chosen to stick with the default settings of the Scikit-learn library.

4.5.2 Development Neural Network model

For the development of the A*NN* model the Keras library was used within Python. With the Keras library A*NN's* can be designed to one's own preference. This involves a lot of decisions to make. To understand these decisions, first will be explained how *ANN's* work and which different parameters are considered in this research. Remember Figure 4, where an example of a neural network design was illustrated. The neurons of different layers are connected with each other. For an *ANN* these connections between neurons are only feedforward, see Figure 25 (Haglin et al., 2019).

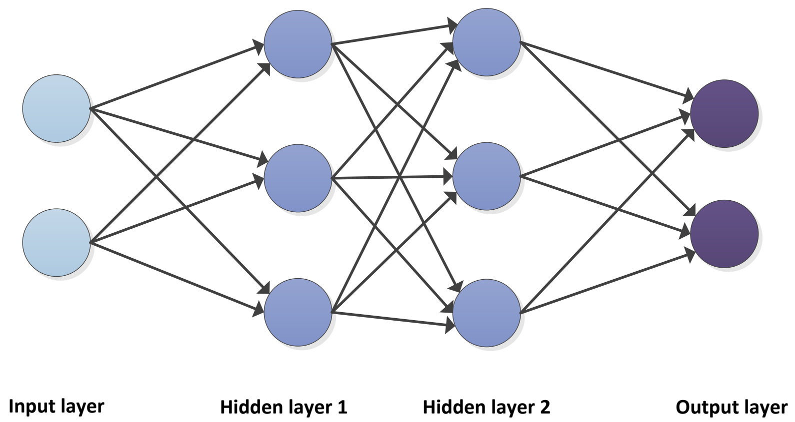

Figure 25: ANN has only feedforward connections between neurons

The connections between these neurons alter the behaviour of the network. To be able to learn the desired behaviour, an ANN makes use of connection weights and activation functions per neuron. To explain this, we will zoom in on the connection between the neurons of an input layer and a neuron in the hidden layer, see Figure 26.

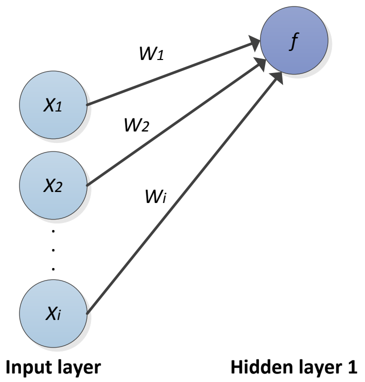

Figure 26: Neurons of input layer are connected with a neuron of the first hidden layer

X_1, X_2 until X_i are the input values contained in the neurons from the input layer, these values are feedforwarded into a neuron of the hidden layer. The hidden layer neuron computes the weighted sum of these values, see equation 4.1.

- $\sum_i x_i * w_i$ (4.1)

The output of the weighted sum goes through an activation function, which determines the final value being forwarded to neurons in the next hidden layer (Kotsiantis et al., 2006). Different activation function exist, in this research the following activations are considered for hidden layers:

- Rectified Linear Unit (ReLU): $f(x) = \max(0, x)$ (4.2)
- Exponential Linear Unit (ELU): $f(x) = \begin{cases} x, & if\ x \geq 0 \\ 1(\exp(x) - 1), & if\ x \leq 0 \end{cases}$ (4.3)
- Tanh: $f(x) = \frac{\sinh(x)}{\cosh(x)}$ (4.4)

The desired behaviour of an *ANN* is obtained by optimizing the weights in the network for all connections between neurons, this is done via an optimization algorithm. Weights are initially chosen at random and updated with each *epoch*. In Keras different optimizers exist ("Optimizers - Keras Documentation," n.d.), the following are considered in this research: *RMSprop, SGD, Adagrad and Adadelta.*
The following parameters are considered for the design of the *ANN* model:

- *Amount of neurons input layer:* equal to the amount variables the model will be trained on.
- *Number of hidden layers:* more hidden layers means more complex relations can be learned, with the risk of overfitting on the training set. More hidden layers also requires more computation time when training the model.
- *Amount of neurons per hidden layer:* more neurons means more complex relations can be learned. At the same costs just mentioned (overfitting and more computation time).
- *Activation function hidden layers:* function that transform the weighted sum of values contained from a previous layer.
- *Amount of neurons output layer:* equal to the amount of classes found in the training set.
- *Activation function output layer:* determines how the output of the model will be presented.
- *Drop-out:* determines what proportion of neurons in the layer are ignored for each layer. This can be done via a drop-out layer, which is added between the hidden layers and output layer. This layer forces the network to learn features in distributed way and also reduces overfitting.

- *Batch size:* number of trainings sample used in each iteration to update the weights of the network.
- *Epochs:* number of complete passes through the entire training set, e.g. for a training set of 100, 10 iterations with a batch size of 10 yield 1 epoch.
- *Optimizer:* used for optimizing the weights within the network.

The variables selected to train the model on, include all relevant variables stated in the data preparation phase (section 4.4). However after some trial and error, it became known that the ANN model scored significantly better (in terms of accuracy and precision) when leaving the following variables out: origin city and destination city.

The next step is to optimize all the hyper parameters of the model. Two grid searches, with 10 fold cross validation, were done to find a parameter setting that yielded good results. The first grid search was done to roughly explore which settings could work, see *Appendix H* for the parameter grid used. Then another grid search was done to fine-tune the parameter settings of the algorithm, see *Appendix I* for the parameter grid used. Table 8 gives an overview of the used parameters in the final model. The design of the final model is presented in Figure 27.

Table 8: Final parameters ANN model used in research cycle 1

Parameter name	**Value**
Neurons of input layer	85
Neurons per hidden layer	500
No. of hidden layers	2
Activation function hidden layers	ReLU
Neurons of output layer	9
Activation function output layer	Softmax
Drop-out rate	50%
Batch size	120
Epochs	100
Loss function	Categorical cross entropy
Evaluation metric	Accuracy
Optimizer	RMSprop with default Keras settings

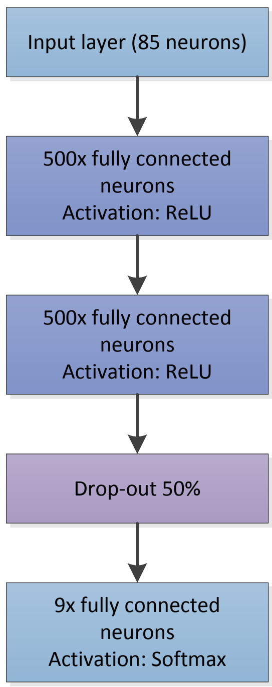

Figure 27: Structure ANN used in research cycle 1

The input layer consists of 85 neurons, the input layers are connected to two fully connected layers with each 500 neurons. Both hidden layers have *ReLU* as their activation function. Next, a dropout layer is added to the model. 0.5% was chosen as the drop-out rate for this layer. The output layer has 9 neurons, corresponding to the 9 different carriers (Table 6). The *Softmax* activation function is used to output the distributions of probability for each class (Xu & Liu, 2020). After the layers were added to the model, a loss function and optimization algorithm is set up. The categorical cross entropy is chosen, since this is a specific loss function for categorical classifications with more than two classes. *RMSprop* is chosen as the optimizer. Since Keras recommends to leave the parameters settings at default, these values are left at their default values. This results in a learning rate of 0.001 and a rho of 0.9.

4.6 EVALUATION

In this section the performance of both the *ANN* model and the *CT* model will be evaluated. Four different metrics, presented in section 3.6, will be used to evaluate performance; *accuracy and precision*, *optimality gap, relative cost savings* and *feasibility*. In section 4.6.1, the simulation approach used to obtain results is briefly described. Then, in section 4.6.2, the results are presented for both models.

4.6.1 Simulation approach

To make sure the results obtained with the machine learning model are valid, a simulation approach is adopted. During 1000 iterations; for each iteration a model is trained on a new random training set (80%) and test set (20%) to then calculate the *accuracy and precision*, *optimality gap, relative cost savings* and *feasibility*. Having the results from each iteration statistics like minimum, mean maximum and standard deviation can be presented, to give a good overview of the performances of the models developed.

4.6.2 Results

The results obtained with the *CT* and *ANN* model are presented in Table 9 and Table 10 respectively.

Table 9: Results CT model for research cycle 1

	Minimum	**Mean**	**Maximum**	**Standard deviation**	**95% Confidence interval**	
					Lower	**Upper**
Accuracy	95.9%	96.4%	98.6%	0.57%	96.28%	97.15%
Relative Costs Savings	32.1%	38.7%	45.0%	3.1%	37.9%	39.6%
Optimality Gap	1.1%	5.2%	10.6%	2.2%	4.6%	5.8%
Feasibility	98.7%	99.1%	99.5%	0.2%	99.0%	99.3%

Table 10: Results ANN model for research cycle 1

	Minimum	**Mean**	**Maximum**	**Standard deviation**	**95% Confidence interval**	
					Lower	**Upper**
Accuracy	95.8%	96.8%	99.1%	0.04%	95.3%	97.3%
Relative Costs Savings	35%	40.1%	44.0%	2%	39.2%	41.0%
Optimality Gap	1.7%	3.9%	8.5%	1.5%	3.2%	4.5%
Feasibility	99.0%	99.4%	99.7%	0.2%	99.3%	99.4%

The results presented in Table 9, show that both models can assign shipments to carriers with high accuracy. Difference in accuracy between both models is 0.4%; 96.4% on average for the *CT* model and 96.8% on average for the *ANN* model. For the precision per carrier obtained with the models see *Appendix J* and *Appendix K*. In terms of cost savings the ANN model outperforms the CT model with 40.1% relative to 38.7% cost savings on average. Both models are able to satisfy hazardous, service and capacity with on average more than 99%. On average 120 shipments every half year will get assigned to a carrier that is not able to satisfy all constraints. The different reasons for infeasible allocations made by both models are visualized in Figure 28 and Figure 29.

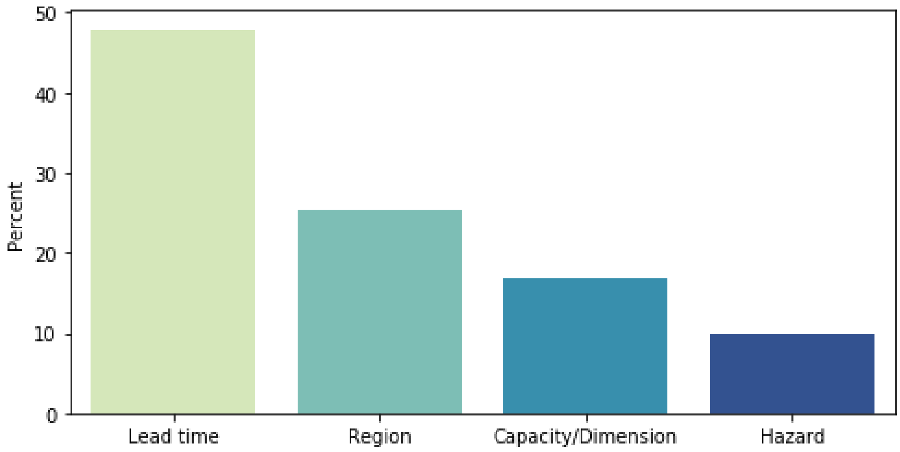

Figure 28: Reasons for non-feasible allocations made by the CT model for research cycle 1

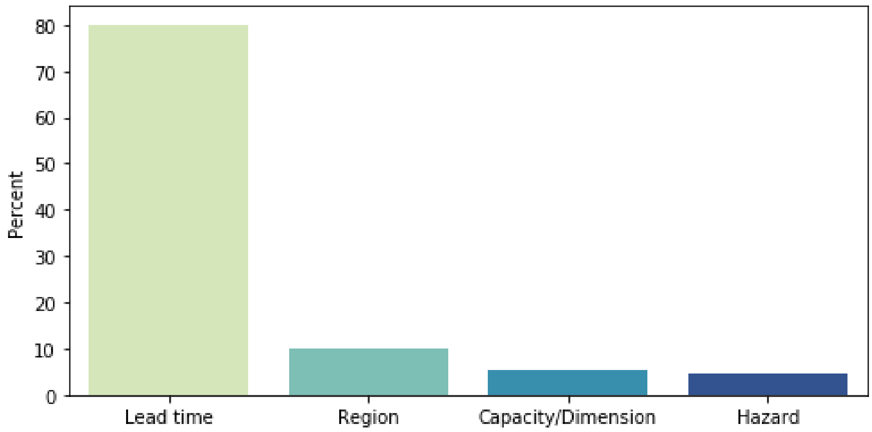

Figure 29: Reasons for non-feasible allocations made by the ANN model for research cycle 1

For both models, most infeasible allocations are due to requested delivery times not being met, followed up with region, capacity/dimension and hazard constraints. To put things in perspective; for the *CT* model around 60 shipments and for the *ANN* model around 100 shipments could be late every half year. In terms of the hazardous constraint; the *CT* model will assign around 10 shipments to a carrier that allowed to transport hazardous shipments every half year, with the *ANN* model this will be around 5 shipments every half year.

5 RESEARCH CYCLE 2: CONSOLIDATION OF SHIPMENTS

In this section, research cycle 2 regarding the consolidation of shipments is discussed. It will be investigated if business rules can be generated for the consolidation of shipments using machine learning. In section 5.1, the business understanding phase for the consolidation of shipments is discussed. Then in section 5.2, a data analysis is done to investigate if consolidating shipments would be economical interesting. In section 5.3 is described how the DT model for the consolidation of shipments is developed, and which results are obtained with the DT model per transport lane. In section 5.4 the data is prepared for modelling. In section 5.5 the development of the machine learning models will be discussed. In section 5.6 the results obtained for both scenarios with both machine learning models are presented.

5.1 BUSINESS UNDERSTANDING

The consolidation problem at hand will be defined in section 5.1.1. In section 5.1.2 the results of the literature review done on consolidation methods used in prior research will be discussed.

5.1.1 Formulation consolidation problem

At the moment, there is only weekly consolidation of shipments on certain transport lanes, as was already mentioned in section 4.2.5. Remember from Figure 7, that a transport lane is defined as a spoke within the hub and spoke transportation network. To be able to design a feasible consolidation program, the following restrictions and assumptions need to be considered:

- First of all; consolidation of shipments is only possible within a transport lane. Every transport lane is considered as a separate consolidation problem. This implies that the consolidation problem is solved for every transport lane independently. Consolidation on the same transport lane, but with a different direction is considered to be a different consolidation route. An example of possible consolidation routes is illustrated in Figure 30. The connections illustrate all different consolidations routes possible. Note that <u>not</u> for all transport lanes a consolidation program is an option, which transport lanes are considered for consolidation is investigated in section 5.2.2.

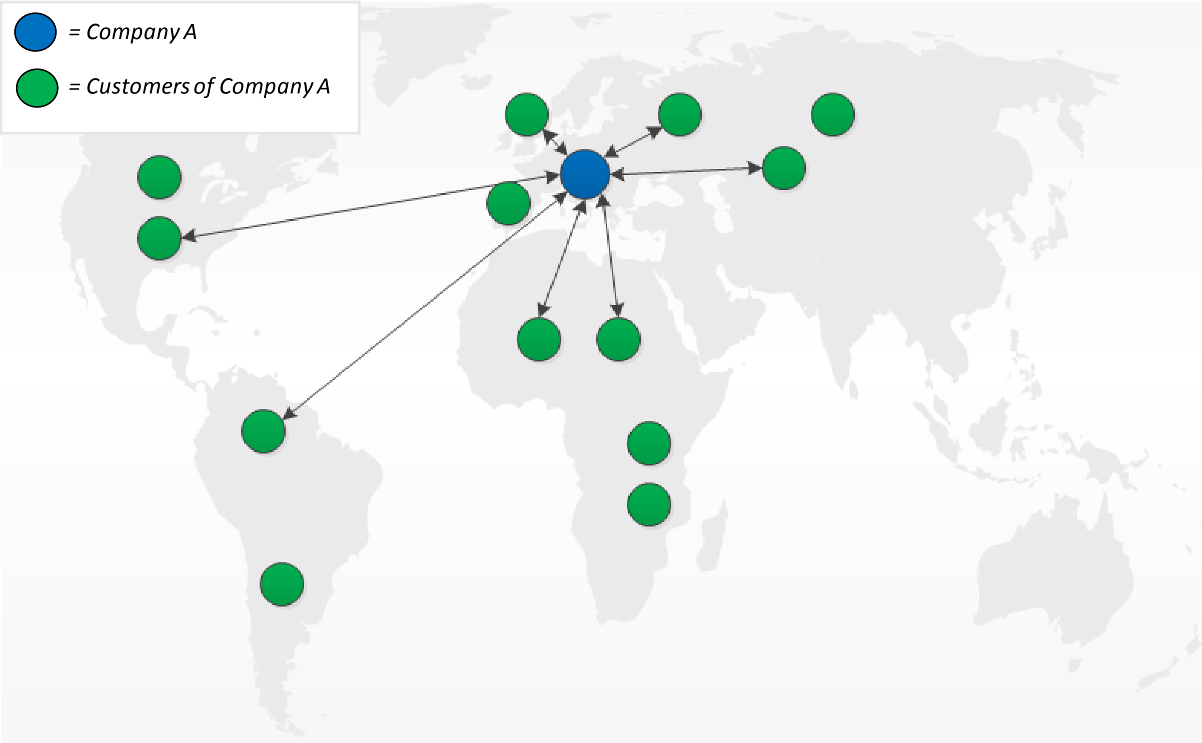

Figure 30: Possible consolidation routes

- Different addresses in the same city (within 20 kilometre) are considered to be on the same transport lane, e.g. there are cities with five different address with regularly deliveries. All of these addresses within the same city are considered to be the same place.
- Shipments with a lead time of 0 days are out of scope for consolidation, because there is no time to hold these shipments for consolidation.
- Shipments are available for consolidation at their requested pick-up day.
- For the DT model, it is assumed that all shipment for a given period are known.
- All shipments that arrive on the same day, can be consolidated at the end of a day.
- Shipments do not have to be shipped immediately, they can wait.
- Hazardous shipments and non-hazardous shipments can be transported together, as long as the chosen carrier is authorized to transport hazardous goods.
- Service constraints are hard constraints, this implies that the consolidation of shipments should not go at the expense of service. Hence, requested delivery times demanded by the customer need to be met.
- For the consolidation of shipments, the same carriers are in scope as for the first part: single shipments.
- Carrier have the same capacity and dimension constraints as applied for single shipments.

Furthermore the decision has been made to consider two different freight consolidation scenarios for this project:

1. *Same day consolidation*: every shipment is shipped at their requested pick-up date as a part of a consolidation. Different consolidation can be formed per requested pick-up date.
2. *Multiple day consolidation*: the best consolidation strategy is chosen. This implies that shipments can be stored for multiple days as long as service constraints are satisfied, with no

restriction on the waiting time for shipments. Different consolidations can be planned on the same day.

It will be interesting to analyse how those two different scenarios influence the cost savings for different transport lanes. Multiple day consolidation yields potentially more cost savings than same day consolidation. Since there are likely more shipments requested over a larger period of time, which yields more economies of scale. Yet, consolidation over multiple days causes uncertainty because of two reasons:

1. Future shipments are not known.
2. When holding a shipment, it could happen that the requested lead time can only be satisfied with a faster and more expensive carrier or worse that the shipment is late at the receiver.

5.1.2 Literature review

The goal of the literature review was to find methods to solve the consolidation problem. For each method found in literature, it was important to know the following:

- The context in which these methods were used to solve the freight consolidation problem,
- The quality of the solutions obtained with these methods, and
- The computation times of these methods.

During the literature review, it became obvious that most research regarding freight consolidation in the past 20 years, was dedicated to formulating integer optimization problems (to solve small instances, 2 papers) with the addition of a heuristic approach (to solve large instances, 7 papers). The presented methods were used in a variety of contexts; from third party logistic providers (Dondo & Mendez, 2014) and (Hanbazazah, Abril, Erkoc, & Shaikh, 2019), to air cargo consolidation problems described in (Leung, Van Hui, Wang, & Chen, 2009), (Huang & Chi, 2007a), (J.H. Bookbinder, Elhedhli, & Li, 2015) and (Wong, Leung, & Hui, 2009). These papers delivered insight in formulating integer optimization problems, e.g. how to formulate operational constraints, like: time-windows (Hanbazazah et al., 2019) and capacity constraints (Tyan, Wang, & Du, 2003), and how to incorporate non-linear costs functions into a linear mathematical formulation (Huang & Chi, 2007a). Techniques presented in these papers to solve real-life sized problems included: column generation with a branch-and-price procedure (Dondo & Mendez, 2014) and (J.H. Bookbinder et al., 2015), or a branch-and-bound procedure (Leung et al., 2009). (Hosseini, Shirazi, & Ghomi, 2014) used a Harmony search optimization algorithm, while (Wong et al., 2009) used a Tabu-search algorithm. Another interesting approach was, to divide the planning horizon into different sub-problems to be solved, and to use the solutions of all these sub-problems to obtain the final solution (Hanbazazah et al., 2019). With all these methods; good quality solutions can be obtained, that are optimal or close to optimality, solved in a reasonable amount of time. It must be noted that optimal solutions can take a lot of time for real-sized problems.

Another researched approach for the freight consolidation problem are the rules based policies, in which researchers try to define rules that form the basis for a policy. In the papers found, formulas are given that could be used in order to come up with decision variables as: optimal dispatch quantity and optimal wait time for consolidations (Çetinkaya & Bookbinder, 2003). (James H. Bookbinder, Cai, & He, 2011) propose an algorithm to test different consolidation policies based upon shipment arrivals with a Markov chain, while (Zhou, Hui, & Liang, 2011) use a simulation approach to test different quantity and time policies, based upon different arrival patterns and shipments quantities. These policies are really practical, which makes them easy to implement in real-life (Çetinkaya & Bookbinder, 2003).

Another method proposed by (Baykasoglu & Kaplanoglu, 2011), uses a reinforcement learning approach based upon the Belief-Desire-Intention structure. They claim that the model can make real-time decisions on its own, and results are of the same quality as optimization algorithms. For a complete overview of the results obtained with the literature review, see *Appendix L*.

5.2 DATA UNDERSTANDING

In this section, it will be investigated if consolidation would be feasible given the historical data and the pricing schemes of the carriers. In order for consolidation to be economical interesting, costs per kg should decrease with higher volumes, this will be investigated in section 5.2.1. Additionally, to be able to consolidate shipments, there must be enough activity on transport lanes within the same time window. This is necessary to be able to consolidate enough goods, while maintaining service constraints. An activity analysis per lane will be discussed in section 5.2.2. In section 5.2.3, will be concluded whether a consolidation program will be feasible. Also high potential lanes for consolidation will be identified based upon the results obtained.

5.2.1 Cost function analysis

See Figure 31 until Figure 36 for the cost functions of the carriers in scope.

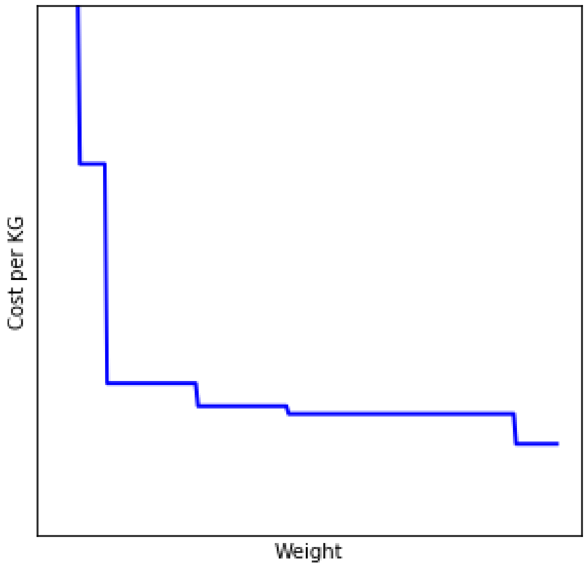

Figure 31: Cost function of Carrier A

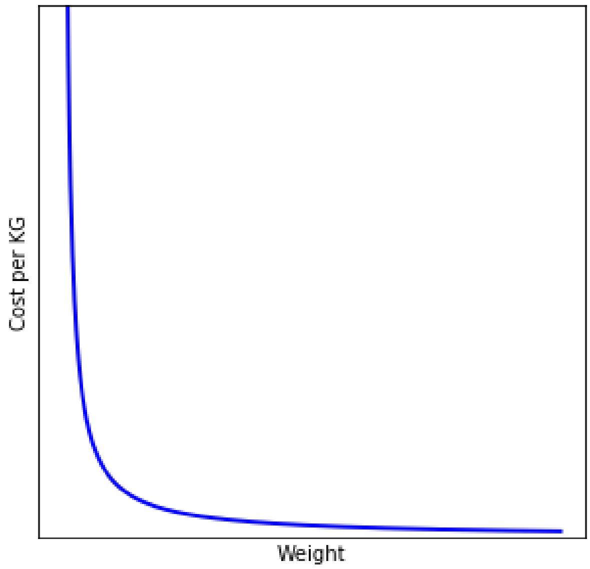

Figure 32: Cost function of Carrier B

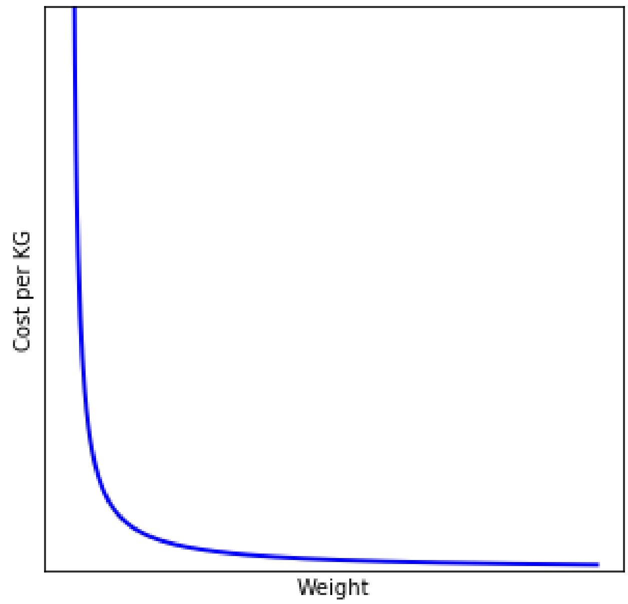

Figure 33: Cost function of Carrier C

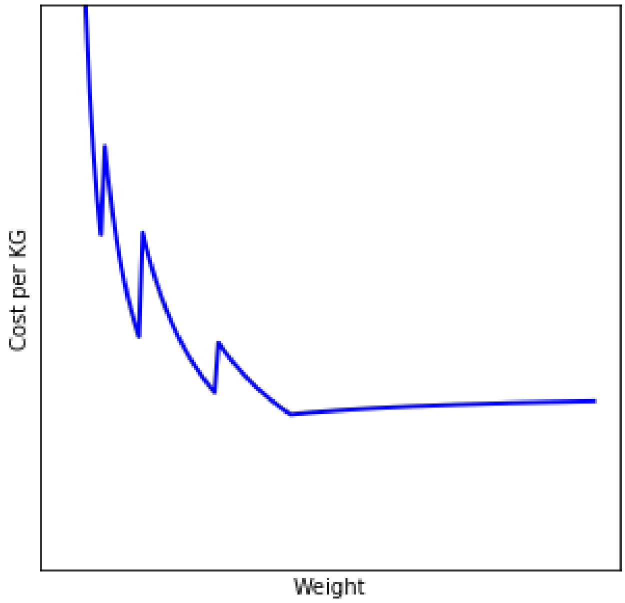

Figure 34: Cost function of Carrier D1

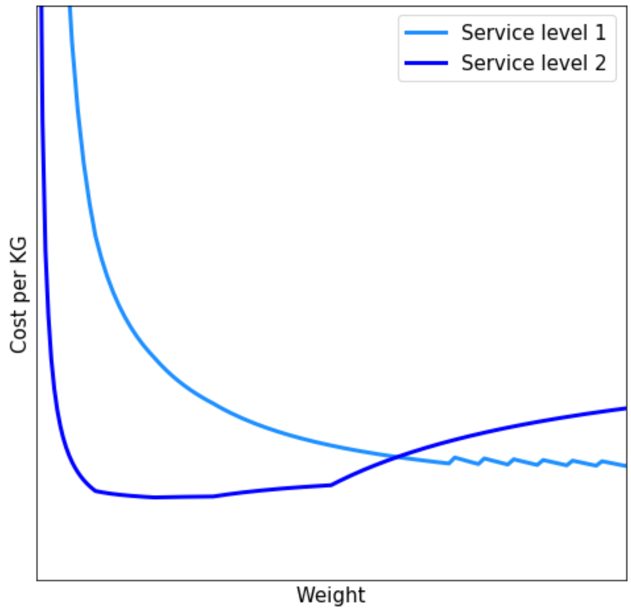

Figure 35: Cost function of Carrier D2,3,4,5

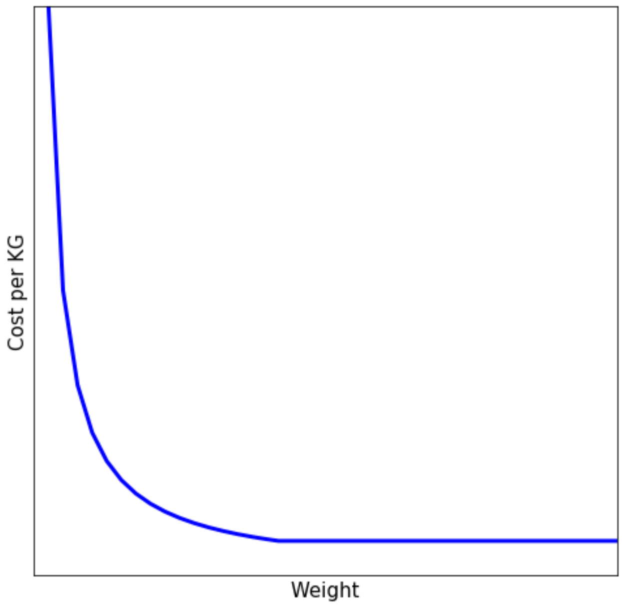

Figure 36: Cost function of Carrier E

For *Carrier A* (Figure 31), *Carrier B* (Figure 32), *Carrier C* (Figure 33) and *Carrier E* (Figure 36) prices are constantly decreasing. The cost function for *Carrier D1* is given in Figure 34, which shows that costs are not monotonously decreasing with more weights. However, until a certain weight, the general trend of the cost function is decreasing. For *Carrier D2, Carrier D3, Carrier D4 and Carrier D5* the cost function is given in Figure 35. These carriers all have a cost function with the same curve that is dependent of the service level (e.g., service level 1 and service level 2). As can be seen in Figure 35, the costs per kg for service level 1 are constantly decreasing with more weight, while the costs per kg of service level 2 only decrease until a certain weight.

5.2.2 Transport lane analysis

Given the historical dataset, there are 833 transport lanes in scope. On each of these transport lanes consolidation could be an option. After a quick examination of the total shipments on all of these transport lanes over the planning period (149 days), it was decided to only consider the 20 transport lanes with the most transport activity in the planning period for consolidation. For the remaining transport lanes, there is so little transport activity observed that consolidation will not be economical interesting. Transport activity is defined as the amount of shipments during a period of time (e.g., a day, a week etc.). Table 11 gives an overview of the transport activity analysis done for the 20 most active lanes over a period of 149 days. For every lane the percentage of active days over a period 149 days are presented. An active day is defined as a day with a least one requested pick-up. Furthermore, the mean and standard deviation of the number of shipments on active days and the requested lead times, are presented in Table 11.

Table 11: Overview transport lane analysis showing transport activity and lead times of shipments

Transport lane	**Total activity**	**% active days**	**Shipments per active day**		**Requested lead times shipments**	
			Mean	**Std.**	**Mean**	**Std.**
Lane 1	857	69%	7.9	4.7	3.8	1.6
Lane 2	617	68%	5.8	2.6	3.0	1.7
Lane 3	270	57%	3.2	2.4	4.7	2.2
Lane 4	205	57%	2.5	1.8	5.4	1.9
Lane 5	191	52%	2.4	1.6	1.5	1.2
Lane 6	173	54%	2.2	1.3	2.5	1.4
Lane 7	160	49%	2.1	1.6	6.1	2.2
Lane 8	139	46%	1.9	1.3	3.1	1.6
Lane 9	138	23%	3.8	3.1	3.1	1.7
Lane 10	137	52%	1.8	1.1	1.9	1.6
Lane 11	131	47%	1.9	1.0	1.1	0.6
Lane 12	126	23%	3.6	2.6	2.5	1.8
Lane 13	119	46%	1.8	1.0	7.4	2
Lane 14	118	42%	1.8	1.2	2.6	1.2
Lane 15	117	46%	1.7	0.9	2.7	1.8
Lane 16	116	42%	1.7	0.9	2.4	2.7
Lane 17	106	39%	1.8	1.0	1.4	0.8
Lane 18	95	38%	1.6	0.8	4.6	1.8
Lane 19	95	36%	1.8	1.2	3.4	1.6
Lane 20	93	40%	1.6	0.8	4.7	2.4

All transport lanes have on average more than 1 shipment a day. This implies that on all 20 transport lanes considered, there is at least some room for the consolidation of shipments. However, from these 20 transport lanes only 9 lanes have on average more than 2 shipments on active days, and only 5 transport lanes have on average more than 3 shipments on active days. On some transport lanes an average of more than 6 shipments can be observed. For the transport lanes averaging just above 1 shipment a day on active days, there is probably not a lot to be gained with consolidation. Especially if there is a low amount of active days for these transport lanes. Also, for most transport lanes a relative large standard deviation compared to the mean can be observed. This is implies high variation of shipments on active days.

For most transport lanes, requested lead times of shipments are on average higher than 1. Since carriers can delivery shipments within 1 day, shipments can be hold in inventory to accumulate more volume for consolidation. To determine if it is evident to let shipments wait, an analysis was done with multi day time intervals. It would be useless to analyse an interval of 10 days when 95 percent of shipments have a lead time between 4 and 6 days. To keep the maximum length of the interval to analyse realistic per lane, a rule of thumb is introduced. The maximum interval length used for analysis per lane was determined based upon the rounded mean lead time of a lane. This rule of thumb will provide a realistic timeframe for the activity analysis for each specific transport lane. To make a fair comparison with the daily activity analysis, only intervals with activity are taken along in this analysis. This approach will give a good estimate on whether waiting longer than 1 day will be pay dividends. The average activity over multiple days is visualized in Figure 37, 38, 39 and 40.

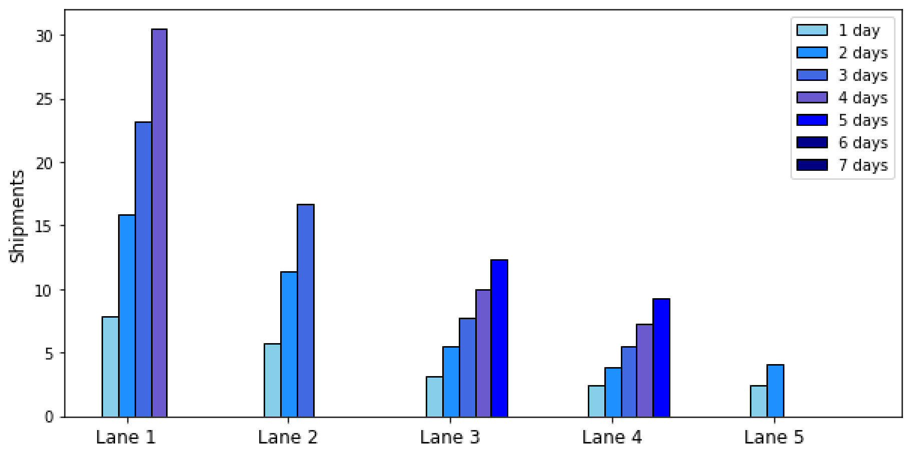

Figure 37: Mean activity per time-interval (1/4)

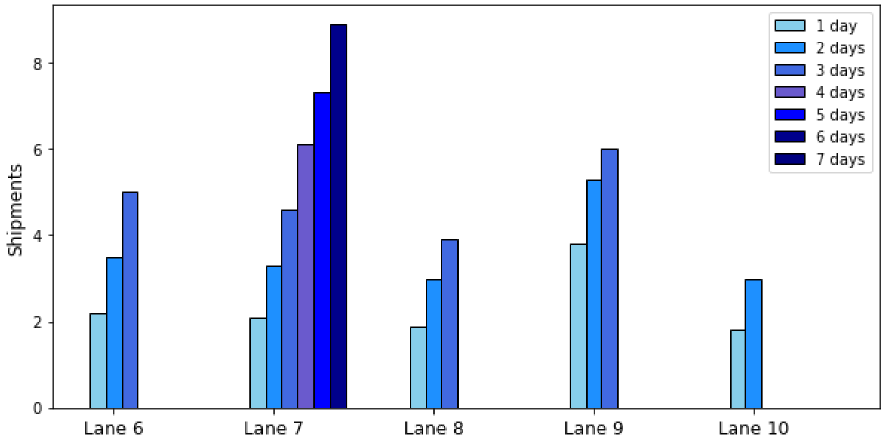

Figure 38: Mean activity per time-interval (2/4)

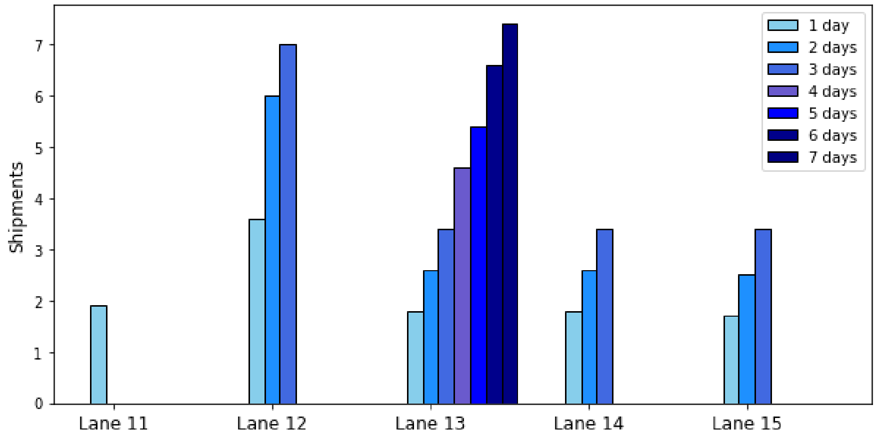

Figure 39: Mean activity per time-interval (3/4)

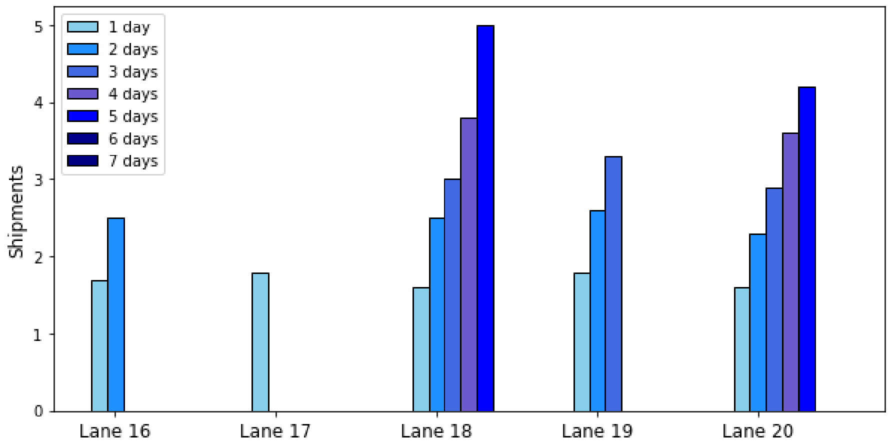

Figure 40: Mean activity per time-interval (4/4)

Looking at Figure 37 till 40, for all transport lanes it is obvious that holding shipments yields significantly more transport activity.

5.2.3 Conclusion

Looking at section 5.2.1 and section 5.2.2, it can be concluded that consolidation should be feasible, and consolidation should result in additional costs savings, since:

- for all carriers the general trend of their cost functions is decreasing, and
- enough transport activity can be observed (economies of scale).

However not for all transport lanes the same magnitude of costs savings can be expected. The more shipments are observed for transport lanes, the more cost savings can be expected on these transport lanes. Since in general, these transport lanes will also have more active days and on average more shipments on these active days/periods of days, which implies more room for combining shipments into consolidations. Based upon this, high potential transport lanes for consolidation were identified: *Lane 1, Lane 2, Lane 3, Lane 4, Lane 5, Lane 6* and *Lane 7.* On these transport lanes, it is likely that significant cost reductions can be gained by consolidating shipments.

There are some transport lanes with less activity over the whole period and a low percentage of active days, but with high average shipments on active days. This implies that most of the days, there are not much shipments on these transport lanes, but if there is activity there is going to be a period with a lot activity. Which makes these transport lanes possibly interesting for consolidation, the following lanes are identified as such: *Lane 9* and *Lane 12*

5.3 DETERMINISTIC MODEL

In this section the DT model is presented to solve the consolidation problem. This DT model will yield optimal planning decisions, which will be used to train machine learning algorithms. Based upon the literature review in section 5.2, it was decided to formulate a mixed integer optimization program to solve the freight consolidation problem. With this approach optimal solutions can be obtained (Tyan et al., 2003b), which is crucial to train good machine learning models. In section 5.3.1, the mixed integer optimization program will be formulated based upon the restrictions, conditions and assumptions mentioned in section 5.1.1. In section 5.3.2 the solving approach for the integer optimization problem is discussed. In section 5.3.3, the results obtained with the DT model will be presented.

5.3.1 Mixed integer linear program

Let S be the set of all shipments in time interval TI: $\{0,1,2...,T_{max}\}$, where T_{max} denotes the maximum length of the time interval. Each $i \in S$ has a couple of attributes. The requested pick-up date of a shipment is denoted by a_i, the requested delivery date of a shipment is denoted by d_i. Weight of a shipment is denoted by w_i. To indicate whether a shipment contains hazardous goods, the binary parameter h_i is declared:

$$h_i = \begin{cases} 1, & if \text{ shipment contains hazardous goods} \\ 0, & otherwise \end{cases}$$

Furthermore; wi_i denotes the maximum width found within a shipment, he_i denotes the maximum height found within a shipment and le_i denotes the maximum length found within a shipment.

Let K be the set of all carriers in scope. Since every transport lane is a separate consolidation problem, carriers per lane can vary, and so does K. For every $z \in K$, the following attributes are declared. The transportation time from origin to destination for a carrier is defined as lt_z. To indicate if a carrier is able to transport hazardous goods, the binary parameter hg_z is declared:

$$hg_z = \begin{cases} 1, & \mathit{if} \text{ carrier can transport hazardous goods} \\ 0, & \mathit{otherwise} \end{cases}$$

Furthermore; wc_z denotes the maximum weight carrier z can handle, vc_z denotes the maximum volume carrier z can handle, wic_z denotes the maximum width a shipment can have to be transported with carrier z, hec_z denotes the maximum height a shipment can have to be transported with carrier z and lec_z denotes the maximum length a shipment can have to be transported with carrier z.

For some carriers volumetric weight is relevant, while for other carriers volume is relevant. Remember that the volumetric weight is the volume of a shipment multiplied by a constant factor dependent of the carrier. To solve this, the parameter v_{iz} is declared which represents either the volumetric weight or volume of shipment i, based upon carrier z.

Let C be the set of consolidations formed on day t and send with carrier z. For every $j \in C, t \in TI, z \in K$, the continuous decision variable cw_{jtz} has been declared; which represents the chargeable weight of consolidation j, send on time t and transported with carrier z. The chargeable weight is defined as the greatest of the weight and volumetric weight/volume of a shipment. To be able to do so, binary decision variable y_{jtz} is declared:

$$y_{jtz} = \begin{cases} 1, & \mathit{if\ the} \text{ volume of consolidation } j, \text{sent on day } t \text{ with carrier } z \mathit{\ is\ greater\ than\ its\ weight} \\ 0, & \mathit{otherwise} \end{cases}$$

For every lane, all shipments during the time interval in the dataset need to be planned. In order to assign each shipment $i \in S$ to a consolidation, binary decision variable x_{ijtz} is introduced:

$$x_{ijtz} = \begin{cases} 1, & \mathit{if} \text{ shipment } i \text{ is in consolidation } j, \text{sent on day } t \text{ with carrier } z \\ 0, & \mathit{otherwise} \end{cases}$$

Since the cost/kg structure is not always decreasing for all carriers. It is necessary to distinguish between consolidations sent on the same day, with the same carrier, by means of index j.

The transportation costs can be calculated by using the chargeable weight cw_{jtz}, of each consolidation formed, as input for the cost function of that particular carrier. However, this would lead to a non-linear cost function that cannot be solved using linear programming. To formulate a linear cost function, the cost functions need to be transformed into a piece-wise linear cost function (Huang & Chi, 2007b). This will result in a mixed integer linear program (MILP), which is solvable. To be able to do so, a new binary decision variable needs to be added. Before explaining the new decision variable and the corresponding constraints, a closer look on the cost functions of the carriers is required. In order to obtain piece-wise linear cost functions, for every cost function fixed breakpoints should be defined. This is best explained on the basis of an example cost function given in Figure 41.

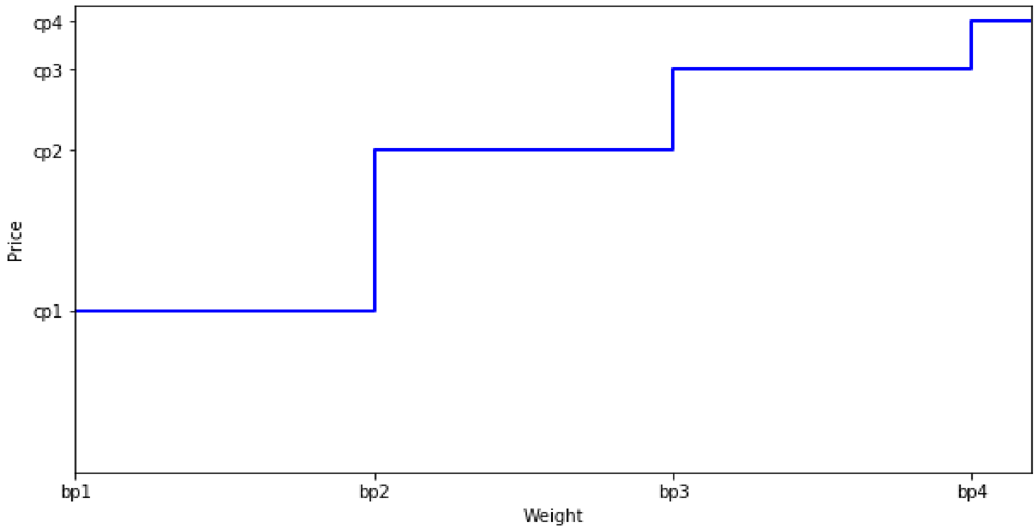

Figure 41: Breakpoints of cost function

All breakpoints bp_1 , …, bp_m correspond with a fixed price cp_1 , …, cp_m, as can be seen in Figure 41. Where bp_m = B, with B chosen sufficiently large, such that the chargeable weight of a consolidation will never exceed bp_m. For all carriers in scope, the cost function can be presented by weight breakpoints and their corresponding prices. To be able to calculate the price of a consolidation as a linear combination of the weight breakpoints and their corresponding prices; the new binary decision variable p_{njtz} is introduced:

$$p_{njtz} = \begin{cases} 1, & if \text{ breakpoint n of costfunction is chosen for consolidation } j, \text{sent on day } t \text{ with carrier } z \\ 0, & otherwise \end{cases}$$

The following integer problem is formulated to solve the consolidation problem per transport lane.

Min $\quad \sum_{n=1}^{m} \sum_{j \in C}^{N} \sum_{t \in TI} \sum_{z \in K} p_{njtz} \; * cp_{nz}$ (5.1)

s.t. $\quad \sum_{j \in C}^{N} \sum_{t \in TI} \sum_{z \in K} x_{ijtz} \; = 1 \qquad \forall\, i \; \in S$ (5.2)

$\sum_{j \in C}^{N} \sum_{t \in TI} x_{ijtz} * h_i \le hg_z \qquad \forall\, i \; \in S, \; \forall\, z \; \in K$ (5.3)

$\sum_{j \in C}^{N} \sum_{t \in TI} x_{ijtz} * wi_i \le wic_z \qquad \forall\, i \; \in S, \; \forall\, z \; \in K$ (5.4)

$\sum_{j \in C}^{N} \sum_{t \in TI} x_{ijtz} * he_i \le hic_z \qquad \forall\, i \; \in S, \; \forall\, z \; \in K$ (5.5)

$\sum_{j \in C}^{N} \sum_{t \in TI} x_{ijtz} * le_i \le lec_z \qquad \forall\, i \; \in S, \; \forall\, z \; \in K$ (5.6)

$\sum_{i \in S} x_{ijtz} \; * w_i \le wc_z \qquad \forall\, j \; \in C, \forall\, t \; \in TI, \forall\, z \; \in K$ (5.7)

$\sum_{i \in S} x_{ijtz} \; * v_{iz} \le vc_z \qquad \forall\, j \; \in C, \forall\, t \; \in TI, \forall\, z \; \in K$ (5.8)

$\sum_{i \in S} x_{ijtz} \; * w_i \le cw_{jtz} \qquad \forall\, j \; \in C, \forall\, t \; \in TI, \forall\, z \; \in K$ (5.9)

$\sum_{i \in S} x_{ijtz} \; * v_{iz} \le cw_{jtz} \qquad \forall\, j \; \in C, \forall\, t \; \in TI, \forall\, z \; \in K$ (5.10)

$\sum_{i \in S} x_{ijtz} \; * w_i + M * y_{jtz} \ge cw_{jtz} \qquad \forall\, j \; \in C, \forall\, t \; \in TI, \forall\, z \; \in K$ (5.11)

$\sum_{i \in S} x_{ijtz} \; * v_{iz} + M * \left(1 - y_{jtz}\right) \ge cw_{jtz} \qquad \forall\, j \; \in C, \forall\, t \; \in TI, \forall\, z \; \in K$ (5.12)

$$a_i \leq \sum_{j \in C} \sum_{z \in K} \sum_{t \in TI} x_{ijtz} * t \leq d_i - lt_z \qquad \forall\, i \in S \qquad (5.13)$$

$$\sum_{n=1}^{m} p_{njtz} * bp_{nz} \geq cw_{jtz} \qquad \forall\, j \in C, \forall\, t \in TI, \forall\, z \in K \qquad (5.14)$$

$$\sum_{n=1}^{m} p_{njtz} = 1 \qquad \forall\, j \in C, \forall\, t \in TI, \forall\, z \in K \qquad (5.15)$$

$$x_{ijtz} \in \{0,1\},\ y_{jtz} \in \{0,1\},\ cw_{jtz} \geq 0,\ p_{njtz} \in \{0,1\} \qquad \forall\, i \in S,$$

$$\forall\, j \in C, \forall\, t \in TI, \forall\, z \in K \qquad (5.16)$$

(5.1) the objective function ensures that the total costs of all consolidations are minimized, where cp_{nz} is the fixed price coupled to breakpoint bp_{nz}, which corresponds to the chargeable weight of the consolidation formed. (5.2) ensures that all shipments are assigned to only one consolidation. (5.3) makes sure that hazardous shipments are transported with carriers that are able to do so. (5.4) to (5.6) inclusive, ensure that dimensions are in specification with the chosen carrier. (5.7) and (5.8) ensure that capacity constraints regarding weight and volume of a consolidation are not violated for the chosen carrier. (5.9) to (5.12) inclusive, make sure that the chargeable weight of a consolidation is defined as the greatest of weight and volumetric weight. Where big M is chosen as the maximum weight observed in one week on a certain transport lane. (5.13) makes sure that shipments are not sent before their requested pick-up date and that shipments arrive before the requested delivery data (Wong et al., 2009). (5.14) makes sure that the chargeable weight of every consolidation send on day t, corresponds with the right breakpoint of the cost function from carrier z. (5.15) ensures that every consolidation gets assigned to only one breakpoint

With this MILP there are no constraints on the maximum waiting time for an order to be dispatched (multiple day consolidation). To be able to the run the same day consolidation scenario, the model must be extended with the following constraint:

$$\sum_{j \in C}^{N} \sum_{t \in TI} \sum_{z \in K} x_{ijtz} * t \leq a_i + fwt \qquad \forall\, i \in S \qquad (5.17)$$

(5.17) ensures that shipments can only be hold for a certain time (days) before they are dispatched. With *fwt,* being the maximum waiting time for dispatching transports. For same day consolidation the *fwt* for shipments is set equal to 0.

5.3.2 Solving approach

The final MILP presented in section 5.4.1, can be classified as a problem that is proved to be NP-hard (Li, Tao, & Wang, 2009). Hence, it does not seem wise to try to solve the program exact for the whole planning period (149 days). Based upon (Hanbazazah et al., 2019), it was decided to divide the whole planning horizon into different sub-problems. This implies that instead of solving the whole planning horizon at once, multiple smaller time-intervals are determined and solved independently of each other. By doing so, smaller sub-problems are created that are faster and easier to solve for an optimizer. The procedure for creating and solving these sub-problems, is described in Figure 42.

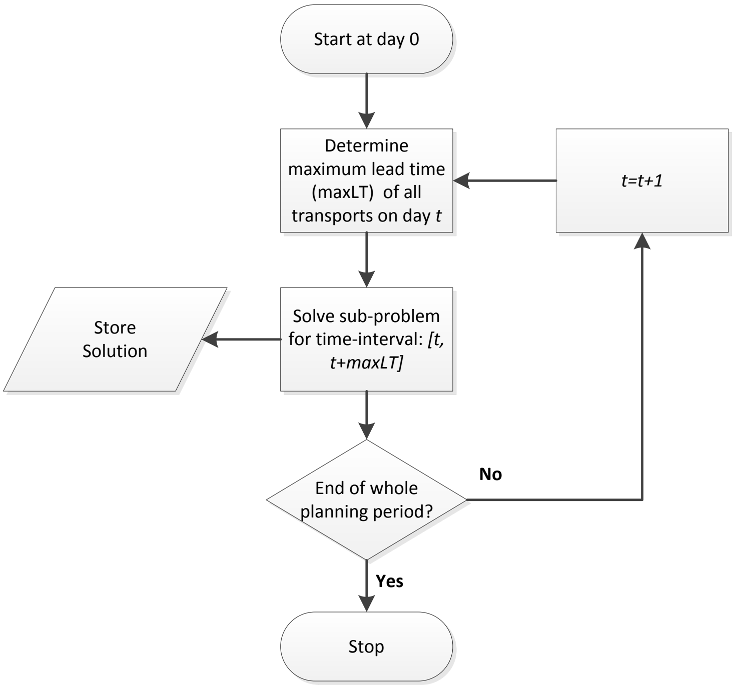

Figure 42: Procedure creating sub-problems

For every day in the planning horizon, a time-interval is determined based upon the maximum lead time of shipments on day *t*. Every iteration yields a sub-problem, this sub-problem is solved and the solution is stored. All of these sub-problems are solved with the final MILP presented in section 5.4.1. After solving all these sub-problems, for each sub-problem the optimal consolidation planning is obtained. However, solutions for different sub-problems can contain the same shipments. Eventually, every transport can be assigned to only one consolidation. Therefore, from all the consolidations generated, the most cost-effective mix of consolidations needs to be chosen, to obtain the final solution. This problem can be formulated as a Set Partitioning Problem (Dondo & Mendez, 2014), where every set is a feasible consolidation. The Set Partitioning Problem will be formulated in the next section.

Let C be the set of all consolidations obtained by solving all sub-problems in the whole planning horizon. Every $j \in C$ has the following attribute: p_j, which represents the total cost of consolidation j. Let S be the set of all shipments (i) contained in the whole planning horizon. To link shipments to consolidations, binary parameter cc_{ij} is introduced:

$$cc_{ij} = \begin{cases} 1, & if\ shipment\ i\ is\ contained\ in\ consolidation\ j \\ 0, & otherwise \end{cases}$$

In order to know which mix of consolidations is optimal; binary decision variable x_j is defined:

$$x_j = \begin{cases} 1, & if\ consolidation\ j\ is\ included\ in\ the\ final\ solution \\ 0, & otherwise \end{cases}$$

The *Set Partitioning Problem* can be formulated as:

$$\text{Min} \quad \sum_{j \in C} p_j * x_j \qquad \forall\ i\ \in S \qquad (5.18)$$

$$\text{s.t} \quad \sum_{j \in C} x_j * cc_{ij} = 1 \qquad \forall\ i\ \in \text{S} \qquad (5.19)$$

$$x_j\ \in \{0,1\},\ cc_{ij}\ \in \{0,1\} \qquad \forall\ j\ \in\ C,\ \forall\ i\ \in S \qquad (5.20)$$

(5.18) is the objective function that minimizes the total costs for all consolidation chosen in the final solution. (5.19) makes sure that every shipment is assigned to exact one consolidation.

5.3.3 Results

For all 20 lanes, the cost savings obtained with consolidation DT model will be presented for the two scenarios defined, see Table 12. To indicate additional savings obtained with consolidation, also the savings obtained with the single shipment DT model are presented. All savings presented are calculated relative to the single shipments base case (current decision making).

Table 12: Overview cost savings with DT models

Transport Lane	Single shipments	Same Day Consolidation (*fwt*=0)	Multiple Day Consolidation Consolidation (*fwt*=∞)
Lane 1	31.7%	68.9%	73.9%
Lane 2	11.4%	40.2%	47.3%
Lane 3	72.7%	84.1%	87.6%
Lane 4	65.4%	71.9%	76.3%
Lane 5	66.4%	74.7%	74.7%
Lane 6	7.8%	18.5%	36.6%
Lane 7	56.9%	62.7%	67.4%
Lane 8	65.1%	70.7%	77.1%
Lane 9	28.1%	58.4%	62.8%
Lane 10	20.3%	28.2%	30.1%
Lane 11	38.6%	47.6%	47.6%
Lane 12	56.3%	75.2%	76.7%
Lane 13	70.0%	75.4%	81.1%
Lane 14	0%	15.3%	15.3%
Lane 15	28.5%	40.2%	50.4%
Lane 16	7.1%	14.2%	16.3%
Lane 17	0%	33%	33%
Lane 18	59.4%	65.6%	71.5%
Lane 19	20.7%	37.1%	46.8%
Lane 20	45.0%	51.4%	56.9%

For both consolidation scenarios, additional cost savings are obtained with respect to the optimal single shipment planning. As expected from the analysis done in section 5.2, waiting longer before dispatching shipments yields more cost savings in most cases. The difference for most transport lanes is no more than 10% when comparing same day consolidation with multiple day consolidation. There are some transport lanes with a 20% difference, while on other transport lanes there is no difference between both scenarios. For the five biggest transport lanes, there is no more than 5% difference between same day consolidation and multiple day consolidation.

5.4 DATA PREPARATION

In this section the data is prepared for modelling in research cycle 2. To give a brief overview of the steps taken, Figure 43 is made. In section 5.4.1 will be explained why additional data needed to be created, and how this additional data was created. In section 5.4.2 the optimal consolidation planning for the two scenarios will be labelled to the dataset. In section 5.4.3 additional variables are constructed, and the final set of variables is selected for both consolidation scenarios. In section 5.4.4 the data is transformed to fit the machine learning algorithms. Finally, the data is split into a training and test set, this is elaborated in section 5.4.5.

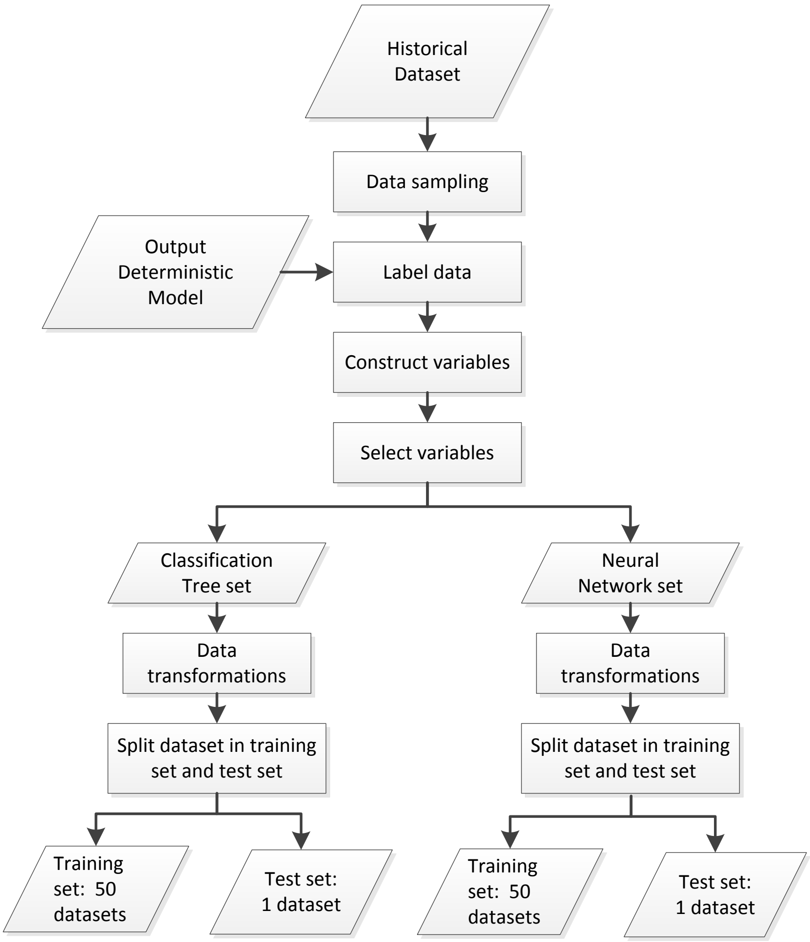

Figure 43: Overview data preparation for research cycle 2

5.4.1 Data sampling

To be able to train a machine learning to make the right consolidation decisions, enough data should be available. Considering the fact that transport lanes have a maximum of 850 shipments in the planning period of 149 days, this is not deemed enough for sufficiently training the algorithms. Therefore, it is required to create additional data, for all transport lanes where a consolidation program is considered.

Whereas the carrier selection for single shipments solely relies on the characteristics of a shipment. The planning of consolidations is more influenced by the mix and total volume of shipments in a certain period of time. It was decided to create for every day with activity found in the historical dataset, a new mix of shipments based upon shipments found in the historical dataset. Since patterns on weekdays regarding daily activity are expected, it was decided to only mix shipments on Mondays with Mondays and Tuesdays with Tuesdays etcetera. A bootstrap approach was adopted, to create extra datasets (Williams, 2013). The exact bootstrap approach used, is explained in *Appendix K.* With this procedure, for all identified active days on a transport lane a different mix of shipments will be selected, creating an extra dataset with each iteration. It was decided to sample 50 additional datasets for every transport lane considered for consolidation.

5.4.2 Data labelling

In order for the machine learning models to learn optimal consolidation planning decisions, all shipments for a given transport lane should be labelled with optimal consolidation planning decisions. This involves:

1. *When* to schedule a shipment for shipping (on which day)?
2. *Which* other shipments should a shipment be combined with?
3. *Which* carrier should ship the consolidated shipments?

It is important to come up with a label structure that contains all of this information, while remaining easy to interpret. To be able to do so, first some more understanding of the consolidation planning obtained with the DT models is required. To be more precise, the following three phenomes need to be investigated:

- *Ph1*: are shipments dispatched at a different date than their requested pick-up date? If yes the label must provide information on how many days a shipment is held before dispatching, otherwise this information is not relevant.
- *Ph2*: are different carriers used for shipping consolidations? If yes, the label must provide information on which carriers was used to transport the consolidation of shipments, otherwise this information is not relevant.
- *Ph3*: are different consolidations send with the same carrier on the same day? If yes, the labels must distinguish between consolidations send on the same day with the same carrier, otherwise this information is not relevant.

The three phenomes listed above are analysed, an overview of this analysis is presented in *Appendix N*. From the analysis, it was decided that a label should distinguish between *phenome 2* and *phenome 3* for same day consolidation:

- Carrier used to transport consolidation (*phenome 2*)
- Distinguish between consolidations send with the same carrier on the same day (*phenome 3*)

If phenome 3 is observed on a transport lane, it happens at most 2 times over 149 days, that two different consolidations are send with the same carrier on the same day. Since the loss is only small (around 10 Euro) when these consolidations on the same day are put together, it is chosen to only distinguish between carriers for same day consolidation. Given the optimal same day consolidation planning determined by the DT model, labels for every shipment in the dataset can be extracted. Since it is decided to only distinguish between carriers. For same day consolidation, every shipment contained in a consolidation that is shipped with carrier *z*, gets labelled with carrier *z*. To illustrate the label extraction from the output of the DT model, a simple example is given in Figure 44.

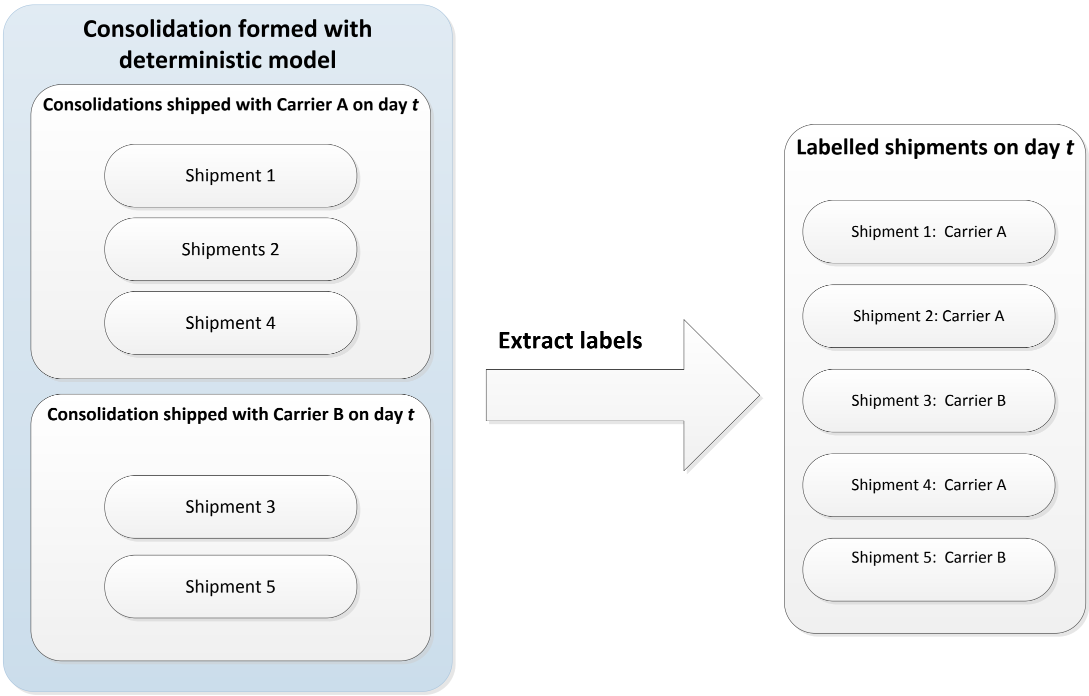

Figure 44: Label extraction same day consolidation

This also works the other way around: from the labels predicted by the machine learning models, the consolidation planning can be determined. To illustrate this, Figure 45 has been made.

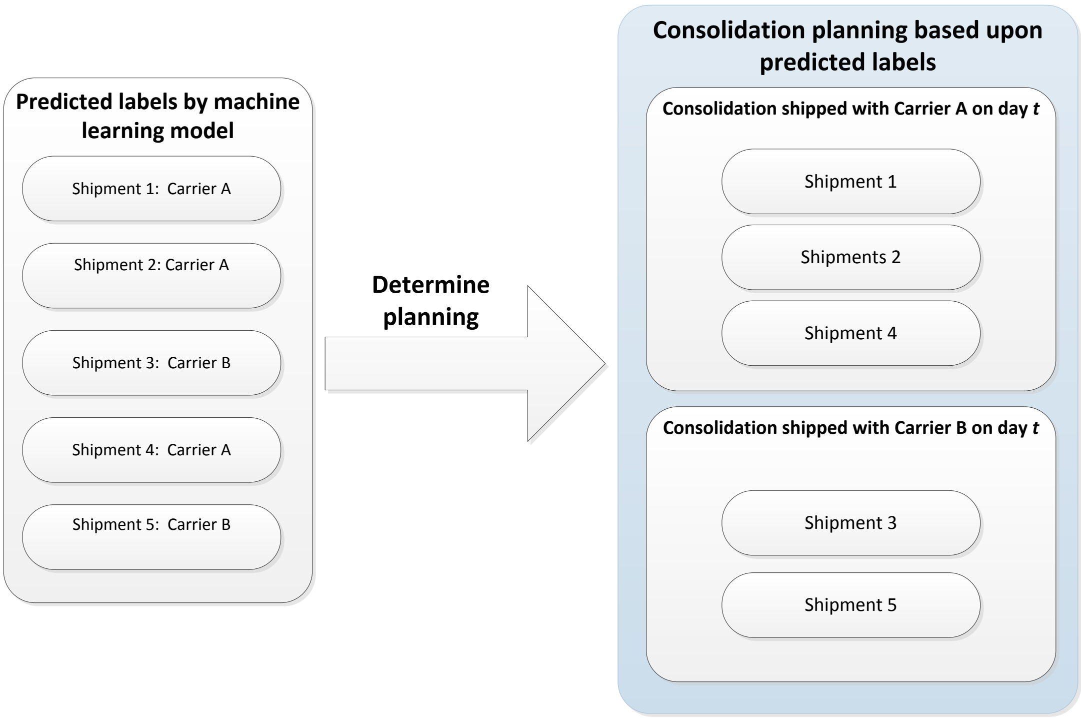

Figure 45: Consolidation planning for same day consolidation based upon predicted labels

For consolidation over multiple days, it was decided that a label should distinguish between *phenome 1* and *phenome 2:*

- the day a consolidation is shipped (*phenome 1*), and
- with which carrier the consolidation is shipped (*phenome 2*).

This implies that a label consist of two parts, namely: information about which carrier is used and information about the day a shipment is shipped within a consolidation. Labels can be structured as following: waitingTime_carrier, e.g. 2_CarrierA. This label implies that a shipment is supposed to be shipped in two days from its requested pick-up date in a consolidation with all shipments scheduled on that particular day, with Carrier A. Given the optimal multiple day consolidation planning, these labels can be extracted for all shipments in the dataset. To illustrate this, a simple example is given in Figure 46.

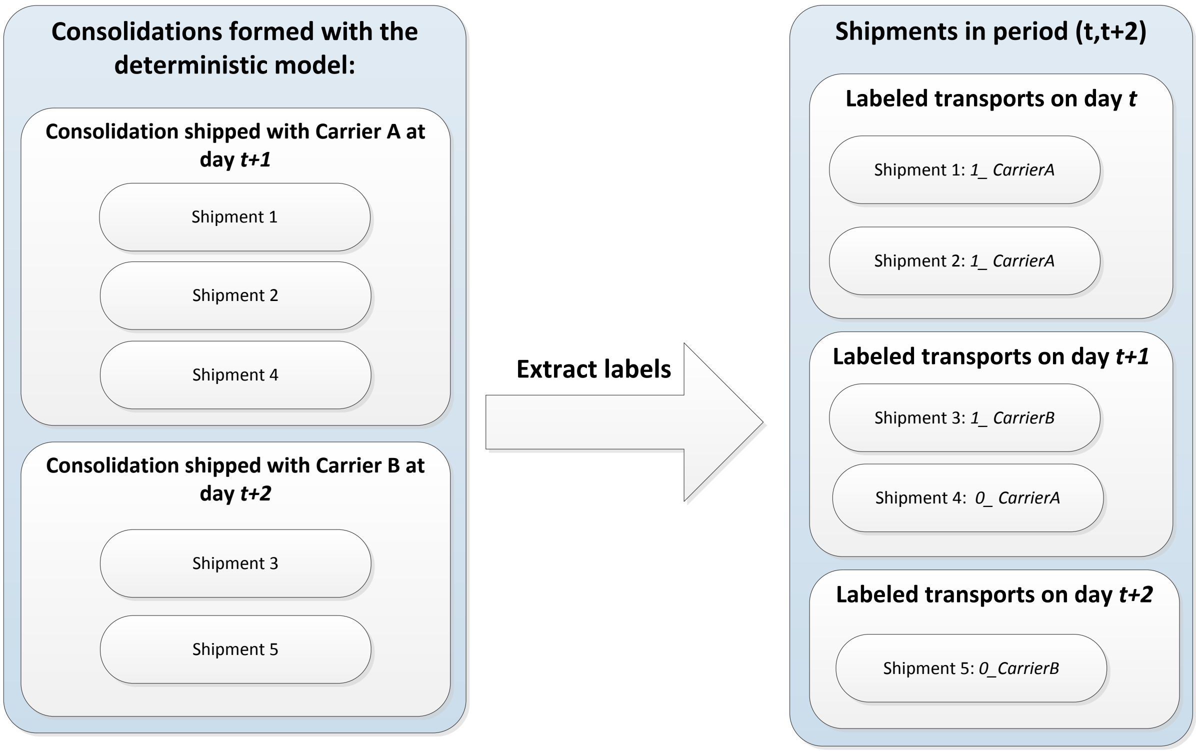

Figure 46: Label extraction multiple day consolidation

This also works the other way around: from the labels predicted by the machine learning models, the consolidation planning can be determined. To illustrate this, Figure 47 has been made.

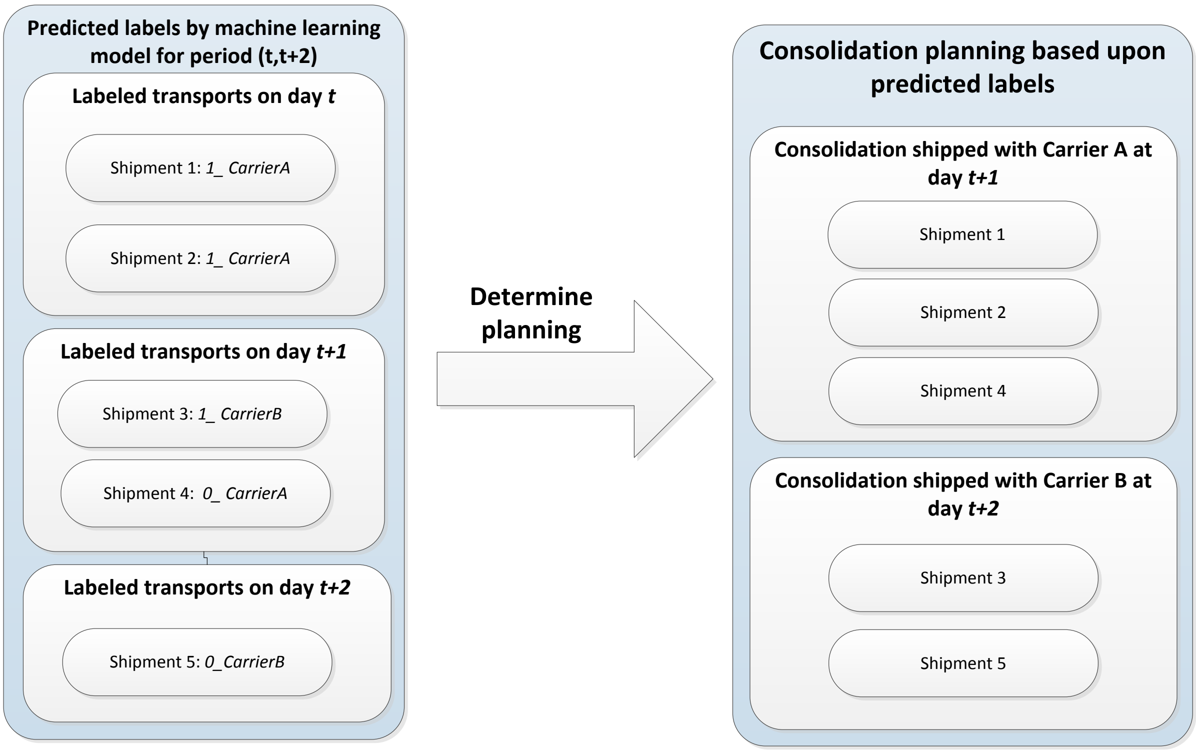

Figure 47: Consolidation planning for multiple day consolidation based upon predicted labels

5.4.3 Constructing and selecting variables

As mentioned before in section 5.2, consolidation planning decisions rely more on the mix and total volume of shipments in a certain timeframe, than on the characteristics of a shipment itself. Therefore, variables containing information about the mix of shipments in a certain time period should be created. The total accumulated volume and weight for all shipments in a certain period could be a good indicator on how to make planning decisions for consolidation.

For same day consolidation, the accumulated volume of weight of all shipments on a day can be used as a variable. However, this provides no information about the mix of shipments obtained on that day (e.g. small shipments, large shipments, hazardous shipments etc.). Therefore, it chosen to distinguish between the total accumulated volume and weight on a day per carrier, where the carrier is based upon the single shipment allocation with the CT model developed in research cycle 1. The single shipment carrier provides a lot of information about an individual shipment. Based upon this, the following additional variables were constructed:

- For every single shipment carrier available on a transport lane, the total accumulated weight (kilograms) per requested pick-up date needs to be determined, based upon the single shipments carrier allocation.
- For every single shipment carrier available on a transport lane, the total accumulated volume (cubic metres), per requested pick-up date needs to be determined, based upon the single shipments carrier allocation.

It must be noted that for all shipments requested on the same pick-up date, the same values for these variables will be obtained. However, in combination with the characteristics of each individual shipment on a day, these variables will provide a lot of information to the model to make consolidation decisions. The following characteristic of an individual shipments are included as a variables:

- Single shipment carrier
- Lead time in days
- Weight of single shipment in kilograms
- Volume of single shipment in cubic metres
- Hazardous good (Yes/No).

For consolidation over multiple days, the construction of additional variables is more complicated. When for example a shipment is scheduled two days from its requested pick-up date, some estimate of future demand should be made. Based upon the core business of the client in scope, it is expected, that demand patterns exist for each transport lane. The idea was to look back for a couple of days, and to make future planning decisions based upon past activity. Since the amount of activity in the last couple of days could provide information about the activity to come in the future. With this reasoning, it was decided to create the following variables. From today, look *n* days back, for every day in this period (including today). Determine the total accumulated weight and volume and distinguish per carrier (based upon the single shipment carrier allocation). Again, the values of these variables from day to day are the same. Since for every shipment on a lane on a given day, the same history can be obtained. However, these variables give the model information about the demand observed in the past, which can have influence on future demand and therefore future planning decisions. Combining this information with the following characteristics of each individual shipment, it is possible to provide a lot of information to the model to learn consolidation planning decisions:

- Single shipment carrier
- Lead time in days
- Weight of single shipment in kilograms
- Volume of single shipment in cubic metres
- Hazardous good (Yes/No)
- Week of the day for the requested pick-up date: planning decision could be influenced by the day of the week, because of patterns in demand.

After some testing, it was decided to look back for 2 days (*n=2)*. Since this yielded the best results in terms of cost savings.

5.4.4 Data transformation

For both the *CT* model and the *ANN* model, the data needs to be transformed. This is done to make sure the algorithms understand the data and to boost performance. For both algorithms different data transformation were done.

- *Classification Tree:* categorical variables are integer encoded. Other methods like binary encoded and binary encoding were tested. However, best results were obtained with integer encoding for the classification tree algorithm in term of cost savings. Numerical variables are not transformed for the classification tree algorithm.
- *Neural Network:* For the encoding of categorical variables different methods like: integer encoding, binary encoding and one-hot encoding were tested. Eventually the best results were obtained with one-hot encoding in term of cost savings. For the transformation of numerical variables, two approaches were tested: standardizing and normalizing. The best results were obtained when numerical variables were normalized in terms of cost savings.

For more elaboration on the data transformation techniques, see section 4.4, where these techniques are already discussed.

5.4.5 Training set and test set

In previous phases of this project, training sets and test sets were random generated based upon a pre-determined distribution between the two. For this phase of the project, this is not possible. To be able to compare the consolidation planning made by the machine learning models, with the optimal consolidation planning made by the DT model. The exact same mix and total volume of shipments in a period needs to be compared. Only then a fair cost comparison between the machine learning models and the DT models can be made. In section 5.4.1, it was explained that for each lane, 50 additional datasets were made which gives a total of 51 datasets per lane (including the original dataset). For both machine learning algorithms, at random 1 of 51 datasets will be chosen as a test set, while the algorithms are trained on the other shipments within the 50 datasets.

5.5 Modelling

In this section the development of the machine learning models, that are able to learn consolidation planning decisions, will be discussed. To learn consolidation planning decisions, the same machine learning algorithms will be used as for the single shipments phase*: CT's* and *ANN's*. Note that both algorithms are already briefly explained in section 4.5.1 and 4.5.2 respectively. We supplied both, same day as well as multiple day consolidation plans to the models. We designed separate models per lane, since:

1. Shipment consolidation is done per lane.
2. Different lanes have different carriers available, which vary in terms of costs, transport capacity, service level, etc.

By employing the separate models approach, no distinction is required for different lanes and carriers available on these lanes. This simplifies the learning process for the algorithms. In section 5.5.1 the development of the *CT* model will be discussed. In section 5.5.2 the development of the *ANN* model will be discussed.

5.5.1 Classification tree model

For the development of the *CT* model, again the Scikit-learn library was used within Python. The optimization of the hyper parameters of the algorithm was done via a grid search with 10 fold cross validation. The used parameter grid can be seen in *Appendix O*. The following changes were with respect to the default parameter setting for each consolidation scenario:

- *Same day consolidation*: it was found that the “Entropy” criterion yielded slightly better results than the default “Gini” criterion. Furthermore no changes were made with respect to the default settings of the Sci-kit library.
- *Multiple day consolidation*: two adjustments were made based upon the grid-search; in the default settings there are no restrictions on the maximum depth of a tree. With the grid-search it was found that a maximum depth of 8 for the classification tree, yielded significantly better results. Furthermore, the default “Gini” criterion was changed to “Entropy”.

After running the model a few times, it became known that false positives for Carrier A and Carrier C were costly for consolidation over multiple days. To compensate for this, it was chosen to give less weight to these carriers compared to other carrier (ratio of 2 to 10). This lead to less false positives and a decrease in costs.

5.5.2 Neural network model

For the development of the *ANN* model in research cycle 2, again the Keras library was used within Python. The optimization of all hyper parameters of the model was done via a grid search with 10 fold cross validation. The first grid search was done to explore a range of feasible parameter settings. For the parameter grid used, see *Appendix P*. Then, another grid search was done on a more detailed level, to fine tune the hyper parameters of the model. This parameter grid is presented in *Appendix Q.* After the second grid search, the final parameters are obtained. Table 13 gives an overview of the used parameters in the final model.

Table 13: Final parameters ANN used in reseach cycle 2

Parameter name	Value
Neurons per hidden layer	200
No. hidden layers	2
Activation function hidden layers	ReLU
Activation function output layer	Softmax
Drop-out rate	50%
Batch size	100
Epochs	60
Loss function	Categorical cross entropy
Evaluation metric	Accuracy
Optimizer	RMSprop with default Keras settings

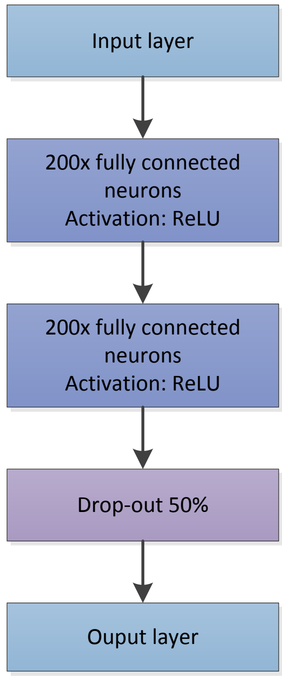

Figure 48: Structure ANN used in reseach cycle 2

The design of the final model is presented in Figure 48, which has a similar structure to the model being presented in section 4.5.2, and is based upon the same logic. The amount of neurons in the input layer varies, depending on the transport lane and the scenario the model is used for. Since the amount input variables is influenced by the amount of available carrier on a transport lane, as stated in section 5.4.3. Furthermore same consolidation yields less input variables, since there is no looking back in time. The input layer is connected to two fully connected layers with each 200 neurons. Both hidden layers have ReLU as their activation function. Next, a dropout layers is added to the model. Just like with the input layer, the output layer can have a different amount of neurons depended on the lane and scenario the model is used for. The biggest difference is due to the difference in scenario. Whereas the same day consolidation labels only consist of carriers to distinguish between, labels for consolidation over multiple days also need to distinguish between the individual waiting times of shipment. The Softmax activation is used to output the distributions of probability for each class (Xu & Liu, 2020). After the layers are added to the model, a loss function and optimization algorithm is chosen. The categorical cross entropy is chosen, since this is a specific loss function for categorical classifications with more than two classes. RMSprop is chosen as the optimizer based upon the grid search done. Since Keras recommends to leave the parameters settings at default, the values are left at their default values. It must be noted that these parameters setting are used for both scenarios.

5.6 EVALUATION

In this section the results obtained with the machine learning models are evaluated. The models are evaluated based upon *additional cost savings* and *feasibility* (see section 3.6). There is explicitly chosen, to not present the results for the metrics *accuracy* and *precision*. Remember that these metrics indicate how accurate labels can be predicted. With the planning of consolidations; a wrong predicted label does not necessary have to result in a bad or costly planning. This could give wrong impressions of the performance of the model. It has been calculated that around 80% of absolute savings are obtained with the 5 most active transport lanes. Based upon this it was decided to only present the results for these 5 transport lanes. The simulation approach used to obtain results is presented in section 5.6.1. Results are presented in section 5.6.2.

5.6.1 Simulation approach

To make sure the results are valid, results will be simulated based upon 1000 iterations. In section 5.4.5, it was elaborated that the test set is 1 data set randomly chosen out of the 51 datasets per lane. With this approach, only 51 iterations could be done. Since every dataset could be the test set only once. Therefore, it was chosen to hold out 5% (at random) of the training data for each iteration. With this approach; each iteration a new model gets trained on a different training set, while testing on 1 of the 51 datasets. This yields almost endless combinations of training sets and test sets. An overview of simulation approach adopted, is illustrated in Figure 49.

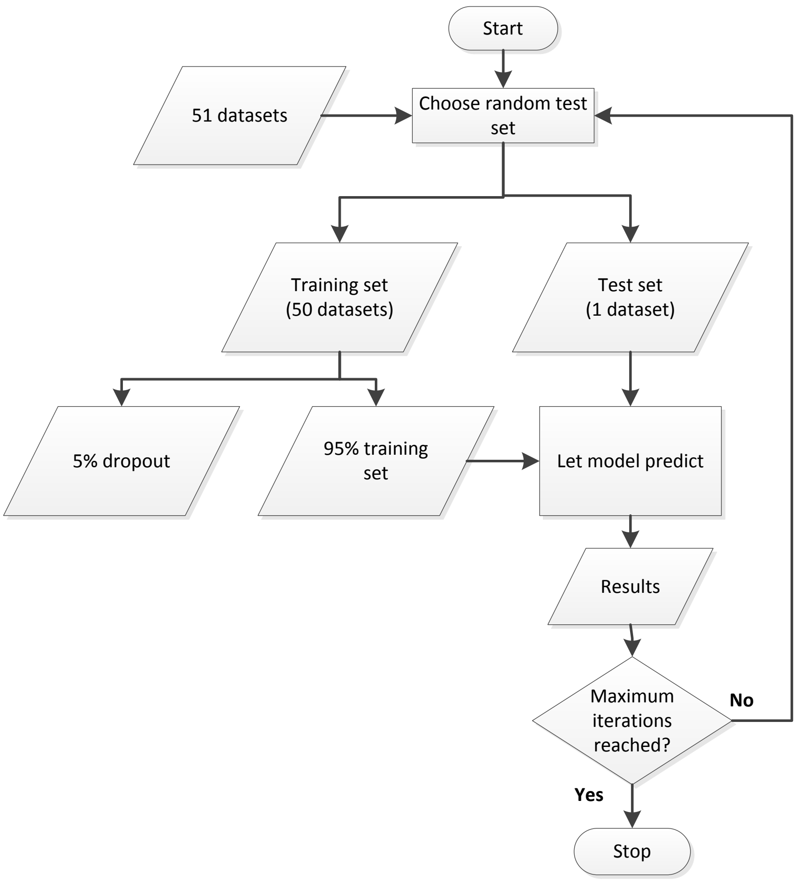

Figure 49: Simulation approach used to obtain results with consolidation models

5.6.2 Results

In Table 14, average additional cost savings per transport lane with consolidation are presented. In Table 15 till Table 18; the average feasibility regarding: hazardous, lead time, capacity and dimension constraints of the planning decisions made, are presented. Note that *fwt* is defined as the fixed (maximum) waiting time for shipments to be dispatched: 0 for *same day consolidation* and ∞ for *multiple day consolidation*.

Table 14: Average additional cost savings obtained per transport lane with consolidation models

	Transport lane				
	1	**2**	**3**	**4**	**5**
Consolidation (ANN, fwt=0)	31.0%	27.7%	12.3%	7.5%	7.2%
Consolidation (CT, fwt=0)	32.2%	29.0%	12.3%	9.3%	7.3%%
Consolidation (ANN, fwt=∞)	26.8%	26.3%	10.4%	9.9%	6.1%
Consolidation (CT, fwt=∞)	28.2%	24.4%	1.9%	8.3%	3.3%

Table 15: Average hazard feasibility of consolidation models

	Transport lane				
	1	**2**	**3**	**4**	**5**
Consolidation (ANN, fwt=0)	100%	100%	100%	100%	100%
Consolidation (CT, fwt=0)	99.9%	99.9%	99.9%	100%	100%
Consolidation (ANN, fwt=∞)	100%	100%	100%	99.9%	100%
Consolidation (CT, fwt=∞)	99.7%	99.9%	99.9%	99.9%	100%

Table 16: Average lead time feasibility of consolidation models

	Transport lane				
	1	**2**	**3**	**4**	**5**
Consolidation (ANN, fwt=0)	100%	100%	100%	100%	100%
Consolidation (CT, fwt=0)	100%	99.9%	99.9%	100%	100%
Consolidation (ANN, fwt=∞)	100%	99.9%	100%	100%	100%
Consolidation (CT, fwt=∞)	99.9%	99.9%	99.9%	100%	100%

Table 17: Average capacity feasibility of consolidation models

	Transport lane				
	1	**2**	**3**	**4**	**5**
Consolidation (ANN, fwt=0)	100%	100%	100%	100%	100%
Consolidation (CT, fwt=0)	99.9%	100%	100%	100%	100%
Consolidation (ANN, fwt=∞)	100%	100%	100%	100%	100%
Consolidation (CT, fwt=∞)	94.5%	100%	100%	100%	100%

Table 18: Average dimensions feasibility of consolidation models

	Transport lane				
	1	**2**	**3**	**4**	**5**
Consolidation (ANN, fwt=0)	100%	100%	100%	100%	100%
Consolidation (CT, fwt=0)	99.9%	99.9%	100%	100%	100%
Consolidation (ANN, fwt=∞)	100%	99.9%	100%	100%	100%
Consolidation (CT, fwt=∞)	94.5%	96.1%	100%	100%	100%

With all consolidation models developed, additional cost savings are obtained compared to single shipments CT model. For 4 out of the 5 transport lanes, using a *CT* algorithm to train same day consolidation planning decisions yielded the most cost savings on average. For only 1 out the 5 lanes (lane 4), using an *ANN* to train consolidations over multiple days yielded the most additional cost savings on average (9.9%). Here, the second best option was to train same day consolidation planning decisions with the *CT* algorithm, with additional cost savings of 9.3% on average. It can be observed that service, hazardous, dimension and capacity constraints are satisfied for all lanes, given all consolidation models developed. For full results see *Appendix R,* which includes minimum and maximum values, standard deviations and 95% confidence intervals obtained with the consolidation models regarding *relative cost savings* and *feasibility*.

q6; Neural network provide better results in mutiple day consolidation and CT better in same day consolidation? Can you explain why that is?

q7: Same data - 50 new data sets, -random sets.

6 Discussion

Now that all the relevant results have been presented, the results can be discussed and reflected upon. In section 6.1, final results and recommendations are given on how to generate improved business rules for the planning of shipments. In section 6.2, limitations of this master thesis will be discussed. In section 6.3, future research directions will be mentioned.

6.1 Results and recommendations

At forehand, a background study was done to determine which machine learning algorithms could be applicable for transportation planning problems (*sub-question 1*). Several algorithms were found (see Chapter 2), eventually it was decided to consider two different machine learning models algorithms in this research: classification trees and neural networks (see section 3.5). Research was done in two different cycles: 1) planning of single shipments and 2) planning of consolidated shipments. Results of both research cycles, and recommendations will be discussed now.

For the development of single shipment models in research cycle 1, knowledge of the current transportation planning (*sub-question 2*) and the transportation network of Company A was required (*sub-question 3*). Main findings included that shipments are planned individually (shipments are not consolidated), and shipments are assigned to carriers based upon their characteristics (e.g. weight, volume etc., see section 4.1.2). Furthermore, the transportation network of Company A can be characterized as a hub-spoke formation (section 4.1.1). A data analysis was done on the historical data provided by Company A, results indicated that most shipments demanded by Company A are small (e.g. less than 10 kg and less than $0.1m^3$). With this knowledge, CT and ANN models were developed that were able to assign single shipments to the (near) optimal carrier in research cycle 1. This resulted in expected savings of 38.7% and 40.1% over the current decision making, for the CT model and ANN model respectively. It was observed, that the cost savings presented with the machine learning models were higher than the optimal cost savings obtained with the deterministic model. Machine learning models results are based upon on 20% of random data each iteration (section 4.6.1), while the results of the deterministic model are based upon the whole historical data set. The different sets (with different shipments) used, explain the differences in cost savings obtained. To estimate the cost savings obtained with the machine learning models for the whole dataset, the mean optimality gap obtained with the machine learning models is subtracted from the savings potential (38.7%, see section 4.3.3). This yields estimated cost savings of (*38.7%- 5.2%=*) 33.5% and (*38.7%- 3.9%=*) 35.8% for the CT model and ANN model respectively. Performance in terms of accuracy is comparable, while differences in cost savings can be observed (see section 4.6.2). For the reason that miss-classifications costs with the CT model are more expensive in general.

In research cycle 2, models were developed to plan consolidated shipments. It was investigated how the consolidation problem should be defined for LLP, to fit within the boundaries of the business (*sub-question 4*). Main findings included that shipments can only be consolidated on pre-determined transport lanes, due to the current rate structure of carriers and the hub-spoke formation (see section 5.1.1 for full definition). Furthermore, two different consolidation scenarios were defined: same day consolidation and multiple day consolidation. A literature review was done to investigate which consolidation methods are used in prior research (*sub-question 5*). Several methods were found, ranging from rule based policies to linear programs (section 5.1.2). It was decided to solve the consolidation problem with a linear program. With a linear program optimal

q3: Experience of the planner, insights, how to incoporate that in your solutions?
q4: Which domain are suitable for machine learning? is logistics one such approach?

consolidation planning decisions can be obtained, which is crucial when training machine learning models. A data analysis was done to determine if consolidation would be feasible for Company A, and on which transport lanes (*sub-question 6*). It was found that rates are decreasing with more volume (economies of scale), and enough transport activity was observed to consolidate shipments on certain transport lanes (section 5.2). Optimizing both consolidation scenarios with the deterministic models, resulted in significant more cost saving for multiple day consolidation. This is logic, since more transport activity could be observed over multiple days which yields more economies of scale (section 5.2.2). Both scenarios were supplied to both machine learning algorithms. After evaluating the results obtained for both scenarios with the machine learning models, it was concluded that significant more cost savings could be obtained with same day consolidation for both algorithms. Due to unknown future demands, multiple day consolidation yields more uncertainty, which leads to more costly decisions. For same day consolidation, all shipments are assumed to be known at the end of a day, which decreases uncertainty. Same day consolidation yields expected additional savings between 7% and 32% for the CT models, and savings between 7% and 31% can be expected for the ANN models, dependent on the transport lane (section 5.6.2).

Results for both algorithms used, are comparable. However, the choice of which model to use, for generating new business rules, is limited to the CT models. Since rule extraction techniques from trained neural networks were not studied in this research, as was already mentioned in section 3.5. Remember that no performance is lost when deriving business rules from CT models, which means performance for the business rules will be equal to the CT models. Based upon the results, the following recommendations are given:

- From the learnt relations within the CT models, it is recommended to extract business rules in the form of *IF THEN* rules. CT models can be translated into a set of rules by creating a separate rule for each path from the root to a leaf (label) in the tree (Kotsiantis et al., 2006). An example of a business rule generated for single shipments is:

  $IF\ ((destination\ country = Australia, Brazil, Canada, China)$

  $AND\ (origin\ country = Germany)\ AND\ (gross\ volume \leq 0.12\ m^3)$

  $AND\ (hazardous = FALSE)\ AND\ (requested\ lead\ time \leq 3\ \text{days}))\ THEN\ Carrier\ D5$

  For the allocation of single shipments to carriers, around 300 *IF THEN* rules will be generated. All 300 rules are mutually exclusive and collectively exhaustive.

- It is recommended to adopt a same day consolidation program, since more cost savings can be expected at lower risks. Extracting business rules in the form of IF THEN rules yields business rules such as:

  $IF\ ((total\ weight\ accumulated\ on\ a\ day\ with\ single\ shipments\ labelled\ as\ Carrier\ D1 \geq 20kg)*$

  $(AND\ requested\ lead\ time \geq 3\ days))\ THEN\ Consolidate\ shipment\ with\ Carrier\ E$

  *Note that information about the mix and volume of single shipments obtained on a day is used to make consolidation decisions (section 5.4.3). Per transport lane around 200 *IF THEN* rules will be generated.

- Following the recommended approach for both scenarios; results in one set of rules (+-300 rules) for single shipments, and five set of rules for consolidation (+-200 rules each) for transport lanes with a consolidation program. In total 6 different sets of rules will be obtained. These sets need to be integrated in one set of rules (+-1300 rules). The following approach is recommended, see Figure 50.

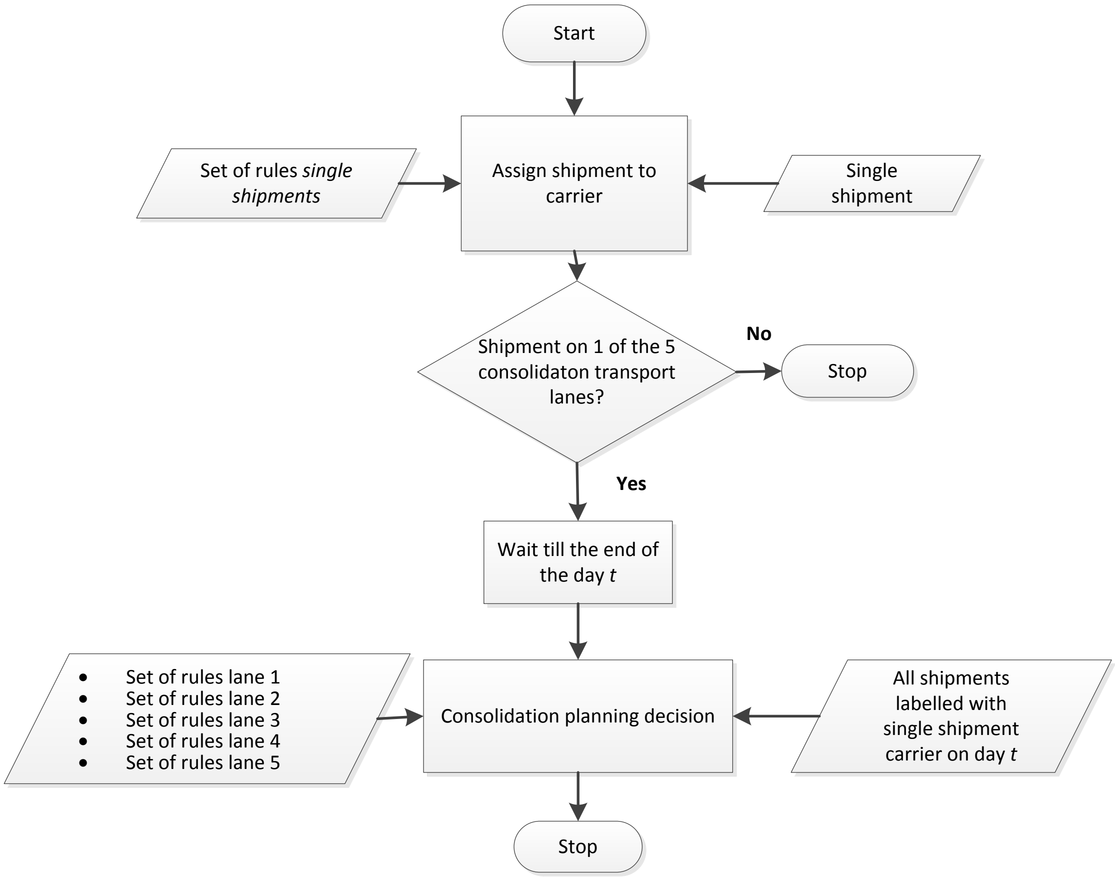

Figure 50: Overview of decision making with the new sets of rules obtained

First, every single shipment is assigned to a carrier. Then, it is decided whether the shipments is requested on one of the pre-determined transport lanes with a consolidation program. If no, the shipment can be shipped immediately. If yes, the consolidation planning process for this shipment will be delayed until the end of the day, until all shipments for that day are known. Finally, consolidation planning decisions per transport lane are made following the business rules. The approach presented, can be used to integrate the new business rules inside the TMS.

6.2 LIMITATIONS

It was decided to use a half year of data. This could be insufficient to give a realistic view of demands required by Company A. With the consequence, that the new business rules could have difficulties to make the right planning decisions for shipments, that the machine learning models have not been trained on. To solve this, the approach adopted in this research could be repeated with more historical data.

It was decided to develop different machine learning models per transport lane for a same day consolidation program. This results in around 200 IF THEN rules per transport lane. It is expected that duplicate rules are generated among the different transport lanes, which implies more rules are generated than necessary. The amount of rules integrated in the TMS, could impact the speed of the decision making. This could limit the use of the business rules in practise.

At forehand, it was assumed that all shipments with the same requested pick-up date can wait till the end of a day to be consolidated. In practise there are certain time slots in which planning decisions need to be made for shipments. This could be a limitation for the use of the new consolidation business rules in practise.

The new business rules generated for consolidation are dependent on the single shipments business rules to make decisions. This implies that the proposed machine learning approach for consolidation cannot be used without the support of the single shipment business rules. This could limit other companies, willing to only adopt a consolidation program only.

The proposed new business rules have not been tested in practise. Therefore, the total cost savings obtained with sets of business rules for single shipments and consolidation, cannot be presented. Furthermore, no research is done on when to update the new business rules again. However, it is expected that the new business rules will be valid when prices and demand patterns remain the same. To identify decreased performance on time, the business rules can be updated each half year to then compare performance.

In this research, also the application of neural networks was investigated. Although, parameters of these models have been modified to increase performance. It is expected that devoting more time to the optimization of these parameters would yield better results.

6.3 FUTURE RESEARCH

The new business rules consist of several sets of business rules that work together as one. Future studies could investigate the possibilities of generating all business rules at once by using one comprehensive machine learning model, e.g. one model that is trained to make single shipment and consolidation planning decisions. This would reduce the amount of business rules generated significantly, since it is expected that duplicate rules exist for consolidation (section 6.2).

Based upon the results obtained with the DT model (5.3.3), it was concluded that multiple day consolidation yields more cost savings than a same day consolidation program. However, with the machine learning models more cost savings could be obtained with a same day consolidation program. It is expected, that this is due to lack of information about future demands provided to the machine learning models. It could be investigated if a forecast of demand (e.g. 1 week) would increase the performance of multiple day consolidation with machine learning models.

Neural networks were investigated for the application of transportation planning. With the neural network models significant cost savings could be obtained. Since results look promising, it is recommended to study rule extraction techniques for future applications of neural networks in the logistics planning domain via business rules.

7 CONCLUSION

In this research, it is investigated how business rules can be developed with machine learning, to select the optimal planning solution for shipments, that benefit DHL LLP customers. A novel approach was adopted, machine learning models were used to generate business rules for the support of transport planning decisions. Machine learning algorithm were trained on optimal planning decisions derived from deterministic models, making it a classification problem. We showed that machine learning algorithms are able to learn transport planning decisions with good performance. By deriving business rules from the trained machine learning models, we opened up the black box of the machine learning models. From the CT models around 1300 *IF THEN* rules were generated. The new set of business rules support planning decisions regarding:

- The planning of single shipments; assigning shipments to a carrier
- The planning of consolidated shipments on transport lanes; forming consolidations of shipments and assigning consolidations to carriers

With these new business rules, cost savings of around 33.5% can be expected for single shipments. On the 5 transport lanes with a same day consolidation program additional cost savings between 7% and 32% can be expected dependent of the transport lane.

The presented approach in this Master thesis could be helpful for other companies willing to adapt machine learning into their operations, without the need to actually implement the models directly in the IT environment. Instead, business rules could be used to support the current decision making regarding transport planning. The approach presented, is not limited to the logistics/transportation sector, and can be used in a variety of sectors to optimize planning via business rules.

REFERENCES

Al Rahhal, M. M., Bazi, Y., Al Zuair, M., Othman, E., & BenJdira, B. (2018). Convolutional Neural Networks for Electrocardiogram Classification. *Journal of Medical and Biological Engineering*, *38*(6), 1014–1025.

Baykasoglu, A., & Kaplanoglu, V. (2011). A multi-agent approach to load consolidation in transportation. *Advances in Engineering Software TA - TT -*, *42*(7), 477–490.

Bhavsar, H., & Ganatra, A. (2012). A Comparative Study of Training Algorithms for Supervised Machine Learning. International Journal of Soft Computing and Engineering (IJSCE).

Bookbinder, J. H., Cai, Q., & He, Q. M. (2011). Shipment consolidation by private carrier: The discrete time and discrete quantity case. *Stochastic Models*, *27*(4), 664–686.

Bookbinder, J. H., Elhedhli, S., & Li, Z. (2015). The air-cargo consolidation problem with pivot weight: Models and solution methods. *Computers and Operations Research TA - TT -*, *59*, 22–32.

Bookbinder, J. H., & Higginson, J. K. (2002). Probabilistic modeling of freight consolidation by private carriage. *Transportation Research Part E: Logistics and Transportation Review*, *38*(5), 305–318.

Brownlee, J. (2019). How to use Data Scaling Improve Deep Learning Model Stability and Performance. Retrieved January 17, 2020, from https://machinelearningmastery.com/how-to-improve-neural-network-stability-and-modeling-performance-with-data-scaling/

Castelvecchi, D. (2016). Can we open the black box of AI? *Nature*, *538*(7623), 20–23.

Çetinkaya, S., & Bookbinder, J. H. (2003). Stochastic models for the dispatch of consolidated shipments. *Transportation Research Part B: Methodological*, *37*(8), 747–768.

Chakraborty, M., Kumar Biswas, S., Purkayastha, B., & Purkayastha biswajit, B. (2019). Rule Extraction from Neural Network Using Input Data Ranges Recursively, *37*, 67–96.

Cheng, J., Yang, B., Gen, M., Jang, Y. J., & Liang, C.-J. (2020). Machine learning based evolutionary algorithms and optimization for transportation and logistics. *Computers & Industrial Engineering*, *143*, 106372.

Choudhary, A. K., Harding, J. A., & Tiwari, M. K. (2009). Data mining in manufacturing: A review based on the kind of knowledge. *Journal of Intelligent Manufacturing*, *20*(5), 501–521. \

Dondo, R. G., & Mendez, C. A. (2014). A branch-and-price approach to manage cargo consolidation and distribution in supply chains. *Industrial and Engineering Chemistry Research*, *53*(44), 17226–17239.

Fosso Wamba, S., Gunasekaran, A., Papadopoulos, T., & Ngai, E. (2018). Big data analytics in logistics and supply chain management. *International Journal of Logistics Management*.

Ghoddusi, H., Creamer, G. G., & Rafizadeh, N. (2019). Machine learning in energy economics and finance: A review. *Energy Economics*, *81*, 709–727.

Haglin, J. M., Jimenez, G., & Eltorai, A. E. M. (2019). Artificial neural networks in medicine. *Health and Technology*.

Hailesilassie, T. (2016). Rule Extraction Algorithm for Deep Neural Networks: A Review, *14*(7), 376–381.

Hanbazazah, A. S., Abril, L., Erkoc, M., & Shaikh, N. (2019). Freight consolidation with divisible shipments, delivery time windows, and piecewise transportation costs. *European Journal of Operational Research TA - TT -*, *276*(1), 187–201.

Hosseini, S. D., Shirazi, M. A., & Ghomi, S. M. T. F. (2014). Harmony search optimization algorithm for a novel transportation problem in a consolidation network. *Engineering Optimization TA - TT -*, *46*(11), 1538–1552.

Hossin, M., & Sulaiman, M. . (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*, *5*(2), 01–11.

Huang, K., & Chi, W. (2007a). A Lagrangian relaxation based heuristic for the consolidation problem of airfreight forwarders. *Transportation Research Part C TA - TT -*, *15*(4), 235–245.

Huang, K., & Chi, W. (2007b). A Lagrangian relaxation based heuristic for the consolidation problem of airfreight forwarders. *Transportation Research Part C: Emerging Technologies*, *15*(4), 235–245.

Hughes, S., Moreno, S., Yushimito, W. F., & Huerta-Cánepa, G. (2019). Evaluation of machine learning methodologies to predict stop delivery times from GPS data. *Transportation Research Part C: Emerging Technologies*, *109*, 289–304.

Kitchenham, B., & Charters, S. (2007). issue: EBSE 2007-001. *Technical Report*, *2*(3).

Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review*, *26*(3), 159–190.

Kourentzes, N., Barrow, D. K., & Crone, S. F. (2014). Neural network ensemble operators for time series forecasting. *Expert Systems with Applications*, *41*(9), 4235–4244.

Lai, Y., Sun, H., & Ren, J. (2018). Understanding the determinants of big data analytics (BDA) adoption in logistics and supply chain management: An empirical investigation. *International Journal of Logistics Management*, *29*(2), 676–703.

Leung, L. C., Van Hui, Y., Wang, Y., & Chen, G. (2009). A 0-1 LP Model for the Integration and Consolidation of Air Cargo Shipments LK. *Operations Research TA - TT -*, *57*(2), 402–412.

Li, Y., Tao, Y., & Wang, F. (2009). A compromised large-scale neighborhood search heuristic for capacitated air cargo loading planning. *European Journal of Operational Research*, *199*(2), 553–560.

Ling, C. X., & Li, C. (1998). *Data Mining for Direct Marketing: Problems and Solutions*.

Mitra, S., Pal, S. K., & Mitra, P. (2002). Data mining in soft computing framework: a survey. *IEEE Transactions on Neural Networks*, *13*(1), 3–14.

Okuda, T., Tanaka, E., & Kasai, T. (1976). A Method for the Correction of Garbled Words Based on the Levenshtein Metric. *IEEE Transactions on Computers*, *C-25*(2), 172–178.

Optimizers - Keras Documentation. (n.d.). Retrieved March 23, 2020, from https://keras.io/optimizers/

Pai, C., & Potdar, K. (2017). A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers. *Article in International Journal of Computer Applications*, *175*(4), 975–8887.

Papadopoulos, T., Gunasekaran, A., Dubey, R., & Fosso Wamba, S. (2017). Big data and analytics in operations and supply chain management: managerial aspects and practical challenges. *Production Planning and Control*, *28*(11–12), 873–876.

Provost, F., & Fawcett, T. (2013). Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking. *Sebastopol, Calif. : O'Reilly.*, 386.

Salzberg, S. L. (1994). C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993. *Machine Learning*, *16*(3), 235–240.

Shanker, M. S., Hu, M. Y., & Hung, M. S. (1996). Effect of data standardization on neural network training. *Omega*, *24*(4), 385–397.

sklearn.tree.DecisionTreeClassifier — scikit-learn 0.22.2 documentation. (n.d.). Retrieved March 23, 2020, from https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html#sklearn.tree.DecisionTreeClassifier

Townsend, J., Chaton, T., & Monteiro, J. M. (2019). Extracting Relational Explanations From Deep Neural Networks: A Survey From a Neural-Symbolic Perspective. *IEEE Transactions on Neural Networks and Learning Systems*, 1–15.

Tyan, J. C., Wang, F.-K., & Du, T. C. (2003a). An evaluation of freight consolidation policies in global third party logistics. *Omega TA - TT -*, *31*(1), 55–62.

Tyan, J. C., Wang, F. K., & Du, T. C. (2003b). An evaluation of freight consolidation policies in global third party logistics. *Omega*, *31*(1), 55–62.

Ülkü, M. A. (2012). Dare to care: Shipment consolidation reduces not only costs, but also environmental damage. *International Journal of Production Economics*, *139*(2), 438–446.

van Heeswijk, W. J. A., Mes, M. R. K., Schutten, J. M. J., & Zijm, W. H. M. (2018). Freight consolidation in intermodal networks with reloads. *Flexible Services and Manufacturing Journal TA - TT -*, *30*(3), 452–485.

Wang, J., Zhang, L., Guo, Q., & Yi, Z. (2018a). Recurrent neural networks with auxiliary memory units. *IEEE Transactions on Neural Networks and Learning Systems*, *29*(5), 1652–1661.

Wang, J., Zhang, L., Guo, Q., & Yi, Z. (2018b). Recurrent neural networks with auxiliary memory units. *IEEE Transactions on Neural Networks and Learning Systems*, *29*(5), 1652–1661.

Williams, J. G. (2013). Bootstrap sampling is with replacement: A comment on Ayllón et al. (2011). *River Research and Applications*, *29*(3), 399–401.

Wirth, R. (2000). CRISP-DM : Towards a Standard Process Model for Data Mining. *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, (24959), 29–39.

Wong, W. H., Leung, L. C., & Hui, Y. Van. (2009). Airfreight forwarder shipment planning: A mixed 0-1 model and managerial issues in the integration and consolidation of shipments. *European Journal of Operational Research TA - TT -*, *193*(1), 86–97.

Xu, X., & Liu, H. (2020). ECG heartbeat classification using convolutional neural networks. *IEEE Access*, *8*, 8614–8619.

Zhou, G., Hui, Y. Van, & Liang, L. (2011). Strategic alliance in freight consolidation. *Transportation Research Part E TA - TT -*, *47*(1), 18–29.

Zhou, G., & Zhang, Y. (2017). Integration and consolidation in air freight shipment planning: An economic and environmental perspective. *Journal of Cleaner Production TA - TT -*, *166*, 1381–1394.

APPENDIX

A: INTERVIEW SUPPLY CHAIN ENGINEER ON 16-10-2019

Interviewer: "Do you have any idea, how other logistic providers make decisions for their daily operations?"

Supply chain engineer: "Yesterday I was at conference and talked with other logistic providers about this. Most confirmed to me, what I was already thinking. Most of them have simple rules based upon logic and expertise to guide daily operations with the support of a TMS. They are definitely not using any form of data driven approaches in their daily decision making. However most companies seemed to be enthusiastic about the concept though. "

Interviewer: "So most companies still use business rules inside a TMS without the use of any data driven approaches?"

Supply chain engineer: "As far as I know, yes this is the case right now. "

B: DHL DATA2MOVE PROJECT PROPOSAL

"Project title

Data-Driven Solutions for Transportation Problems with Machine Learning (ML) techniques

Objective

The LLP core operations are guided by Transport Management Systems (TMS) .The 'heart' of the TMS is the Business Rule – concept. These come down to rules such as "Any package made of type X under 70 kg has to be transported via Y". Due to the nature of the sector, adjusting, changing and creating Business Rules is a necessity. However, it can take a long time before Business Rules are changed and this process takes a lot of manpower. A study performed by DHL LLP showed that a rudimentary model which uses Machine Learning already could make the right decisions in 98% of the cases (given a certain set).

DHL LLP would like a student to find a way/model that can create Business Rules and maintain them, in order to press costs and improve efficiency. The student is urged to use machine learning methods to improve the overall efficiency in the client's Supply Chain via Business Rules.

Possible directions

- Identify different kind of decisions that are made in practice regarding parcel weight/volume and transportation method
- Develop a model that can determine optimal solutions taking into account these decisions

Timeline and Practical Information"

C: SAVINGS POTENTIAL

The numbers presented in Table 19 are based upon 6 months of shipments in 2019, and are calculated based upon the data provided by LLP.

Table 19: Savings potential single shipments

Costs current solutions	618,018 EUR
Costs optimal solutions	387,711 EUR
Absolute savings	232,716 EUR
Relative savings	38.72%

D: HIGH LEVEL OVERVIEW DATA 2019 OF COMPANY A

Table 20: Overview data Company A

No. Shipments	10129
Total Volume	3139 m^3
Total Weight	34256 kg
Origin – Destination pairs	835

E: DESCRIPTION OF VARIABLES IN DATASET

21: Overview of variables found in dataset

Variable	Description	Difference
TR Creation Date/Time	Date and time of creation of shipment	n/a
TR Code	Code of shipment	n/a
TR SU Code	Code of unit within shipment	n/a
Origin Country	Origin country of shipment	n/a
Origin City	Origin city of shipment	n/a
Origin Location Name	Origin location name of shipment	n/a
Origin Location Code	Origin location code of shipment	n/a
Origin Location Address	Origin location address of shipment	n/a
Origin Post Code	Origin postal code of shipment	n/a
Dest Country	Destination country of shipment	n/a
Dest City	Destination city of shipment	n/a
Dest Location Name	Destination location name of shipment	n/a
Dest Location Code	Destination location code of shipment	n/a
Dest Location Address	Destination location address of shipment	n/a
Dest Post Code	Destination postal code of shipment	n/a
Actual Pickup	Actual pickup date of shipment	n/a
Requested Pickup	Requested pickup date of shipment	n/a
Actual Delivery	Actual delivery date of shipment	n/a
Requested Delivery	Requested delivery date of shipment	n/a
TR Business Entity	Company that issued shipment	n/a
TR Name	Service level of shipment (AOG, CRT or RTN) for outbound shipments	Empty column for inbound shipments
TR Created By	Name of person who created the shipment in the system	n/a

TR Type	Service level of shipment (routine or non-routine) for inbound shipments	Empty column for outbound shipments
Incoterm	Incoterm used for shipment	n/a
Forwarder	Empty Column	n/a
Shp Forwarder Name	Allocated carrier for shipment by planning	n/a
TR Shipment Mode Short Code	Empty column	n/a
TR Hazardous (Y/N)	States if shipment contains hazardous goods (yes or no)	n/a
Gross Weight (KG)	Gross weight of shipment in kilograms	n/a
Net Weight (KG)	Net weight of shipment in kilograms	n/a
Gross Volume (M3)	Gross volume of shipment in m^3	n/a
Net Volume (M3)	Net volume of shipment in m^3	n/a
Special Instructions	Special instructions and information about shipment	n/a
TR SU Length (CM)	Length of unit in shipment in cm	n/a
TR SU Width (CM)	Width of unit in shipment in cm	n/a
TR SU Height (CM)	Height of unit in shipment in cm	n/a
TR SU Gross Weight (KG)	Weight of unit in shipment in cm	n/a

F: MISSING VALUES PER VARIABLE

Table 22: Missing values found per variables in dataset

Name of variable	Missing values
TR Creation Date/Time	0
TR Code	0
TR SU Code	0
Origin Country	0
Origin City	0
Origin Location Name	0
Origin Location Code	0
Origin Location Address	0
Origin Post Code	0
Dest Country	0
Dest City	0
Dest Location Name	1
Dest Location Code	0
Dest Location Address	0
Dest Post Code	0
Actual Pickup	917
Requested Pickup	0
Actual Delivery	934
Requested Delivery	0
TR Status	0
TR Business Entity	9034
TR Name	3603
TR Type	10249

TR Created By	0
Service Code	3506
Service Level	9034
Incoterm	0
Forwarder	12635
Shp Forwarder Name	457
TR Transport Mode Short Code	12637
Nb of Ship Units	0
Nb of Packaged Items	0
TR Hazardous (Y/N)	332
Gross Weight (KG)	0
Net Weight (KG)	0
Gross Volume (M3)	0
Net Volume (M3)	0
Customer Reference 1	3247
Customer Reference 2	3797
Customer Reference 3	3885
Shp Linked	454
Ord Linked	10898
Special Instructions	9760
TR SU Count	0
TR SU THU	327
TR SU Length (CM)	68
TR SU Width (CM)	119
TR SU Height (CM)	119
TR SU Gross Weight (KG)	5

G: PARAMETER GRID CLASSIFICATION TREE RESEARCH CYCLE 1

Table 23: Parameter grid classification tree research cycle 1

Parameter	Interval/type	Steps
Criterion	Gini, Entropy	-
Splitter	Best, Random	-
Max. depth	3,100	5
Min. samples split	2-20	2
Min. samples leaf	2-20	2

H:

PARAMETER GRID 1 ANN USED IN RESEARCH CYCLE 1

Table 24: Parameter grid 1 ann used in research cycle 1

Parameter name	Interval/type(s)	Steps
Neurons per layer	50-800	50
No. hidden layers	1-3	1
Activation function hidden layers	ReLU, ELU, Tanh	-

Drop-out rate	0%-70%	10%
Activation function output layer	Softmax	-
Optimizer	SGD, RMSprop, Adagrad, Adadelta	-
Batch size	0-300	100
Epochs	0-200	100

I: PARAMETER GRID 2 ANN USED IN RESEARCH CYCLE 1

Table 25: Parameter grid 2 ann used in research cycle 1

Parameter name	Interval/type(s)	Steps
Neurons per layer	450-550	10
No. hidden layers	2	-
Activation function hidden layers	ReLu	-
Drop-out rate	40%-50%	5%
Optimizer	RMSprop	-
Batch size	50-150	10
Epochs	50-150	10

J: ACCURACY AND PRECISION CT MODEL RESEARCH CYCLE 1

In this section the results obtained with the classification tree model will be presented and discussed. As can be seen in Table 26, the accuracy of the model is 96.8% on average. When looking at the mean, the standard deviation and the confidence intervals, the accuracy and the precision for almost all carriers is decent and stable. Except for the following carriers: Carrier B, Carrier D2 and Carrier D4, these results lack behind. This is mainly due to the low amount of samples for these classes in the dataset. Literature states that a technique like oversampling of these minority classes in the training set could be used to improve learning in some cases(Ling & Li, 1998). This was tried, but it seemed to have no effects on the results. The effect of the number of training samples on precision, can be seen by the fact that most carriers with high precision also have a high number of samples in the training set.

Table 26: Overview results CT model research cycle 1 showing accuracy and precision

	Mean	**Standard deviation**	**95% Confidence interval Lower**	**Upper**	**Average samples of label in test set**
Precision Carrier A	87.77%	2.94%	87.73%	88.09%	155
Precision Carrier B	74.34%	20.24%	73.60%	76.11%	7
Precision Carrier C	79.33%	4.34%	78.92%	79.46%	93
Precision Carrier D1	99.62%	0.19%	99.61%	99.63%	1126
Precision Carrier D2	45.14%	32.18%	42.37%	46.37%	4
Precision Carrier D3	98.28%	0.62%	98.20%	98.28%	610

Precision Carrier D4	73.18%	13.02%	72.48%	74.09%	13
Precision Carrier D5	98.34%	0.50%	98.30%	98.36%	815
Precision Carrier E	89.77%	1.81%	89.66%	89.89%	282
Accuracy	96.4%	0.57%	96.28%	97.15%	3105

K: ACCURACY AND PRECISION ANN MODEL RESEARCH CYCLE 1

As can be seen in Table 27, the *ANN* model obtains an accuracy of 96.36% on average. When looking at the mean, the standard deviation and the confidence intervals, the accuracy and precision for almost all carriers are decent and stable. Except for the following carriers: Carrier B and Carrier D2. The precision of Carrier B (48.99%) and Carrier D2 (51.03%), are low compared to the overall accuracy of 96.36%. Also the standard deviation for these carriers regarding precision is high compared to the other carriers. Except for Carrier D4, which also has a high standard deviation for precision, but a decent mean precision of 85.51%. This is mainly due to the low amount of samples for these classes in the dataset, as said before.

Table 27: Overview results ANN model research cycle 1 showing accuracy and precision

	Mean	**Standard deviation**	**95% Confidence interval Lower**	**Upper**	**Average samples of label in test set**
Precision Carrier A	91.51%	3.65%	90.78%	92.24%	155
Precision Carrier B	48.99%	41.13%	40.78%	57.19%	7
Precision Carrier C	83.65%	5.03%	82.65%	84.65%	92
Precision Carrier D1	99.36%	0.43%	99.27%	99.45%	1128
Precision Carrier D2	51.03%	35.06%	44.03%	58.02%	4
Precision Carrier D3	96.33%	1.18%	96.10%	96.57%	610
Precision Carrier D4	85.51%	11.90%	83.14%	87.89%	13
Precision Carrier D5	97.77%	0.90%	97.59%	97.95%	815
Precision Carrier E	88.91%	2.57%	88.40%	89.42%	281
Accuracy	96.8%	0.04%	95.29%	97.33%	3105

L: RESULTS LITERATURE REVIEW

In Table 28 below, the following abbreviations are used for the column “Method”:

- H: Hybrid method
- IOP: Integer Optimization Problem
- RBP: Rule Based Policy
- O: Other approaches

Table 28: Overview results literature review

Index	Article	Method	Description (Abstract citation)
1	(Baykasoglu & Kaplanoglu, 2011)	O	"In this paper we propose a multi-agent based load consolidation decision making approach. In the proposed approach the load consolidation decisions for the less-than-truckload orders are made by the software agents. The less-than-truckload orders are assigned/consolidated to the trucks by the negotiation mechanism constructed within the model."
2	(Çetinkaya & Bookbinder, 2003)	RBP	"In the present paper, we apply renewal theory to two strategies commonly utilized in practice. For the case of a quantity policy we obtain the optimal target weight before dispatch, while for a time policy, we calculate the optimal length of each consolidation cycle (maximum holding time for any order). These strategies are analysed for private carriage and then for common carriage."
3	(Dondo & Mendez, 2014)	H	"We present a methodology for finding near-optimal solutions to a problem related to LTL-shipping by using column generation combined with a customized branch-and-price procedure. The approach rapidly provides near-optimal solutions, since it solves the column generation sub-problems approximately and does not necessarily consider all unexplored nodes in the search-tree. We also present computational results on numerous test problems of varied topologies and on a real case study."
4	(Hanbazazah et al., 2019)	H	"This paper studies the freight consolidation problem for a third party logistics (3PL) provider that tranships products from multiple suppliers to a single business customer over a contracted multi-period horizon" … "A mixed integer programming model is developed for this problem, which employs piecewise cost functions to capture the economies of scales that are common in transportation. To speed up obtaining solutions, an exact solution methodology is proposed."
5	(Hosseini et al., 2014)	H	"This article presents a new harmony search optimization algorithm to solve a novel integer programming model developed for a consolidation network. In this network, a set of vehicles is used to transport goods from suppliers to their corresponding customers via two transportation systems: direct shipment and milk run logistics. The objective of this problem is to minimize the total shipping cost in the network, so it tries to reduce the number of required vehicles using an efficient vehicle routing strategy in the solution approach."
6	(Huang & Chi, 2007a)	H	"In this paper, the consolidation problem is first transformed into a well-known set covering problem by treating a feasible consolidated shipment as a set. Lagrangian Relaxation is used as the backbone to develop a recursive heuristic algorithm."
7	(J.H. Bookbinder et al., 2015)	H	"We formulate a mixed integer program, and propose four solution methodologies for the air-cargo consolidation problem under the pivot-weight scheme (ACPW). These are exact solution approaches based on branch-and-price, a best-fit decreasing loading heuristic, and two extended local branching heuristics (a multi-local tree search and a relaxation-induced neighbourhood search)"

8	(James H. Bookbinder & Higginson, 2002)	RBP	“Here we employ probabilistic modelling to choose the maximum holding time and desired dispatch quantity. We obtain practical decision rules for temporal consolidation for transportation in one’s own truck.”
9	(James H. Bookbinder et al., 2011)	RBP	“This article studies the dispatch of consolidated shipments. Orders arrive to a depot at discrete time epochs following a discrete time batch Markov arrival process (BMAP). A discrete time Markov chain for the accumulated weight of orders in the system is introduced and analysed.”
10	(Leung et al., 2009)	H	“This paper addresses the problem of determining the optimal integrations and consolidations of air cargo shipments. The problem of assigning shipment activities to processing units is formulated as a linear 0-1 program. Exploiting the special structure of this model, we design a solution procedure that includes heuristics and a branch-and-bound algorithm”
11	(Tyan, Wang, & Du, 2003a)	IOP	“This paper examines a special class of freight consolidation at an integrated global logistics company in global supply chain (GSC). A mathematical programming model has been developed to assist the evaluation of consolidation policies.”
12	(van Heeswijk, Mes, Schutten, & Zijm, 2018)	O	“We subsequently evaluate consolidation opportunities for the k best routes by applying a decision tree structure, taking into account reload operations, time tables, and synchronization of departure windows.”
13	(Wong et al., 2009)	H	“In this paper, we formulate a forwarder’s shipment planning problem as a mixed 0–1 LP. Effects of integration and consolidation – on the timely delivery of shipments during any phase of the shipping process – are explicitly addressed.”
14	(Zhou & Zhang, 2017)	IOP	“This study investigates the effects of carbon emissions within the framework of air freight integrator shipment planning, with integration and consolidation. A mixed 0-1 Linear Programming (LP) model is formulated to examine the effects of operational adjustments on carbon emissions and cost performance, as well as the effects of carbon emissions regulations on operational decisions.”
15	(Zhou et al., 2011)	RBP	“We adopt a simulation approach to find the optimal shipment dispatching plan in collaborative freight consolidation.”

M: BOOTSTRAP APPROACH

To explain exactly how additional data was created, the procedure is described in detail and supported with figures. To keep things simple, the approach is illustrated with a planning period of 10 days (2 weeks). Note that all steps described now, are normally done for the planning period of 149 days found in the historical dataset.

1. Identify all days with transport activity for the planning period of 10 days. Every day with transport activity is made green.

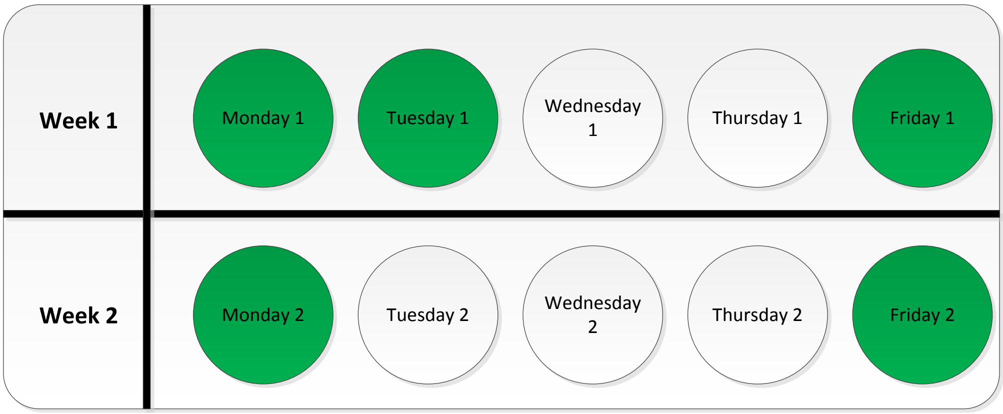

Figure 51: Identify days with transport activity

2. Collect all active days found in the planning period (149 days), and sort them by day of the week.

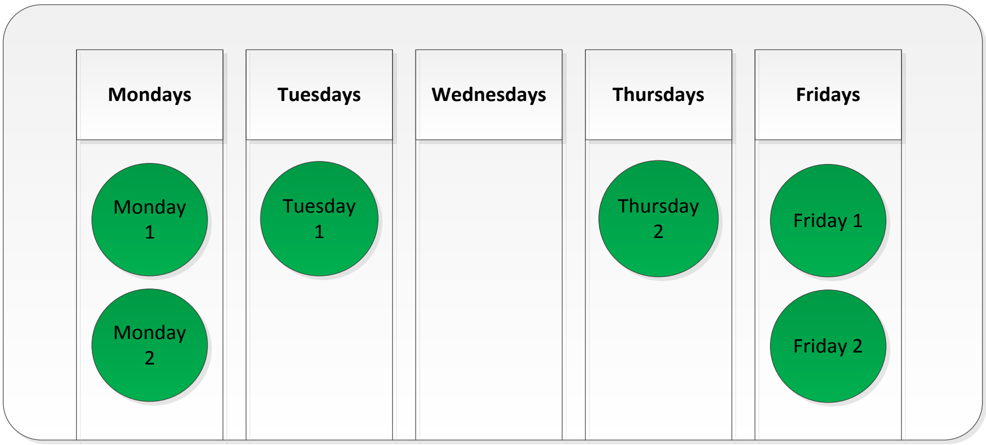

Figure 52: Sort days with transport activity by day of the week

Based upon this, days and shipments can be swapped around on active days to create new datasets. The following steps are taken to do so:

For all active days on a lane, identified in the planning period [t,t+10]:

1. Identify weekday for day t (E.g., Monday, first day of the planning period of 10 days)
2. Choose a random weekday (here, Monday) from all collected weekdays (here, Mondays). Let that be WeekdayX (say, Monday 2).
3. Get number of shipments X on this WeekdayX, see Figure 53

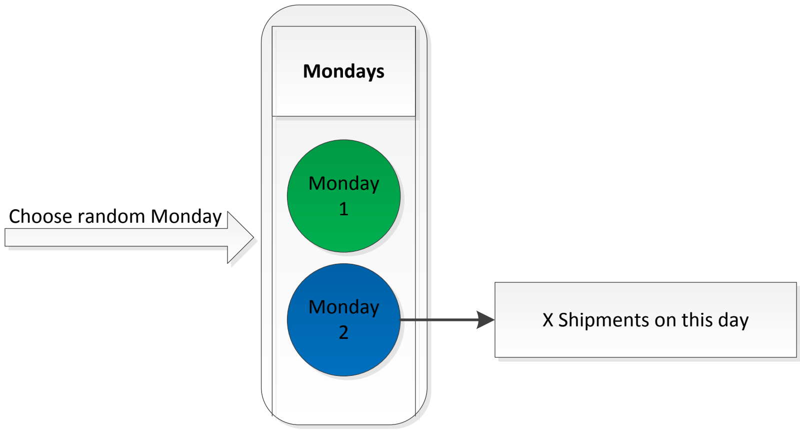

Figure 53: Get number of shipments on randomly chosen Monday: weekdayX

4. From all collected weekdays (here, Mondays). Choose *X* shipments at random. These *X* shipments are now scheduled to have their requested pick-up date at day *t.*

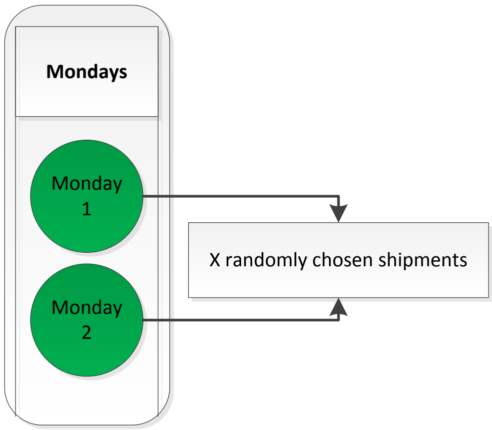

Figure 54: Choose x shipments from all mondays

5. Put all chosen X shipments back to their original day, so these shipments can again be randomly chosen again.

N: DATA LABELLING ANALYSIS

Every cross in the table below indicates a yes for the following phenomes:

- *Ph1*: are shipments dispatched at a different date than their requested pick-up date? If yes the label must provide information on how many days a shipment is held before dispatching, otherwise this information is not relevant.
- *Ph2*: are different carriers used for shipping consolidations? If yes, the label must provide information on which carriers was used to transport the consolidation of shipments, otherwise this information is not relevant.
- *Ph3*: are different consolidations send with the same carrier on the same day? If yes, the labels must distinguish between consolidations send on the same day with the same carrier, otherwise this information is not relevant.

See Table 29, for an overview of the results obtained per consolidation scenario for each phenome.

Table 29: Overview label analysis

Transport Lane	Same day consolidation			Multiple day consolidation		
	Ph1	Ph2	Ph3	Ph1	Ph2	Ph3
Lane 1		x	x		x	x
Lane 2		x	x		x	x
Lane 3		x	x		x	x
Lane 4		x	x		x	x
Lane 5		x	x		x	x
Lane 6		x			x	x
Lane 7		x	x		x	x
Lane 8		x			x	x
Lane 9		x	x		x	x
Lane 10		x	x		x	x
Lane 11		x	x		x	x
Lane 12		x	x		x	x
Lane 13		x	x		x	x
Lane 14		x	x		x	x
Lane 15		x			x	x
Lane 16		x	x		x	x
Lane 17		x			x	x
Lane 18		x			x	x
Lane 19		x			x	x
Lane 20		x			x	x

O: Parameter grid used CT for reseach cycle 2

Table 30: Parameter grid used CT for reseach cycle 2

Parameter	Interval/type	Steps
Criterion	Gini, Entropy	-
Splitter	Best, Random	-
Max depth	2,50	2
Min samples split	2-50	2
Min samples leaf	2-50	2

P: PARAMETER GRID 1 FOR ANN USED IN RESEACH CYCLE 2

Table 31: Parameter grid 1 used ANN for reseach cycle 2

Parameter name	Interval/type(s)	Steps
Neurons per layer	1-500	50
No. hidden layers	2	-
Activation function hidden layers	ReLU, ELU, Tanh	-
Dropout	30%-70%	10%
Optimizer	SGD, RMSprop, Adagrad, Adadelta	-
Batch size	0-200	20
Epochs	0-200	50

Q: PARAMETER GRID 2 FOR ANN USED IN RESEACH CYCLE 2

Table 32: Parameter grid 1 used ANN for reseach cycle 2

Parameter name	Interval/type(s)	Steps
Neurons per layer	150-250	10
No. hidden layers	2	-
Activation function hidden layers	ReLU	-
Dropout	40%-50%	5%
Optimizer	RMSprop	-
Batch size	50-150	10
Epochs	50-150	10

R: FULL RESULTS CONSOLIDATION MODELS

Full results for all transport lanes, are presented in the upcoming sections. Cost savings are relative to the current decision making.

Lane 1

The results regarding the cost savings are presented in Table 33, and results regarding the constraint feasibility are presented in Table 34.

Table 33: Relative cost savings for lane 1

	Min.	Mean.	Max.	Std.	95% confidence interval	
					Lower	Upper
Single Shipments Optimal	38.2%	45.2%	50.5%	3.4%	44.6%	46.4%
Consolidation Optimal (fwt=∞)	73.8%	79.0%	82.2%	2.2%	78.4%	79.7%

Consolidation Optimal (fwt=0)	69.6%	75.3%	78.6%	2.9%	74.4%	76.1%
Single Shipments (CT)	32.0%	41.6%	48.1%	3.7%	40.5%	42.6%
Consolidation (ANN, fwt=0)	64.8%	72.6%	77.2%	3.1%	71.7%	73.5%
Consolidation (CT, fwt=0)	68.3%	73.8%	77.5%	2.8%	73.2%	74.4%
Consolidation (ANN, fwt=∞)	60.4%	68.4%	75.1%	4.1%	67.2%	69.6%
Consolidation (CT, fwt=∞)	63.0%	69.8%	75.2%	2.8%	69.2%	70.3%

Table 34: Constraint feasibility lane 1

	Hazard Feasibility		**Lead Time Feasibility**		**Capacity Feasibility**		**Dimensions Feasibility**	
	Mean	**Std.**	**Mean**	**Std.**	**Mean**	**Std.**	**Mean**	**Std.**
Model Consolidation (NN, fwt=0)	100%	0%	100%	0%	100%	0%	100%	0%
Model Consolidation (CT, fwt =0)	99.9%	0%	100%	0%	100%	0%	99.9%	0%
Model Consolidation (NN, fwt =∞)	100%	0%	100%	0%	100%	0%	100%	0%
Model Consolidation (CT, fwt= ∞)	99.7%	0%	99.9%	0%	100%	0%	94.5%	1.9%

In Table 33, can be seen that all models yield extra savings compared to the single shipments model. Same day consolidation yields for both machine learning models the most cost savings compared to multiple day consolidation. Looking at the difference between both models for same day consolidation; the *CT* model (73.8%) yields more cost savings than the *ANN* model (72.6%) on average. Which is on average only 1.5% from the optimal solution for same day consolidation. For multiple day consolidation; the *CT* model (69.8%) also yields more cost savings on average than the *ANN* model(68.4%). The standard deviations and 95% confidence intervals for all models, are in line with the results obtained with the deterministic models. This means the models perform stable. Only the *ANN* model for consolidation over multiple days has a significantly higher standard deviations compared to the other models. In terms of constraint feasibility (Table 34), it can be seen that the *ANN* model performs slightly better compared to the *CT* model. However, even with the classification tree only 0.1% and 0.3% on average are not feasible for hazard and lead time constraints respectively.

Lane 2

The results regarding the cost savings are presented in Table 35, and results regarding the constraint feasibility are presented in Table 36.

Table 35: Cost savings lane 2

	Min.	**Mean.**	**Max.**	**Std.**	**95% confidence interval**	
					Lower	**Upper**
Single Shipments Optimal	3.9%	9.6%	16.6%	3.1%	9.1%	10.3%
Consolidation Optimal (fwt=∞)	37.8%	47.1%	54.4%	3.8%	46.0%	48.2%
Consolidation Optimal (fwt=0)	26.2%	37.2%	45.3%	5.7%	35.6%	38.8%
Model Single Shipments (CT)	2.6%	4.9%	13.3%	2.7%	4.4%	5.5%
Model Consolidation (NN, fwt=0)	23.3%	32.6%	42.5%	7.8%	30.3%	35.0%
Model Consolidation (CT, fwt=0)	23.1%	33.9%	42.4%	5.6%	32.8%	35.0%
Model Consolidation (NN, fwt=∞)	16.3%	31.2%	41.5%	6.7%	29.3%	33.2%

Model Consolidation (CT, fwt=∞)	17.0%	29.3%	39.9%	5.3%	28.2%	30.3%

Table 36: Constraint feasibility lane 2

	Hazard Feasibility		**Lead Time Feasibility**		**Capacity Feasibility**		**Dimensions Feasibility**	
	Mean	**Std.**	**Mean**	**Std.**	**Mean**	**Std.**	**Mean**	**Std.**
Model Consolidation (NN, fwt=0)	100%	0%	100%	0%	100%	0%	100%	0%
Model Consolidation (CT, fwt =0)	99.9%	0%	99.9%	0%	100%	0%	99.9%	0%
Model Consolidation (NN, fwt =∞)	100%	0%	99.9%	0.1%	100%	0%	99.9%	0.1%
Model Consolidation (CT, fwt= ∞)	99.9%	0%	99.9%	0%	100%	0%	96.1%	1.3%

In Table 35 can be seen that all models yield extra savings compared to the single shipments model. When comparing the two different scenarios; same day consolidation yields on average more cost savings for both models. The *CT* model (33.9%) has on average higher cost savings than the *ANN* model (32.6%). The standard deviation of the *CT* model is in line with the standard deviation of the deterministic model, whereas the standard deviation from the *ANN* model is slightly higher than the deterministic model. For consolidation over multiple days; the *ANN* model outperforms the *CT* model with 31.2% versus 29.3% cost savings. In Table 36, it can be seen that the feasibility regarding all constraints for every model is good with almost a 100% overall score.

Lane 3

The results regarding the cost savings are presented in Table 37, and results regarding the constraint feasibility are presented in Table 38.

Table 37: Cost savings lane 3

	Min.	Mean.	Max.	Std.	95% confidence interval Lower	Upper
Single Shipments Optimal	69.9%	72.7%	75.0%	1.3%	72.4%	73.0%
Consolidation Optimal (fwt=∞)	83.2%	86.3%	88.4%	1.4%	85.9%	86.7%
Consolidation Optimal (fwt=0)	80.4%	83.6%	85.4%	1.3%	83.2%	83.9%
Model Single Shipments (CT)	66.9%	70.8%	74.0%	1.5%	70.4%	71.1%
Model Consolidation (NN, fwt=0)	77.0%	83.1%	85.2%	1.8%	82.6%	83.6%
Model Consolidation (CT, fwt=0)	80.2%	83.1%	85.4%	1.5%	82.8%	83.4%
Model Consolidation (NN, fwt=∞)	76.3%	81.2%	83.7%	2.0%	80.5%	81.9%
Model Consolidation (CT, fwt=∞)	67.6%	79.1%	83.7%	3.5%	78.3%	79.7%

Table 38: Constraint feasibility lane 3

	Hazard Feasibility		Lead Time Feasibility		Capacity Feasibility		Dimensions Feasibility	
	Mean	**Std.**	**Mean**	**Std.**	**Mean**	**Std.**	**Mean**	**Std.**
Model Consolidation (NN, fwt=0)	100%	0%	100%	0%	100%	0%	100%	0%
Model Consolidation (CT, fwt =0)	99.9%	0%	99.9%	0%	100%	0%	100%	0%
Model Consolidation (NN, fwt =∞)	100%	0%	100%	0%	100%	0%	100%	0%
Model Consolidation (CT, fwt= ∞)	99.9%	0%	99.9%	0%	100%	0%	100%	0%

In Table 37 can be seen that with all consolidation models higher savings can be obtained than with the single shipment model. Both same day consolidation models have on average 83.1% cost savings, and therefore outperform both models with multiple day consolidation. The standard deviations for the same day consolidation models are in line with the standard deviation obtained with the deterministic model, which means the models perform stable. While the standard deviation for the models with consolidation over multiple days is higher, compared to the results obtained with the deterministic model for this scenario. From Table 38 becomes clear that all constraints are met for all models almost 100% of the time.

Lane 4

The results regarding the cost savings are presented in Table 39, and results regarding the constraint feasibility are presented in Table 40.

Table 39: Cost savings lane 4

	Min.	Mean.	Max.	Std.	95% confidence interval	
					Lower	**Upper**
Single Shipments Optimal	51.3%	65.8%	74.3%	5.1%	64.4%	67.3%

Consolidation Optimal (fwt=∞)	73.8%	79.0%	82.2%	2.2%	78.4%	79.7%
Consolidation Optimal (fwt=0)	57.4%	71.5%	79.1%	6.5%	69.6%	73.3%
Model Single Shipments (CT)	48.8%	63.6%	73.2%	5.5%	62.0%	65.2%
Model Consolidation (NN, fwt=0)	56.6%	71.1%	78.9%	5.9%	69.4%	72.8%
Model Consolidation (CT, fwt=0)	56.9%	72.9%	82.8%	6.6%	71.6%	74.2%
Model Consolidation (NN, fwt=∞)	55.3%	73.5%	82.7%	5.5%	72.0%	75.1%
Model Consolidation (CT, fwt=∞)	53.2%	71.9%	82.6%	6.7%	70.6%	73.3%

Table 40: Constraint feasibility lane 4

	Hazard Feasibility		**Lead Time Feasibility**		**Capacity Feasibility**		**Dimensions Feasibility**	
	Mean	**Std.**	**Mean**	**Std.**	**Mean**	**Std.**	**Mean**	**Std.**
Model Consolidation (NN, fwt=0)	100%	0%	100%	0%	100%	0%	100%	0%
Model Consolidation (CT, fwt =0)	100%	0%	100%	0%	100%	0%	100%	0%
Model Consolidation (NN, fwt =∞)	99.9%	0.1%	100%	0%	100%	0%	100%	0%

Model Consolidation (CT, fwt= ∞)	99.9%	0.1%	100%	0%	100%	0%	100%	0%

In Table 39 can be seen that with all consolidation models higher savings can be obtained than with the single shipment model. The most cost savings (73.5%) on average can be obtained with multiple day consolidation with the *ANN* model. With the *CT* model (71.9%) significantly less savings are obtained on average compared to the *ANN* model. For same day consolidation; the *CT* model scores slightly better with cost savings of 72.9% compared to cost savings of 71.1% with the *ANN* model. The standard deviations for same day consolidation obtained with the machine learning models, are in line with the standard deviations obtained with the deterministic models. This implies that the machine learning models perform stable. However, for consolidation over multiple days the standard deviation obtained with the machine learning models is almost twice as high, as the standard deviation obtained with the deterministic model. And therefore, these models have a variation of quality in their consolidation planning decisions. In Table 40 can be seen that all constraints are met for all models almost 100% of the time.

Lane 5

The results regarding the cost savings are presented in Table 41, and results regarding the constraint feasibility are presented in Table 42.

Table 41: Cost savings lane 5

	Min.	**Mean.**	**Max.**	**Std.**	**95% confidence interval** **Lower**	**Upper**
Single Shipments Optimal	53.6%	62.9%	70.7%	3.7%	62.2%	63.7%
Consolidation Optimal (fwt=∞)	62.3%	71.1%	79.2%	4.0%	70.3%	71.9%
Consolidation Optimal (fwt=0)	62.0%	69.9%	77.8%	4.1%	69.1%	70.8%
Model Single Shipments (CT)	49.0%	60.2%	68.1%	4.0%	59.4%	61.0%
Model Consolidation (NN, fwt=0)	57.2%	67.4%	76.2%	4.6%	66.1%	68.8%

Model Consolidation (CT, fwt=0)	57.3%	67.5%	74.9%	3.8%	66.8%	68.3%
Model Consolidation (NN, fwt=∞)	58.4%	66.3%	71.7%	3.9%	64.5%	67.2%
Model Consolidation (CT, fwt=∞)	51.3%	63.5%	73.2%	4.9%	62.6%	64.5%

Table 42: Constraint feasibility lane 5

	Hazard Feasibility		**Lead Time Feasibility**		**Capacity Feasibility**		**Dimensions Feasibility**	
	Mean	**Std.**	**Mean**	**Std.**	**Mean**	**Std.**	**Mean**	**Std.**
Model Consolidation (NN, fwt=0)	100%	0%	100%	0%	100%	0%	100%	0%
Model Consolidation (CT, fwt =0)	100%	0%	100%	0%	100%	0%	100%	0%
Model Consolidation (NN, fwt =∞)	100%	0%	100%	0%	100%	0%	100%	0%
Model Consolidation (CT, fwt= ∞)	100%	0%	100%	0%	100%	0%	100%	0%

In Table 41 can be seen that all models yield extra saving over the single shipments model. Most cost savings are obtained with the same day consolidation scenario. With both models about the same cost savings can be obtained; 67.4% for the *ANN* model and 67.5% for the *CT* model on average. For consolidation over multiple days; the *ANN* model approaches the cost savings obtained with same day consolidation, with cost savings of 66.3% on average. Whereas with the *CT* model cost savings of 63.5% on average are obtained. For all models the variation is in line with variation obtained with the deterministic model. This implies that the models run stable. In Table 42 can be seen that for all models the constraints are satisfied with 100%.

S: RESULTS PER TRANSPORT LANE

Table 43: Costs and relative savings per transport lane given a single shipment policy and consolidation policy

	Costs			Relative savings
	Single shipments (current business rules)	**Single shipments (new business rules)**	**Consolidation (new business rules)**	**Consolidation (new business rules)**
Lane 1	15,440 EUR	10,567 EUR	4,045 EUR	73.8%
Lane 2	10,273 EUR	9,113 EUR	6,790 EUR	33.9%
Lane 3	12,913 EUR	3,529 EUR	2,182 EUR	83.1%
Lane 4	13,696 EUR	4,740 EUR	3,711 EUR	72.9%
Lane 5	35,605 EUR	11,996 EUR	11,571 EUR	67.5%