Eindhoven University of Technology

MASTER

Predicting delivery time windows at a freight distribution company

a case study

Bourkha, W.

Award date:
2017

Link to publication

Disclaimer
This document contains a student thesis (bachelor's or master's), as authored by a student at Eindhoven University of Technology. Student theses are made available in the TU/e repository upon obtaining the required degree. The grade received is not published on the document as presented in the repository. The required complexity or quality of research of student theses may vary by program, and the required minimum study period may vary in duration.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 16. Jul. 2025

Department of Industrial Engineering & Innovation Science
Information Systems Group

Predicting delivery time windows at a freight distribution company

- A case study

by:
W. (Walid) Bourkha
Student number: 0822519

Supervisors:

dr. ir. R.M. (Remco) Dijkman	*TU/e IE&IS Information Systems Group*
dr. O. (Oktay) Türetken	*TU/e IE&IS Information Systems Group*
dr. ir. H. (Rik) Eshuis	*TU/e IE&IS Information Systems Group*
MSc. N. (Niels) Hoefs	*Manager Staff Operations Cargo*

in partial fulfilment of the requirement for the degree of:

Master of Science in Operations, Management and Logistics

Eindhoven, Monday 17th April, 2017

TUE. School of Industrial Engineering
Series Master Theses Operations Management and Logistics

Subject headings: delivery time window prediction, estimated time of arrival.

Abstract

This research focusses on the prediction of delivery time windows. These predictions are important for companies since part of their customer satisfaction relies on delivery time reliability. In order to increase the reliability of the time window estimates two different techniques are tested in a case study. The case study is done at the Cargo department of a freight distribution company, this department is specialized in last mile delivery of cargo. The performance in terms of the number of orders delivered within a three-hour time window is analysed and improved by using data-driven techniques.

Based upon the related work in the field of delivery time prediction, Multi-Linear Regression and Artificial Neural Network (ANN) are used to obtain time window estimates. Before the techniques are tested in a case study, the initial situation is analysed in order to evaluate the performance of the techniques. In addition, the current performance is evaluated using naïve and adjusted time windows. Next, prediction models are developed based on the two techniques. Different parameters are selected and tested in the model to improve the prediction of delivery time windows.

The extension using adjusted time windows in the initial situation resulted in an average performance improvement of 18% for the three-hour time window. For the case study, implementing the two techniques results in a great performance improvement. The regression technique performs 20% better compared to the initial performance. Implementing the ANN model increases the performance with 24%. Finally, for the case study company a tool is developed that measures the delivery performance.

Executive Summary

Businesses expect that their orders are delivered at the exact time promised by their carrier. Delayed shipments are considered unacceptable, since this can affect for example the production processes. Delivery time reliability is therefore one of the most important factors in the B2B industry. It helps that in B2B markets the buyer and seller operate as a business partner. This results in planned and structural deliveries. In order to improve the delivery estimates, this project develops two models for the prediction of delivery time windows.

A literature study is conducted to get insights in the different techniques available for predicting delivery time windows. Based upon an overview of the related works in the field of delivery time window prediction, two techniques are selected to apply in a case study. Machine learning methods as Multi-Linear Regression and Neural Network are applied in a case study to predict future delivery time windows. The characteristic of the first technique is that its representation is simple and can be easily implemented in daily practice. Neural Networks are used, since literature suggest that they have a major application in forecasting, and are well suited for predictions if there is enough data available.

Before the proposed techniques are tested, the initial situation at the case study company is analysed using a naïve- and an optimal adjusted time window. The three-hour naïve time window builds a time window arounds the expected time of arrival (ETA) at the customer by adding 1.5 hours and extracting 1.5 hours from the ETA. For the adjusted time window an optimal setting of the window is calculated in order to achieve the highest percentage of orders delivered in this three-hour window. The current situation is analysed because the performance of the techniques are evaluated using the initial situation as a benchmark. The case study company uses a partner network for delivering their orders, therefore the performance per partner is analysed separately. The performance metrics of the initial situation are given in *Table 1*. It can be seen that for some partners the difference between implementing a naïve or an adjusted time window has a great impact on the performance. Overall, the adjusted window performs 18% better compared to the naïve window. This is due to the fact that the distribution between early and late delivered orders is not equally divided.

Partner	Three-hour naïve Window	Three-hour Adjusted Window	MAE	MAPE
Partner A	88%	88%	46.4	3.22%
Partner B	51%	76%	102.6	7.13%
Partner C	66%	69%	82.1	5.7%
Partner D	49%	66%	108.4	7.53%
Partner E	66%	77%	79.1	5.5%

Table 1 - Results for all the Partners in the initial situation

Next, two models are developed based on the two techniques. These models are built using different parameters for each partner, since the initial performance shows great variance between the partners. The most important performance metric used for evaluating the performance of the models is the percentage of orders delivered within a three-hour time window. Implementing a multi-linear regression model results in an average performance improvement of 20% compared to the current situation. The mean absolute error (MAE) stands for the number of minutes the predicted time of arrival varies from the actual time of delivery. Using the regression model the MAE decreases with 23.1%.

The next model built is based on the Artificial Neural Network (ANN) technique. This model uses two hidden layers for every partner. However, the number of hidden nodes used per layer differs per partner. This is done because the performance improved when using different number of nodes per layer. Compared to the current situation, the developed Neural Network model performs on average 24% better. The MAE decreased on average with 25.5%. The summarized performance improvement compared to the initial situation is presented in *Figure 1*.

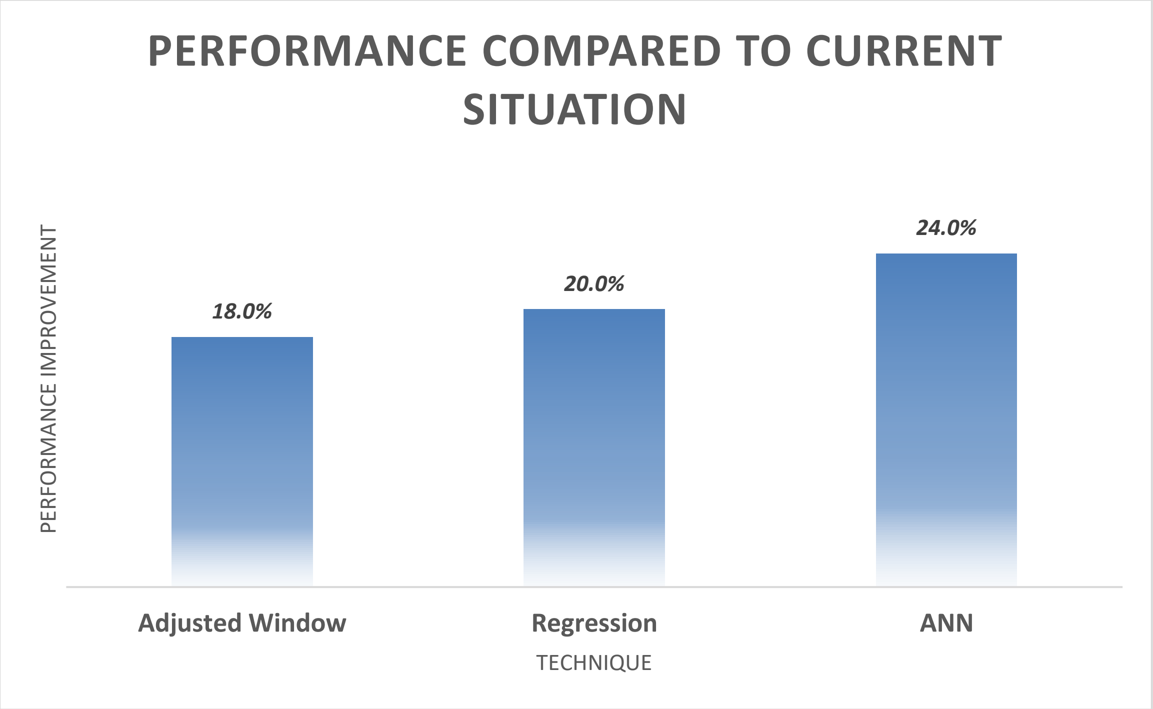


Figure 1 - Comparison between the techniques and initial performance

The performance improvement of the implemented techniques suggest that there is a great opportunity for the case study company to improve their delivery performance. Therefore, a tool for measuring the delivery performance of the partners is built using VBA. The choice for this software is because it is the most practical and suitable way, since other people will work with this tool. The main function of the tooling is to present what the performance is, and how to improve it.

Finally, practical implications for the case study company are presented. An interesting insight is the fact that when implementing an adjusted time window instead of using a naïve time window, the performance gets close to the output of the regression technique. Therefore, implementing an adjusted time window would result in a great performance improvement. In addition, its implementation in daily practice is easy. Second, the tooling can be used on daily basis to communicate possible improvements with the partners. Since, this gives them the possibility to see what specific orders are delivered outside the time window.

Preface

This thesis is the end result of six months hard work and research and also marks the end of my student career. This result would not have been possible without the support of many people.

First of all, I would like to thank Remco Dijkman for mentoring me during my graduating period. During our meetings you provided me with valuable insights, constructive feedback and support to point me into the right direction. It was a pleasure to work under your supervision to this end result. Also I would like to thank Oktay Türetken as the second supervisor for my thesis.

I also want to thank my company supervisor, Niels Hoefs. He provided me with good practical ideas and made me feel like a full team member. You supported me by giving me the resources, the opportunity to implement my ideas, and involving me whenever possible. Thank you for all of this.

Next, I would like to give my special thanks to my parents for supporting me during my whole student career and giving me the opportunity to pursue a good education. The rest of my family – especially my brothers – tanks you for always supporting me and being there for me whenever I needed it most.

The last thing I want to say is that I really enjoyed writing this thesis. The project made me very ambitious since it covers all the aspects I aspire. It gives great pleasure when a model works and the implementation pays off directly. All in all, doing this project was a valuable experience I would carry on for the rest of my life.

I wish you much pleasure reading this!

Walid Bourkha

17th April 2017

Table of Contents

List of Figures

List of Tables

List of Acronyms

ACS	Ant Colony System
ADR	Agreement Concerning the International Carriage of Dangerous Goods by Road
ANN	Artificial Neural Network
AVL	Automatic Vehicle Location
B2B	Business-to-Business
B2C	Business-to-Consumer
CRISP-DM	Cross-Industry Standard Process for Data Mining (Crisp-DM)
DARP	Dial-a-Ride-Problem
DSRP	Design Science Research Process
DVRP	Dynamic Vehicle Routing Problem
ETA	Expected Time of Arrival
GA	Genetic Algorithm
GGA	Grouping Genetic Algorithm
IS	Information Systems
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MAS	Multi-Agent-System
MIP	Mixed Integer Programming
MLP	Multi-layer Perceptron
MPE	Mean Percentage Error
MSE	Mean Squared Error
MSSG	Multi-strategy Grouping Genetic Algorithm
PDF	Probability Density Function
PDP	Pickup and Delivery Problem
PDPTW	Pick-up and Delivery with Time Windows
RC	Roll Container
RMSE	Root Mean Squared Error
TMS	Transport Management System
VRP	Vehicle Routing Problem

Chapter 1

Introduction

This section elaborates the context of this project. First the motivation for this research is given, followed by background information about the research topic, which describes the research context. Thereafter the research goals and the research design of this project are discussed. Finally, a short introduction to the case study company is given, and the thesis structure is presented.

1.1 Motivation

Nowadays companies are continuously trying to reduce costs and improve their performance. One way to achieve this, is by improving customer satisfaction. Aranko (2013) stated that companies and consumers increasingly purchase goods online, resulting in a growth for expressed delivery service. The increased demand for instant and accurate delivery status is driving a dire need to ensure shipment integrity in the parcel supply chain. Therefore, streamlined efficiency in the parcel industry is more important than ever. Effective construction of transport planning allows companies to limit costs and be more competitive on the market.

Increasing customer satisfaction is priority number one in the parcel and cargo industry (Li et al., 2006). Part of this customer satisfaction relies on delivery time reliability. Delayed shipments are considered unacceptable by consumers, making customer service one the most important factor in delivery operations. To achieve this, prediction of delivery time windows should be done more accurately. Research in predicting delivery time windows should be done in order to get more insights in the possibilities of implementing reliable and shorter delivery time windows for the customer. Despite this importance, only a few research activities cover the perspective of road freight carriers and develop measures to manage the rising challenges (Elbert et al., 2016). Therefore, this report will focus on generating a method for predicting delivery time windows in the transportation industry. These predictions will be data-driven, rather than based on human experience.

Nowadays, businesses accumulate a lot of data with potential value that must be extracted. The Transport Management System (TMS) saves a lot of information about the process details, routings and durations of each step in the process. However, humans cannot cope with this large amount of data, but computers can. Therefore, this research will try to improve the prediction of delivery time windows at a freight distribution company based on historical data. One of the drivers of the research is the desire of the case company itself to gain insights in their current performance of time window estimates and improve this. The goal is not solely to research the problems at hand but also to provide a valuable backbone research of the basics that are related to delivery time window prediction.

1.2 Background

The goal of this research is to gain insights into the techniques and methods for predicting delivery time windows and apply them in a case study. Therefore, it is important to understand the general overview of the encompassing fields and the target segment.

There are two different target segments for parcel service providers; the business-to-consumer (B2C) and business-to-business (B2B) market. According to Duin et al. (2015) there are considerable differences between these segments. In B2C-networks the demand is more fluctuating and therefore less predictable. In contrast, in B2B markets the buyer and seller are operating as a business partner resulting in planned and structural deliveries. Therefore, it must be stated that the delivery efficiency must be much higher in B2B- than in B2C-markets. As stated before, this research focusses mainly on B2B deliveries.

The research takes place in the context of delivery time windows. The concept of delivery time window is described as “the earliest acceptable delivery time and the latest acceptable delivery time” by Bushuev & Guiffrida (2012). Within the transportation sector delivery time windows are provided in order to indicate the estimated window of arrival at the customer. When an order is placed, the customer is typically given a fixed due time for expecting the delivery. Under the concept of a delivery time window, the customer is supplied with a timeslot. Here, the deliveries can be classified as being early, on-time, and late. In order to illustrate this in *Figure 2* a visualization of the delivery time window concept is provided.

When considering the modelling of time window problems in operations research, these classified deliveries can be seen as ‘soft time windows’. Which actually are presented as a service to the customers and therefore need to be adhered to as much as possible. On the other hand, ‘hard time windows’ describe the constraints which have to be met in order to reach a valid solution. For instance, the restriction imposed by the customer for the opening hours of the site.

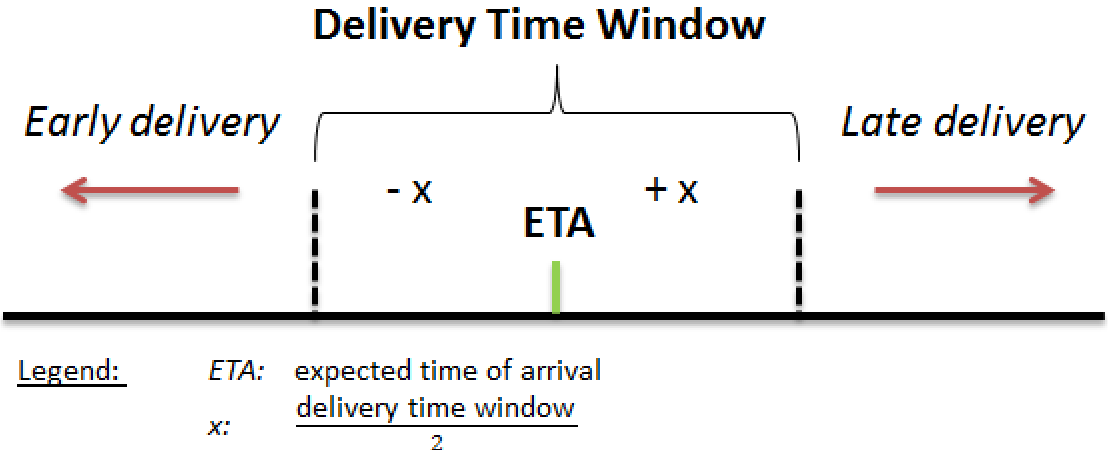


Figure 2 - Time-window delivery concept

The attractiveness of delivery time windows was researched by Corbett (1992) who demonstrated how a furniture manufacturer placed a higher value on delivery reliability, which refers to the on-time delivery within a window, than on delivery speed. Boyer et al. (2009) identified delivery time windows as a key factor in modelling the “last mile challenge” in supply chains. To achieve this the expected time of arrival (ETA) should be as close as possible to the actual arrival time. This implies that the error of the ETA should be minimized. Besides the effect on the customer service level, time window constraints also have a growing influence on the costs of road transport services. Therefore the focus of this research is on improving the delivery time window estimates.

Time window estimates are a performance measure for delivery processes. An example of a parcel delivery service process is provided in *Figure 3*. This process is divided in three steps; collecting, sorting and delivery (Jung et al., 2006). The collecting process refers to the pickup of the order a the shipper and transport it to a terminal (Departure Terminal). Second, the goods are transported by a linehaul, which refers to the tranport between two terminals, to another terminal (Arrival terminal). Lastly, the 'last mile' delivery is started by the service provider. This research focuses only on the red part of the figure, which represents the distribution process. In order to achieve streamlined effiency between these processes, the parcel delivery services need to focus on logistic planning for operating efficiently.

Figure 3 - General overview of parcel delivery service process

For the prediction of delivery times there are different techniques available. These techniques can be classified in *operations research* and *data mining*. Data mining is used as a synonym for knowledge discovery from data, it is associated with terms as patterns analysis, knowledge extraction, data archaelogy, and many more (Han et al., 2011). Data mining is an essential process that uses techniques to extract patterns from data companies have accumulated in the past. On the other hand, operations research can be seen as a scientifiq method of providing departments with a quantitative basis for complex decision making (Morse & Kimball, 2012).

Another important consideration is the model in which the time window predictions must be done. In this research, the focus is on pickup and delivery problems (PDP). The case study company is an example of a freight carrier that handles this problem in daily practice. However, there are many variations and approaches developed of the PDP. Therefore, the focus is on the PDP with time windows (PDPTW). This variation arises mainly in the context of urban courier services. The PDPTW concerns a vehicle that must satisfy a set of transportation requests on his route. The delivery process starts at the distribution centre, where the forwarder loads his orders. After loading the orders, the courier satisfies a set of nodes in a specific time window and returns to the depot when finished. This process is applicable for all express delivery companies if capacity constraints are neglected.

1.3 Research Goals

The goal of this research is to improve the prediction of reliable time window estimates for the delivery of customer orders in the freight transportation industry. The applied method should generate reliable time windows for the customer. The end result is validated empirically in a case study. From this research goal the main research question can be derived. The main research question in this thesis is:

"How can accurate and reliable prediction estimates for delivery time windows be generated?"

In order to answer the main research question the following sub-research questions are formulated:

[1] Which methods are currently available for prediction of delivery time windows?

[2] Which factors influence the actual ETA?

[3] How can the prediction of delivery time windows be made more accurate?

[4] How does the implemented method perform in a case study?
- What is the current performance?
- What is the performance of the imposed method?
- Which practical implications are there for the case study company?

The scope of this research is limited to the prediction of delivery time windows. The resource planning is out of this scope for this research.

1.4 Research Design

The research design aims to answer the research questions mentioned before. This research design is based on the Design Science Research Process (DSRP) according to Peffers et al., (2006). This conceptual design science process model helps to produce and present high quality research in information systems (IS). This design is combined with the CRISP-DM framework, which will be used in the demonstration phase for data analysis.

The DSRP consists of six steps, where the first two steps are already described in section 1.2 and 1.3 and the last step is not relevant. The other steps are presented as the research design in *Figure 4*. First a literature study is conducted to get insights in the different techniques available for predicting delivery time windows. Multiple generic techniques and specific methods are uncovered to get an overview of the academic research field. After clear insights in the current state of research are provided, two techniques are proposed and reviewed. This will be done in the Design & Development part of the DSRP, where research question [1] will be answered.

The second research question will be answered with input from literature and experienced employees from the case study company. Using these two methods for assessing which factors will influence the actual ETA will result in a good trade-off between theory and practice. This is necessary because not every variable that literature suggests, will be available in the provided dataset by the case study company. This answers research question [2].

After assessing which techniques will be used and which variables could influence the actual ETA, the optimal setting for applying the techniques in a case study is evaluated. This is done by building the models with data from high quality and testing different combination of parameters. The modelling part of the CRISP-DM will give insights on how to make the prediction of delivery windows more accurate and results in answering research question [3].

In the demonstration process the proposed techniques are applied in a case study. First the actual performance at the case study company is reviewed. Thereafter, the performance of the proposed techniques is measured on a real dataset. In order to do this, the CRISP-DM framework is used as a guide. Next, the tool for measuring the delivery time window performance is presented.

Finally, the proposed method is evaluated by comparing the generated delivery time windows with the actual performance. In this way research question [4] will be answered. In addition, feedback about the overall research and practical implications are provided.

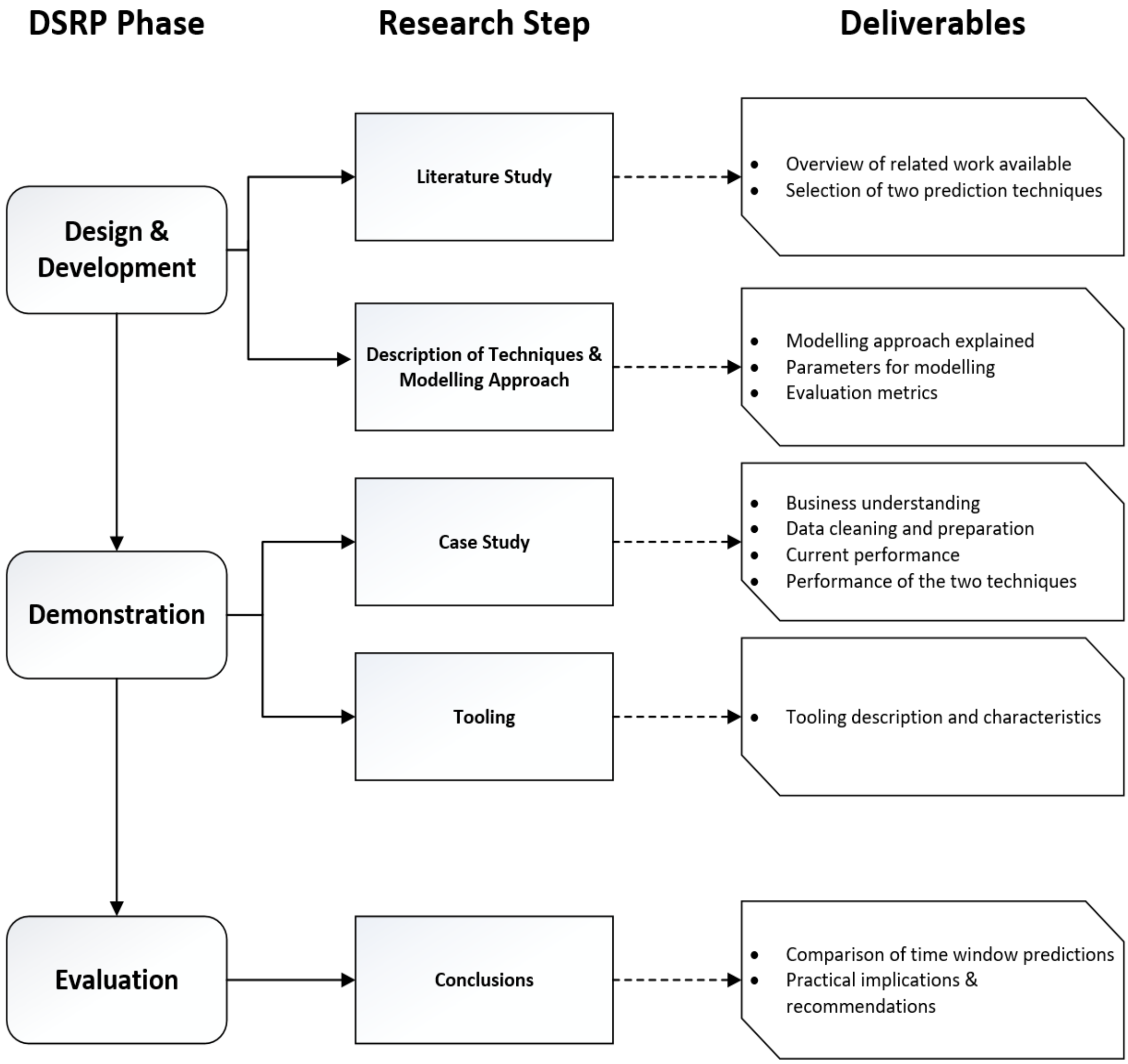


Figure 4 - Research design

1.6 Thesis structure

Figure 4 represents the design of the remainder of this thesis. In Chapter 2 a literature review of the existing techniques for predicting delivery time windows is presented. The two selected techniques and their modelling approach are discussed in chapter 3. Chapter 4 consists of a case study to validate the proposed techniques. Besides, a detailed evaluation is done by comparing the performance of the proposed method with actual performance. In addition, Chapter 5 explains the characteristics of the developed tool. After all research steps are discussed an overall conclusion and potential applications for the case company are presented in Chapter 6. This chapter also covers the limitations of this research and directions for future research.

Chapter 2

Techniques for Time Window Prediction

This chapter handles the Design & Development phase of the DSRP. In this section a cross-analysis of related works on the field of techniques for predicting delivery time windows is presented. The goal of this chapter is to give an overview of the techniques and methods used in past research to predict delivery time windows. The related works are found using the literature study. It provides a good overview of the available literature on delivery time windows. Lastly, two methods for predicting delivery time windows are discussed.

2.1 Approach

In this section, the approach of the literature study is presented. In order to give a complete overview of the encompassing field, the review approach is explicitly elaborated.

2.2.1 Sources

Regardless of the subject you study, there is no single source of information that will contain everything you need for your literature search. Therefore, to achieve the goal of this literature review, multiple search engines are used to create an overview of the research area. The databases contain a large range of academic subjects that will facilitate in finding information to the relevant subject of interest. These databases (*Table 2*) are selected to complement each other. Web of Science and JSTOR cover the journals in the research area, ACM and Springer the conferences. Science Direct and Springer have a focus on books. The combination of these five search databases gives a good overview of the research area in which this literature review is interested in.

Search Engine	Indexes[1]
Web of Science	- 12,000+ journals - 160,000 conference proceedings
Springer journals	- 9,000,000+ journals - 2,400 Journal Articles
ACM Digital Library	- 407,367 Full-text articles - 2,000+ Proceedings Volumes
JSTOR	- 2,000+ journals - 50+ disciplines
Science direct	- 3,800+ journals - 35,000+ book titles

Table 2 - Search engines and their indexes

---

[1] As per October 10th, 2016

2.2.2 Search Terms

For the creation of the search terms, the research goal of the project is used as a starting point. From there the following keywords are used:

1. Delivery Time Window (Including synonyms and variants
2. Prediction (including synonyms and variants)
3. Transport (including synonyms and variants)
4. Freight (including synonyms and variants)

The term Delivery Time Window (1) aims to the search of articles that are conducted in this research area, which is the topic of interest. The goal of the literature review is to find reviews that focus on the prediction of delivery time window given historical data and the expected delivery time, therefore search element (2) is included. In addition, the focus should be on the transportation of freight, search element (3) and (4) satisfies this. All these elements should exist in the potential articles. The 'AND' operator is used to connect these elements. The extension with the transport and freight variant leads to very specific research in this industry. However, the results are broad enough. In addition, all the separate elements mentioned in *Table 3* are combined in a single search query.

Index	Keyword	Synonyms and variants	Source
1	Delivery Time Window	Delivery Time Window	*n.a.*
2	Prediction	Forecast(ing), estimate, indicator	Thesaurus (Dictionary.com, 2016)
3	Transport	Logistics, shipping	Thesaurus (Dictionary.com, 2016)
4	Freight	Shipment, carriage, packages	Thesaurus (Dictionary.com, 2016)

Table 3 - Search terms

2.2.3 Selection Criteria

The study selection criteria are used to determine which studies are included or excluded from a systematic literature review. It is also intended to identify primary studies that provide direct evidence about the research question. The criteria are split into two categories; accessibility and content criteria.

Accessibility criteria

The accessibility criteria are associated with:

1. Use of English language
2. Accessibility to full text

The accessibility criteria exclude articles that are not written in the English language and documents that are not available in full text. The first criteria only includes articles written in English. Language problems could arise when considering articles written in other languages. However, this restriction should not be a problem since English is used as the standard language for scientific literature.

The second criteria is added because of access reasons. Articles are excluded if there is no access to the full text. Without the full text, important information regarding the topic of interest could be missing in the article.

Content criteria

The content criteria chosen are:

1. Recent articles, not older than ±20 years (year of publication 1995 or later)
2. Goal of the article is to predict delivery time (windows) by using a method or technique
3. The article contains an evaluation
4. The article is applicable to the transportation industry

The first content criteria limits the search space to only recent articles, since the area of transportation has evolved largely over the years, which makes older articles quickly outdated. The second criteria is that the goal of the literature should be aligned with finding methods or techniques used for predicting delivery time windows, such that the applied literature can be used to address the research question. Criteria (3) and (4) are from a content point of view. The goal of this literature study is to give an overview of the methods and techniques used for predicting delivery time windows in scientific literature. To be able to get an overview of the practical implementations, the literature should be applicable to the transportation industry. Besides, the article should include an evaluation of the addressed method or technique used in practice. In this way, the performance and insights can be easily evaluated with other articles.

2.3 Global Classification

The previous section elaborated on the search protocol. This section presents the results that are obtained by using the search terms in the chosen search engines. The goal is to get a general overview of the complete research area. To achieve this, multiple search engines are used. The combination of these five search databases gives a reasonable overview of the research area. There are in total 56 hits found in the five different databases. The abstract and conclusion of these articles are read to decide whether to the articles are good enough to answer the research questions. Additionally, to that, the criteria for inclusion/exclusion are applied to decide which articles are included for further analysis.

After applying the selection criteria, 7 of the 56 hits yielded in the previous section are added to the article pool. By looking into the selected articles, the article pool can be extended by looking at the references of the reference. For example, this can be useful if a review article is found that evaluates other literature in the field. Another way to extend the article pool is by looking at other works created by the author cited. This in order to find relevant articles in the field of interest. Based on the selection criteria stated in *section 2.2.3*, another 7 papers were added. With the addition of these relevant articles, the total number of articles comes to 14.

The research question specifies to find the methods and techniques used in predicting delivery time windows, therefore the techniques used in the articles are first analyzed. Because of this, all these methods and techniques will be structured in a model to give an overview of the different techniques that are used in predicting delivery time windows. The next step is to classify the dimensions used for analyzing the articles, which will be done after classifying the different techniques.

As stated in the introduction, only a few research activities cover the perspective of road freight carriers. This is also the fact with research about techniques used in the prediction of delivery time windows. Therefore, it is difficult to give an overview of all the techniques used in the encompassing field.

The research has as main goal to get insights in the techniques used for predicting delivery time windows. The articles from the article pool in *Table 4* show surprising results concerning the methodology used. After analyzing every article, it can be stated that the article pool contains a great variance of techniques used. Almost every paper focuses on his own technique, this results in a broad overview of the techniques used for predicting delivery time windows. It can also be distinguished that some articles use hybrid methods, this means that they use a combination of two or more techniques in their research. The categories classified are presented in a tree model in *Figure 5*. The top row shows the general techniques used in the articles found and functions as a source for the rest of the tree.

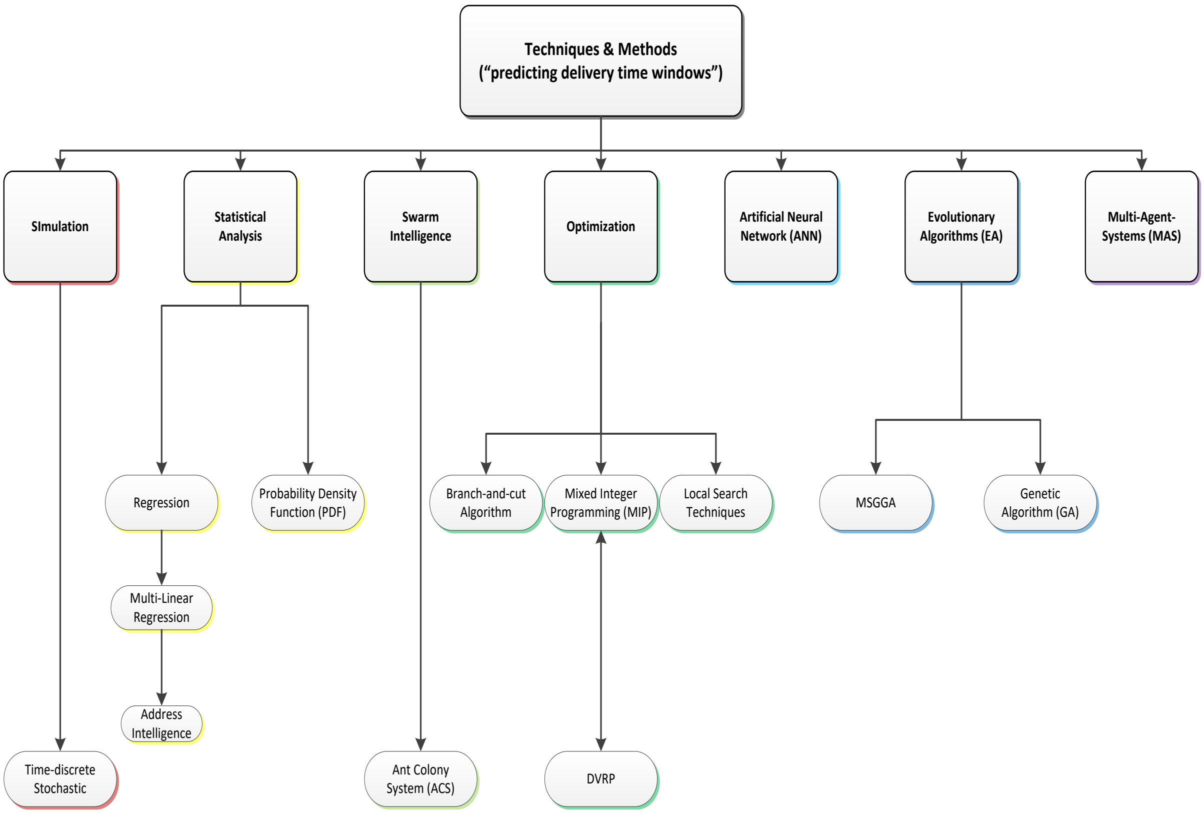


Figure 5 - Used methods and techniques tree model

2.4 Selection of Methods and Techniques

For the selection of methods and techniques that are used in this research, the articles are assessed on the following dimensions:

- Requirements
- Prediction technique used
    - Regression or ANN
- Performance of the technique and evaluation
- Implementation in practice

The tabular model created with the analysis of the articles (*Table 4*), describes the content of the articles in detail. The first column stands for the different articles. Next, the other columns assess all the articles on the different dimensions named above. Since the focus in this research is on predicting delivery time windows using historical based data, the first dimension should state operational data as a requirement. Besides, the scope of the research is on the prediction of time windows, rather than on the resource planning of the network. Therefore, articles that use a network as a requirement for the method or technique can be neglected.

A distinction is made when considering the methods and techniques that are used in the articles. The focus in this research will be on two specific predicting techniques; regression and ANN. These techniques aim at making predictions for any target variable by learning from a dataset. Regression is an attractive model because the representation is simple and can be easily implemented in daily practice. ANNs are used for the fact that they have a major application in forecasting (Sharda, 1994), and are well suited for predictions if there is enough data available.

Since this research will consider a case study implementation, it would be an advantage if the proposed method or technique is already implemented in practice. In addition, the performance of the technique in combination with the experiment used for evaluation has also influence on selecting the initial method or technique to use for this research. This implies that articles that achieve a low performance after implementing their technique will be neglected. The article of van Duin et al. (2015) makes a good technique to use for an operational business case (like predicting of delivery time windows). It uses historical delivery data to predict future delivery results by address intelligence. In addition, this technique was also implemented at a logistics parcel service company (DHL). As can be seen in *Table 4*, Bushuev & Guiffrida (2012) scores exactly the same evaluation points as van Duin et al. (2015). However, the paper of van Bushuev & Guiffrida (2012) focusses on the prediction of the time window for the whole supply chain, instead of the last mile delivery only. Therefore, this paper is excluded in this research.

On the other hand, the paper of Jeong & Rilett (2004) uses a Regression Model, and a Artificial Neural Network for predicting arrival times. The techniques were also tested in a case study which makes the paper suitable for practical implications in this research. In addition, the paper of Zhang et al. (1998) deals with the interesting application of ANNs in forecasting. The goal of this paper was to provide a synthesis of published research in the area of forecasting, and give insights on ANN modelling issues. Since this literature selection focusses on the prediction of delivery time windows, implementing ANN in a case study can be interesting. In the next session the tehniques are more elaborated to give insights in their practical applications.

Table 4 - Overview of the related work analysis

				Prediction Technique						
Article	**Requirement**	**Case Study**	**Performance**	**Statistical Analysis**	**Simulation**	**Optimization**	**Swarm Intelligence**	**ANN**	**EA**	**MAS**
Van Duin et al. (2015)	O	X	X	**X**						
Elbert, Thiel, & Reinhardt, (2016)	O	/	/		X					
Jeong & Rilett, (2004)	O	X	X	**X**				**X**		
Bushuev & Guiffrida, (2012)	O	X	X	**X**						
Zhang et al., (1997)	O	/	X					**X**		
Kozlak et al., (2006)	N	/	/							X
Hosny & Mumford, (2007)	N	/	X						X	
Du, Li, & Chou, (2005)	N	/	/			X				
Ropke, Cordeau & Laporte, (2007)	N	/	/			X				
Fabri & Recht, (2005)	N	/	X			X				
Ghannadpour et al, (2014)	N	X	/						X	
Genhong et al. (2009)	N	/	X						X	
Sprenger & Mönch, (2009)	N	X	X				X			
Lin, (2011)	N	/	/			X				

Legend:

Requirements column	O: Operational data N: Network
Case study column	X: Implemented in practice /: No implementation in practice
Performance column	X: Technique evaluation and performance good /: Technique evaluation or performance not sufficient

2.5 Regression Technique

This section aims to elaborate on the regression technique. For this research, this technique will function as an initial starting point for predicting delivery time windows. After the arguments and practical applications using input from article(s) are clear, the proposed technique can be implemented in a case study.

The first model that is developed for application in the case study is the multi linear regression model. This model establishes a relationship between the target variable and the variables used for prediction. Van Duin et al. (2015) uses the application of multiple linear regression to support the development of address intelligence. The focus in this paper is on improving the home delivery efficiency. However, the usage of address intelligence can also be applied in this research for predicting delivery time windows. By using the concept of address intelligence, the relation between zip code areas can be analysed. In addition, the regression model can be used for predicting the delivery time windows.

The regression model developed will be implemented in a case study. Before the final regression model can be applied, the parameters used must be evaluated on different dimensions. This research uses the same evaluation on the regression model as (Jeong & Rilett, 2004). The researchers used 0.15 as the cut-off value of the correlation coefficient for the input parameters. In addition, the parameters are tested on statistical significance before the final model is developed. The parameters that are not significant are excluded from the final prediction model that is used in the case study.

2.6 Artificial Neural Network (ANN)

The second technique that is used for predicting the delivery time windows is ANN. This technique emulates the biological learning process of the human brain (Jeong & Rilett, 2004). In addition, ANNs have powerful pattern recognition, pattern classification capabilities, and are able to learn from and generalize from experience. Zhang et al. (1998) made a ‘state of the art analysis’ related to forecasting using neural networks. Some of their findings are used to develop and improve the models.

ANNs are used for the fact that they have a major application in forecasting (Sharda, 1994). Lapedes (1987) and Faber (1988) reported one of the first successful applications of ANNs in forecasting. They used a multi-layer network and their results showed that ANNs can be used for forecasting time series with very high accuracy. In this research, the goal is to forecast delivery time windows, which are difficult to specify. However, ANNs are well suited for these objectives if there are enough data or observations. This modelling approach with the ability to learn from experience is very useful in predicting the time windows.

For the prediction of delivery time windows, the inputs of the ANN are the independent variables. The functional relationship estimated by the ANN is written in equation (2.1), where $x_1, x_2, \dots, x_n$ represent the n independent variables and y is the dependent variable to predict. In this context, the ANN is functionally equivalent to a nonlinear regression model. However, considering a time series forecasting problem, the input data are typically the past observations and the output is the predicted value. This can be written as the mapping function visualized in equation (2.2), where y_t represents the observation at time t.

$$y = f(x_1, x_2, \dots, x_n) \qquad (2.1)$$

$$y_{t+1} = f(y_1, y_2, \dots, y_{t-n}) \qquad (2.2)$$

Before the model is built, the total data available is usually divided into a training set, test set, and a validation set. The training set consists of the sample data for training the model. The test set refers to the hold-out sample that is used for testing the performance of the model. On the other hand, the validation set consists of data is not used for training or testing the model and therefore has not been show to the network before. The ratio used for dividing the dataset is 70:15:15.

The model that will be developed will use the Levenberg-Marquardt optimization algorithm as training function, which is the same function that Jeong & Rilett (2004) used in their research. Their objective what to predict bus arrival times using Automatic Vehicle Location (AVL) data. The measure of effectiveness used to qualify the accuracy is the difference, expressed as the MAE, between the actual and predicted arrival times. The same measurement will be used to evaluate the different training models that will be developed.

For successful application of Neural Networks, it is very important to specify the number of hidden layers and nodes. The hidden nodes allow the neural network for pattern recognition, detect the feature, and perform mapping between input and output variables. Literature suggests the use of different number of hidden layers. Barron (1994) states that two hidden layer networks may provide more benefits than one. In accordance, several authors address this problem and consider more than one hidden layer in their network. However, Lapedes & Farber (1988) state that a network never needs more than two hidden layers to solve most problems including forecasting. Therefore, in this research two hidden layers will be used.

Determining the optimal number of nodes to use in a neural network is a complicated issue. There is no theoretical basis or a systematic approach for selecting these parameters. The problem with too few hidden nodes is the fact that the network may not have enough power to model and learn the data. However, to many nodes could result in a overfitted model. Some researchers including Lippmann (1987) provided an empirical rule to restrict the number of nodes, these includes using "$2n + 1$", where n is the number of input nodes. The problem is that none of these rules works well for all problems. Therefore, in practice the most common way to determine the optimal number of hidden nodes is via trial-and-error.

2.7 Parameters

In this research, the combination of the prediction techniques and the method of address intelligence are implemented. The first step in developing the model is specifying the relevant input variables. According to the method of Van Duin et al. (2015), address intelligence could be used to find case specific variables. In application with the target variable, the zip code could have influence on the time window delivery. Therefore, in this research a combination of the techniques and the method proposed by van Duin et al. (2015) will be used for predicting the delivery time windows.

The other relevant independent variables for the prediction models will be chosen using input from experienced employees at the case study company. Literature can suggest many different potential parameters that influence the ETA. However, for the case study only a select number of parameters can be extracted from the TMS. Therefore, the input from employees is valuable in order to assess which relevant parameters to use when developing the models.

Chapter 3

Method for Predicting Time Windows

This chapter is part of the Design & Development phase of the DSRP and explains the proposed techniques used in this research. First a multi linear regression model is built. This technique combines different variables that can influence the prediction of the delivery time window. Second, a ANN model is built for the same purpose. Finally, the attributes used for developing the model are specified.

3.1 Modelling Approach

This section elaborates on the approach of the models. First the regression model is elaborated followed by the ANN. This section explains the key constructs of the two imposed techniques. First, a multi-linear regression model is built as this is done in the paper of van Duin et al. (2015) and Jeong & Rilett (2004). Second the ANN model with input from Jeong & Rilett (2004) and the state for the art analysis of Zhang et al. (1998) si developed. Finally, the evaluation of the techniques is discussed.

3.1.1 Initial Regression Model

Regression analysis involves identifying relationships between the independent variable and the dependent variables. It is the most widely used analysis of all statistical techniques. In this research, different parameters are used to develop an estimated regression equation for predicting delivery time windows. As an illustration of the regression technique, the equation for computing the predicted value of Y using linear regression is visualized below:

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \ldots + \beta_n \cdot X_n + \varepsilon \qquad (3.1)$$

Y denotes the dependent variable, also named as the target variable to predict, and $X_1, \ldots, X_n$ are the independent variables from which Y is being predicted. β_0 stands for the so-called intercept, and $\beta_1, \ldots, \beta_n$ are the coefficients of the variables. The parameters (intercept and the coefficients) are estimated by least squares, resulting in the model's prediction errors to be typically normally distributed. Finally, ε represents the unexplained variation in the predicted variable.

At the centre of the regression technique stands the relationship between the independent variable and the dependent variables. As stated before, in this research the independent variable represents the predicted delivery time window. The dependent variables are the parameters used for the prediction of the time window. In order to analyse the dependency between the predicted variable and the parameter used for prediction, the correlation coefficient is used. Correlation is related to the sense with relationships among variables. It is a measurement of linear association between two variables, where the correlation coefficients are always a value between -1 and +1. A high positive value indicates a good relation in a positive linear sense, where in contrast a high negative value indicates a negative linear sense. Consequently, a correlation coefficient of 0 indicates no linear relationship between two variables. This research uses the same cut-off (0.15) value as Jeong & Rilett (2004).

3.1.2 Artificial Neural Networks (ANN)

In this research, Neural Networks is used as a computational tool for analysing data and developing models that help to identify structures and patterns in the data. First the available data is divided in a training, testing and validating sample. Which is respectively distributed in 70% training data, 15% testing data, and 15% validating data. The next step is to develop the model.

The input layer of a ANN consists of different parameters. Choosing the right parameters to use is prevail for achieving a high performance. Therefore, in order to get the relevant variables to use in the model construction, feature selection is used. This process consists of selecting a subset of relevant variables or predictors to use for developing the model. In this research, the performance of the parameters will be tested by using an iterative process. First, the model is tested using one parameter for 100 iterations. Next, the best variable is selected and the model is tested using this variable as initial starting point in combination with the other variables. This process is tested over again until all variables are included in the model. The performance of the different combinations will be visualized in a graph and the subset of variables with the highest performance will be selected as input layer.

The next step is to decide how many hidden layers to use in the model. As stated before, the paper of Zhang et al. (1998) suggests that using two hidden layers may provide more benefits than one. Therefore, the model is developed using two hidden layers. In addition, every hidden layer should contain an optimal number of neurons. Previous research provided different backbones to use for finding the optimal number of neurons in each layer. However, since the model will be developed with two layers, different combinations of neurons are applicable. In order to find the optimal combination, different solutions are tested. To get a representative output, the solutions run 10 iterations. For this research, the table below is tested to get the optimal number of neurons in each hidden layer.

		Layer size hidden layer 2						
		3	5	7	10	15	20	25
Layer size hidden layer 1	3							
	5							
	7							
	10							
	15							
	20							
	25							

Table 5 - Different layer sizes used for testing

Finally, when the model is developed the performance is obtained by running the network on the validation data. This gives a representative output, since this sample is not used in training or testing the model.

3.2 Relevant prediction variables

An important consideration when modelling is the fact that the modeler should aim at keeping the model as simple as possible, while yielding maximum results. This implies that the model should only contain variables that improve the prediction performance. Therefore, only a few variables available in the dataset are used for modelling. In this subsection, the different variables are presented. First the parameters dealing with the method of van Duin et al. (2015) are presented. Next, the relevant parameters suggested by employees at the case study company are presented.

3.2.1 Address Intelligence

According to the method of van Duin et al. (2015), address intelligence is used to build their regression model. For the application in this research, the following two parameters are used for predicting the delivery time window using a regression model:

1. Receiver zip code
2. Consigner zip code

The receiver zip code denotes the address of delivery. In the Netherlands, this zip code consists of four integer values and two strings. The same counts for the consigner zip code, which is the zip code from where the shipment originally comes from. This could be an address in the Netherlands or in a foreign country. *Table 6* visualizes different postcode numbers in the Netherlands specified with zip code area.

Zip code[2]	*Zip code area*
1000-2100	Amsterdam and Noord-Holland
2200-2400	Zuid-Holland
2500-3100	Den Haag and municipality Rotterdam
3200-4200	Area of Utrecht
4200-5600	Noord-Brabant
5700-6600	Heerlen, Kerkrade
6700-8300	Gelderland, most of Drenthe and Overijssel
8300-9900	Friesland and Groningen

Table 6 - Zip code areas per municipality

3.2.2 Cargo Handled

The service time is associated with the specific type of cargo handled and the form type. The specific type of cargo is usually specified as a product code. In practice, almost every company uses these codes to specify what type of cargo is being transported. The specific type of cargo handled can have influence on the delivery time window, since different products have different characteristics. For example, ADR cargo has to be handled very carefully which could result in delivery delays.

The form type is related to the form in which the cargo is delivered. When transporting cargo, the following classes are mainly used to deliver the order(s):

- Number of Pallets
- Number of Collo's
- Number of Roll Containers (RC)

---

[2] http://postcodebijadres.nl/postcodes-nederland

3.2.3 ETA

The parameter ETA is associated with the time the partner estimates to deliver the order. This variable is crucial in predicting the delivery time window, since the window of delivery is built around the ETA.

The last variable that is analysed is the specific day of delivery of the week. Some orders are for example delivered at one day of the week on structural basis. In addition, a delivery could be planned two times a week but at a different time. In order to build a representative model with a high performance, it is important to test this parameter.

3.3 Evaluation Method

From a business perspective, only the final result is assessed in terms of orders that are delivered within the proposed time window. Using historical data, the difference between the actual delivery times and the predicted delivery times by the different models can be calculated. This prediction is evaluated with the help of a several performance measurements. Starting with the Mean Absolute Error (MAE), where $\hat{y}_j$ is the predicted value (ETA) and y_j is the actual value. It indicates the difference between the predicted value and the actual value for all n number of observations. The MAE takes the average absolute deviation, which is calculated in the following way:

$$Mean\ Absolute\ Error = \frac{1}{n}\sum_{j=1}^{n}\left|\hat{y}_j - y_j\right| \qquad (3.2)$$

In addition, the Mean Percentage Error (MPE) is also measured to assess the number of orders that are delivered within a certain time window. The y_j is the actual delivery time and the $\hat{y}_j$ represents the ETA. This performance measure looks at the percentage of prediction improvement in a proposed time window as a function of the number of orders n. This measure consists of three different metrics; 180, 120 and 60 minutes time window. To give a good overview of the accuracy of the models, all these time window intervals are measured. The MAPE is calculated as followed:

$$Mean\ Absolute\ Percentage\ Error = \frac{100\%}{n} \cdot \sum_{j=1}^{n}\left|\frac{y_j - \hat{y}_j}{y_j}\right| \qquad (3.3)$$

Finally, since there is a difference in impact on customer satisfaction between too early and too late delivery, a distinction between these two is made. The orders that are not delivered within the certain time window can be further analysed by classifying them in earliness and lateness. The earliness assesses the percentage of orders that are delivered too early, y_j^E represents a binary variable that is 1 if the actual delivery time is smaller than the ETA minus the time window duration w divided by two, and zero otherwise. In this research, the orders that are classified as too early are evaluated according to the following equation:

$$Earliness = \frac{100\%}{n} \cdot \sum_{j=1}^{n} y_j^E \qquad (3.4)$$

$$y_j^E = \begin{cases} 1, & if\ \ y_j < \hat{y}_j - \frac{w}{2} \\ 0, & otherwise \end{cases}$$

The lateness assesses the percentage of orders that are delivered too late. In the equation below y_j^L represents a binary variable that equals 1 if the actual delivery time is greater than the ETA plus he time window duration w divided by two, and zero otherwise. The orders that are classified as too late are evaluated according to the following equation:

$$Lateness = \frac{100\%}{n} \cdot \sum_{j=1}^{n} y_j^L \qquad (3.5)$$

$$y_j^L = \begin{cases} 1, & if \;\; y_j > \hat{y}_j + \frac{w}{2} \\ 0, & otherwise \end{cases}$$

Chapter 4

Case study

In the previous chapter the chosen techniques were presented. In this section, based on the demonstration phase of the DSRP, the evaluation of the method on real generated data is presented using a case study at a freight distribution company. First, the elements of the CRISP-DM framework are used as a guide for performing the case study. Next, the initial performance at the case study company is presented. Finally, two models are developed and tested.

4.1 Business Understanding

Understanding the business and industry is necessary before the data can be used. In this section, the problem is analysed from a business perspective, and underlying business mechanisms. First a rough understanding of the business case and goals are explained. Second, the target variable is defined. Finally, a process description is made to get a better understanding of the overall process that impacts the expected time of delivery.

4.1.1 Business Relevance

Companies aim at the highest possible service level for the delivery of the outgoing orders. From a business point of view the reliability of the delivery in the provided time window should be 100%. A business could align his production processes with the expected time of delivery of the goods. Being too late has negative effects on the overall process of the company and results in financial consequences. Which means that a company should rely on the provided time for delivery. However, in practice this is infeasible because there is always a random event involved. For example, a road that is blocked, bad weather, or high waiting times at the previous customer. Therefore, the case study company implements a specified time window around the ETA in order to give a reliable estimate for delivery. In the initial situation, a naïve three-hour time window is communicated with the customer. The desired service level to achieve is to deliver 95% of the orders within the specified time window. In the following section the current state of performance will be presented.

4.1.2 Target Variable (ETA)

Currently, the ETA is used as an indicator for delivery. Based on the provided ETA a naïve three-hour time window is communicated with the customer. However, currently there is no KPI that gives insight in the actual performance for the three-hour window. Therefore, in this research different time windows are considered. A two hour and a four-hour time window are also implemented to give practical insights for the case study company.

The ETA counts as the target variable since the time window will be built around the predicted ETA. Currently, the ETA is set by the partner and on the day of delivery it is communicated with the TMS of the case study company. The customer can then track his delivery using a Track&Trace

application. This platform presents a three-hour naïve time window around the ETA, for example; when the ETA is set on 10:30, a delivery time window between 09:00 and 12:00 is communicated with the customer. However, it must be stated that the application of using ETA's and time windows is implemented a year ago. Previous the customer only knew that the delivery would take place on a specific day, without knowing the ETA or time window. Due to customer preferences, the case study company choose to implement using ETA's and time windows on a regularly basis. But measuring the performance in the form of KPI's and communicating this with the partner has never been done before. Therefore, the data contains orders without ETA. It is impossible to measure the initial performance of the partner delivery if the ETA is not present in the data. In addition, building a prediction model without using the ETA as a parameter would result in very poor predictions.

To give an idea on how bad the performance will be if the model will be built without the ETA parameter *Table 7* is created. It can be seen that building a regression model without using the ETA parameter the performance will decrease with 47% for the three-hour time window. In addition, the MAE and MAPE will increase with more than 100%. This implies that the ETA should be used in order to develop a representative model.

KPI	Without ETA parameter	With ETA parameter	Difference
In Time Window	46.6%	87.7%	***-47%***
MAE	104.4 minutes	46.4 minutes	***+125%***
MAPE	7.25%	3.22%	***+125%***

Table 7 - Performance model without ETA

4.1.3 Process and Partner Description

In this subsection, the overall delivery process of cargo at the freight distribution company is described. In order to do that, the process characteristics are discussed and explained.

The Cargo division is part of the Logistic Solutions department. This department consists of many networks with a separate specialism for delivering goods. The focus at the Cargo division is on delivery of orders with more than six packages, and on value added services like; ADR, Pharma & Care and Special services. The customers come from very different industries varying from retailers to hospitals.

The delivery of the orders is done using the partner network of the company. In the Netherlands they use five partners to deliver their orders. All these partners use their own applications for communicating with the IT environment of the case study company. In addition, the planning of the routes and predicting the ETA's are the responsibilities of the partner.

The trivial process model at the case study company is presented in *Figure 6*. After the distribution at the cross-dock warehouse, the goods are transferred to a partner. From there on, the partner loads his truck with the goods that have to be distributed according to a given route that day. When all these activities are completed, the delivery process can be started. Each truck has to deliver a set of orders on a route. The route consists of orders that are delivered on structural basis, but also on one-time orders that are requested by a specific customer.

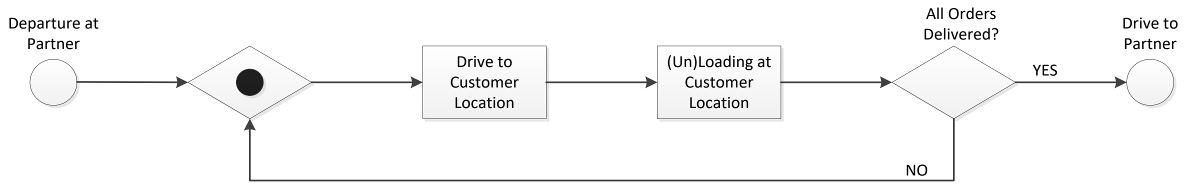


Figure 6 - Trivial process model case study company

4.2 Data Understanding

This subsection aims at creating an understanding of the data used in this research for analysis. Next to the characteristics of the data also the context is addressed to give a clear overview of the dataset.

4.2.1 Data Collection

The dataset obtained from the case study company contains all the deliveries of orders conducted in the period of 5 September 2016 to 28 November 2016, which is equal to 12 weeks of data. The data is extracted from the TMS of the case study company. The exact number of orders delivered in this period is confidential.

4.2.2 Data Description

The data consists of many columns with information about the order including the ETA. In order to assess if the delivery is done in the provided time window, both the ETA and the actual time of delivery should be specified in the order. The actual time of delivery of the order is specified as the i01event.

In addition, this section looks at the different relevant predictor variables and how they are presented in the current situation. By using only the relevant parameters, the final performance of the final model will be higher than by including all relevant data. In theory, machine learning techniques select the important variables on their own. However, this is time consuming and not even necessary. The human selection of relevant predictors in combination with suggestions from literature, leads to a simpler initial starting model. This implication improves the performance of the machine learning algorithm used to derive the final model.

Relevant predictors

The first category of relevant predictors is the application of address intelligence. In this research, this refers to the use of the following two predictor variables:

1. Receiver zip code
2. Consigner zip code

In the modelling part the receiver zip code can be seen as the ‘to zip code’, and the consigner zip code to the ‘from zip code’. The receiver zip code always consists of four integers and two string values, since the scope of this research are only orders delivered in the Netherlands. The consigner zip code refers to the client sender that the case study company serves. Every day the orders that have to be delivered the next day are consolidated at a specific partner warehouse, depending on the area of delivery. The ‘from zip code’ is the zip code from where the order originally comes from, before it was received at the warehouse of the partner. This could also be a zip code from a foreign country.

The second category of predictor in the dataset is the type of product. The type of product is specified as a four-integer value. The freight distribution company uses these product codes to show what kind of product it transports. In addition, the product is transported in a specific form; RC, pallet, and collo. These specific transport forms are specified as an integer value ≥ 0, and can be seen as relevant predictor variables. The service time associated with one or more pallets is different than the time for one or more collo's. The same applies for the relation between the other transporting forms. Therefore, in building the final prediction models, the specific product code and transporting form are considered as relevant parameters.

The dataset contains two other relevant predictor variables which are the target variable (ETA), and the day of the week the delivery takes place in. The target variable is earlier discussed in *Chapter 4.1.2*. The day of delivery can be found by translating the delivery date to the specific day of the week. Some of the orders are delivered on structural basis, and therefore planned each specific day of the week at the exact same time. Therefore, this parameter should be extracted from the data and considered as one of the relevant predictor.

4.3 Data preparation

This section of the Crisp-DM framework covers all steps to construct the final dataset from the initial raw data. Since all the relevant data is extracted from one TMS, the output is a unified dataset that can be fed in excel or MATLAB. The regression technique is done in Excel and therefore most of the data preparation too. The modelling of the ANN is done in the programming language of MATLAB. Therefore, some of the specific data preparation activities required for the ANN must be done in MATLAB.

4.3.1 Data Cleaning

Before the data can be used for analysis, data cleaning activities should be done. First, irrelevant information has to be excluded from the raw data set. This means that some filters need to be applied in order to get the dataset ready for analysis. The filters in *Table 8* are applied to the raw dataset. In addition, the table describes the percentage of orders that is excluded when applying the specific filter to the raw dataset. It can be stated that most of the data loss is due to filter 2 and 5.

Filter #	Description	Percentage Decline
1	*Orders that are not delivered in the Netherlands*	**-12.5%**
2	*Orders without a I01 event (physical delivery scan)*	**-60.8%**
3	*Only consider LB type of orders*	**-1.7%**
4	*Orders with irregularities are deleted*	**-9.7%**
5	*Orders that have no ETA*	**-64.7%**
6	*Only considered orders delivered with MAE between 0 and 300 minutes*	**-15%**

Table 8 - Data cleaning statistics

Filter (1) is applied because orders that are delivered in another country are not part of the partner network. The partners only deliver orders in their delivery area, which is in the Netherlands. The second filter ensures that only the orders that have a physical delivery scan are considered. Without this scan, it is impossible to measure what the performance is, because the actual time of delivery is unknown. Filter (3) filters on the LB type of orders, which are the distribution orders. Pick-up orders and collection orders are not in the scope of this research. Not all the orders are delivered smoothly, sometimes the receiver is closed or the order is rejected. Filter (4) ensures that all the orders with irregularities are deleted from the dataset. *Chapter 4.1.2* discusses the importance of the parameter ETA, therefore orders with no ETA are excluded from the dataset by filter (5). The last data filter ensures that orders that are delivered more than five hours too late, are excluded. These orders can be seen as outliers, which would have negative effects when training the prediction model.

After applying these filters, the dataset can be analysed. First, the performance of the three-hour time window for every partner is analysed. Since there is a difference in the effect of too early and too late delivery, a distinction between these two is made. Therefore, the analysis contains the performance considering a three-hour time window with too early and too late deliveries. In addition, the performance of a two-hour time window is analysed, since this would be an improvement compared to the current situation.

4.3.2 Data Quality

This section intends to assess the quality of the data. Elements as missing values, inconsistent values, and typos among the dataset are analysed. The data report extracted from the TMS consist of many columns with specific information about the order.

One of the columns has an issue, which is the i01 event. This event specifies the exact time of delivery of the order and sometimes it is measured a few hours later than the given ETA. This implies that the order is delivered too late, however after analysing most of the data it can be concluded that this is an irregularity that occurs on structural basis. This data cannot be used to learn a model, because the actual value is unknown. Therefore, an assumption has to be made in order to omit these unreliable values from the final dataset.

Both the ETA and the i01 event should be available in order to assess the delivery performance. However, after analysing the data and applying the other filters, it comes clear that around 20% of the orders has no ETA specified. These orders should be deleted from the final dataset. On overall, the data is of high quality and needs to be filtered to the specific case study.

4.3.3 Format Data

After assessing the quality of the data and cleaning missing values the data will also be prepared using some calculations to extract more details about each order. There is a difference between Partner D and Partner E. In the initial dataset extracted from the TMS, Partner D represents these two units. Therefore, the first step in preparing the data is to divide these two partner networks. This is done by filtering on a specific product code and insert a new partner selection column. In this way, it is easy to oversee the performance per partner network.

Consequently, the zip codes in the dataset contain integer and string values. In order to analyse the data and build a prediction model, the data should contain only integer values. Therefore, the string values in the zip codes are deleted. This is done for both the ‘to zip code’ and ‘from zip code’ parameter.

For this research, the delivery time window is determined by adding and extracting 1.5 hours to the ETA, which represents respectively the upper- and lower bound of the time window. In order to assess the orders delivered in the time window, the actual time of delivery should be ≥ the lower bound, and ≤ the upper bound. Furthermore, assessing if the order is too early delivered the i01is set < than the lower bound and for assessing if the order is too late the i01 is set > than the upper bound. Therefore, a column with ‘In time window’, ‘too early’, and ‘too late’ is added to the dataset for analysis.

4.4 Current Performance

Before the proposed technique is evaluated in a case study, the current performance at the case study company is measured. The analysis contains the results of the partners using the available data. To give representative insights in the current performance, three months of data is used. In accordance with the manager of operations, a specific time frame for data usage is chosen. The exact date period used is from September 5th, until November 28th, 2016.

The following partners are considered for analysis; Partner A, B, C, D, and E. The performance is measured according to the different evaluation methods named in *Chapter 3.5*. In addition, a naïve time window of respectively four, three and two hours is used to compare the differences between the partners. The naïve time window uses the average as the centre of the time window and an equal width to each side. This three-hour time window is used as a starting point for analysis, since this is the current way of communicating time windows with the customer.

Next, the time window performance could be optimized by maximizing the number of orders delivered in the three-hour window. This could be done by setting the window instead of naïve, unequal to each side. This technique is being referred as the adjusted time window. The case study company wants insight in this performance since this could be implemented easily in practice. For the adjusted time window only a three-hour window is considered.

4.5.1 Partner A

According to different conversations with employees of the case study company, they assumed that this specific partner would have the best performance in time window delivery. The argumentation for this assumption is the fact that they have the most advanced IT system compared to the other partners.

Figure 7 represents the performance using a naïve time window of three hours and *Figure 8* using an adjusted time window of minus 1.75 hours and plus 1.25 hours. It can be stated that the performance is high and that it makes no difference using the adjusted time window. In both cases the performance stays almost the same (87.3% for naïve window & 88.3% adjusted window).

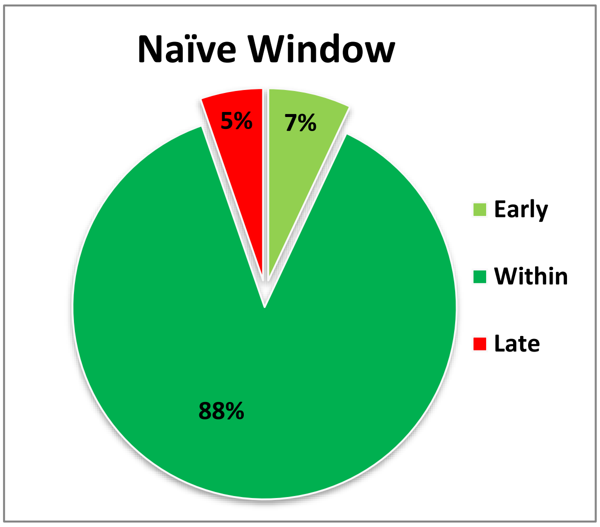


Figure 7 - Performance of the naïve window current situation Partner A

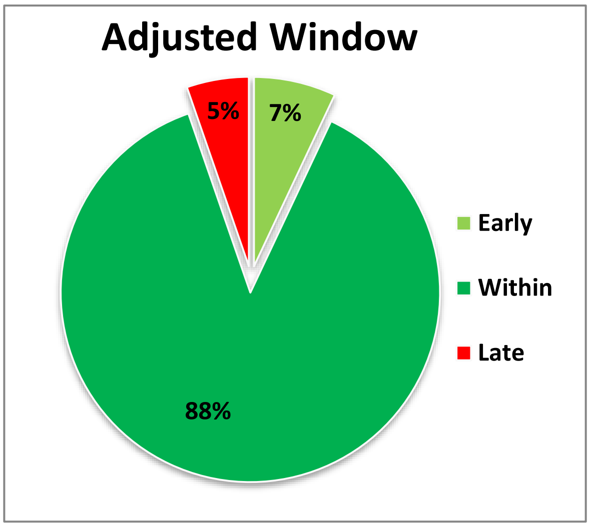


Figure 8 - Performance of the adjusted window current situation Partner A

Using the naïve window, the absolute deviation per time unit can be calculated. *Figure 9* shows the distribution between the difference of arrival time and the actual arrival. The average error of the partner is -6.33 minutes and the standard deviation is 64.95 minutes. This implies that most of the orders are delivered within the time window or too early. From a business point of view, too late delivery is a more severe problem than too early delivery. When summing the orders delivered within the time window and too early, the required service level of 95% is achieved.

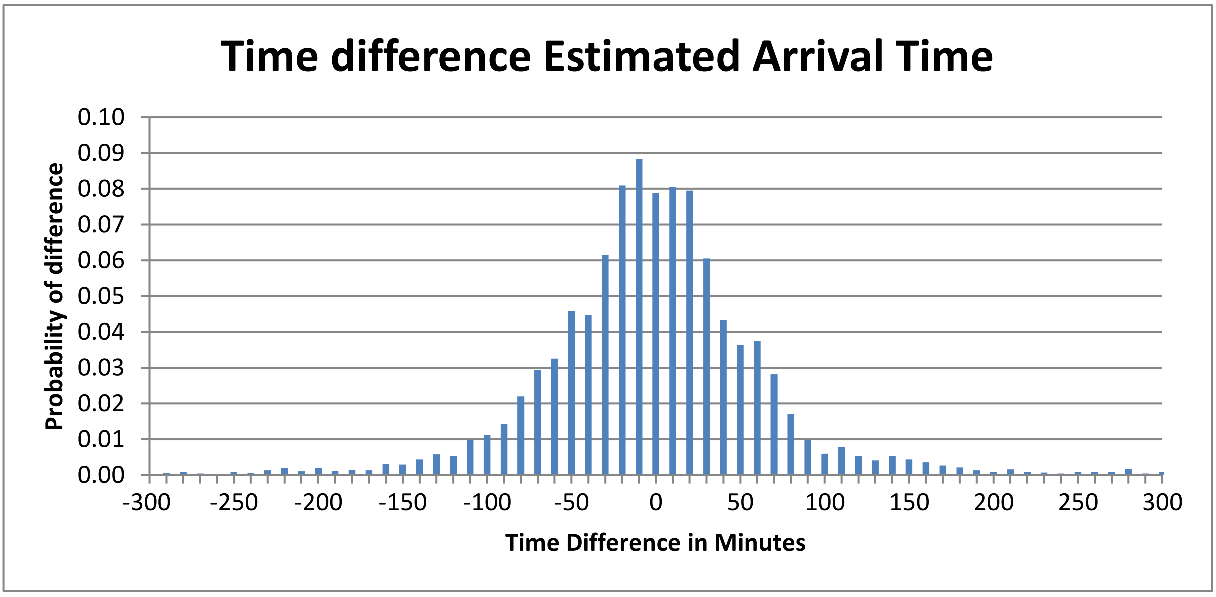


Figure 9 - Difference between estimated arrival time and actual arrival time distribution of Partner A

The performance for the different naïve windows are visualized in *Table 9*. It can be concluded that implementing a higher time window (four hour) has less impact on the performance than implementing a lower time window (two hour). A logical reason for that could be the fact that the MAE is 46.4 minutes, resulting in a higher percentage of orders falling out of the window when implementing a naïve two-hour time window. The last performance measurement is the MAPE, this is used to evaluate the performance compared to different partners and the use of different prediction techniques. On overall, it can be stated that the performance of this partner is good.

Performance Measurement	
Four Hour Window	93.1%
Three Hour Window	87.7%
Two Hour Window	73.8%
MAE	46.4 minutes
MAPE	3.22%

Table 9 -Performance measurements current situation Partner A

4.5.2 Partner B

The second partner for analysis is Partner B. This partner is responsible for the deliveries around the area of West-Brabant. *Figure 10* represents the performance using a naïve time window of three hours. It can be stated that the performance using the naïve time window is low. Almost half of the total orders are delivered outside the time window. This suggests that when using an adjusted time window this could have a great impact on the overall performance. *Figure 11* represents the performance using the adjusted window. In this case, the window is set using the ETA as the lower bound and adding three hours as the upper bound. This results in an increase of more than 25% in performance for the three-hour time window.

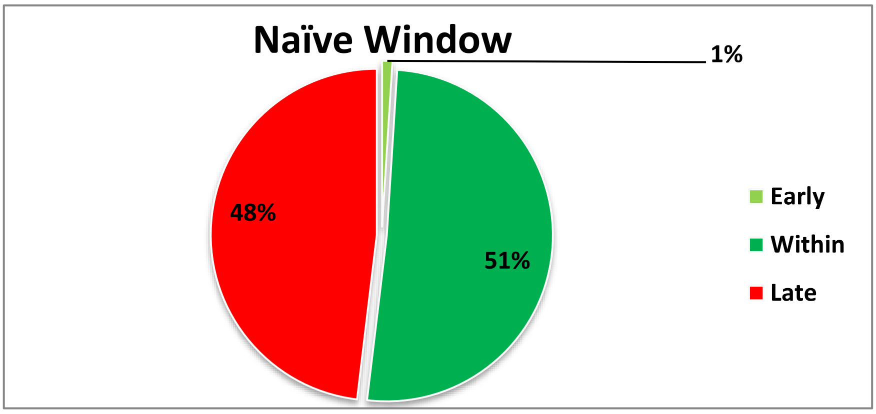


Figure 10 - Performance of the naïve window Partner B

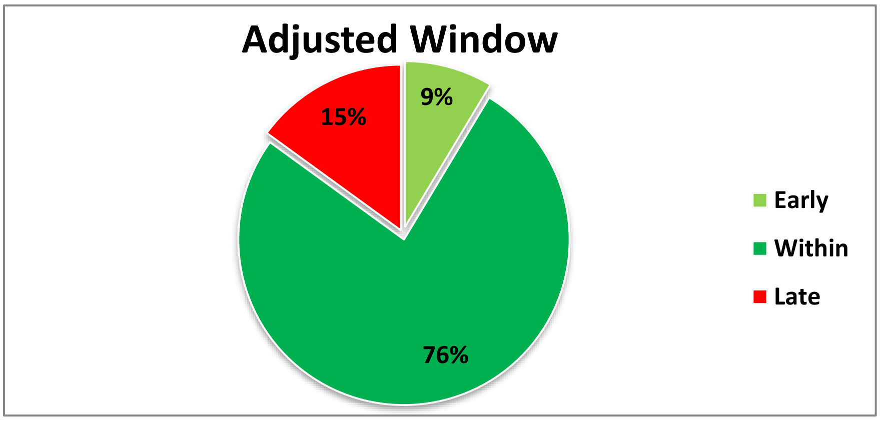


Figure 11 - Performance of the adjusted time window Partner B

Using the naïve window the absolute deviation per time unit can be calculated. *Figure 12* shows the distribution between the difference of arrival time and the actual arrival. The average error of the partner is 94.9 minutes and the standard deviation is 80.44 minutes. This implies that most of the orders are delivered outside the time window. According to the distribution most of the errors occur in the right side of the figure, resulting in a low performance due to too late delivery.

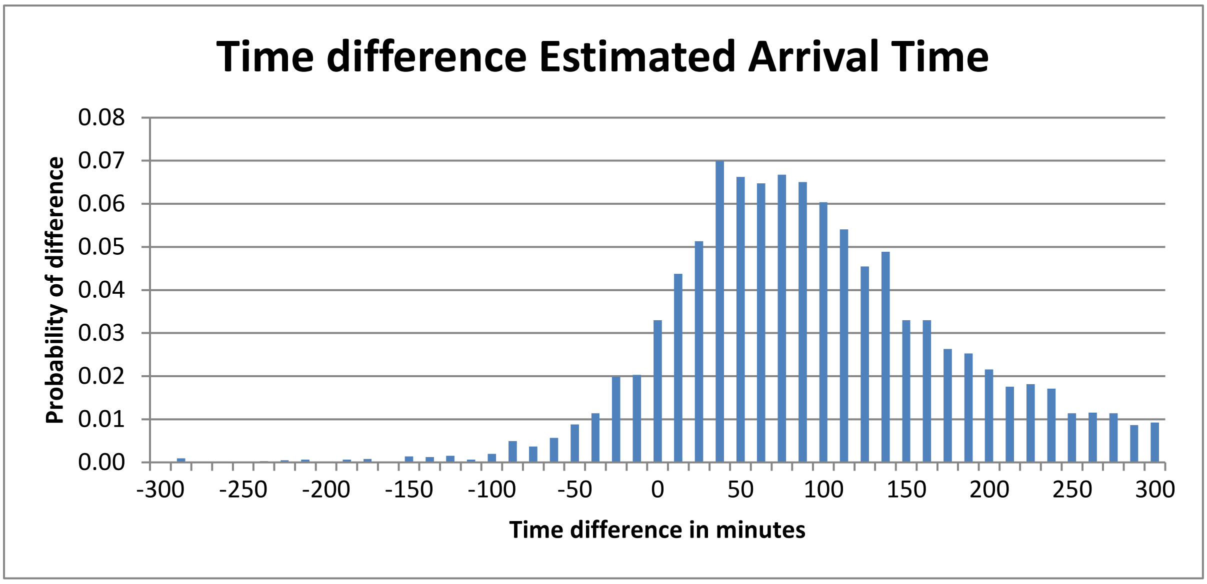


Figure 12 - Difference between estimated arrival time and actual arrival time distribution of Partner B

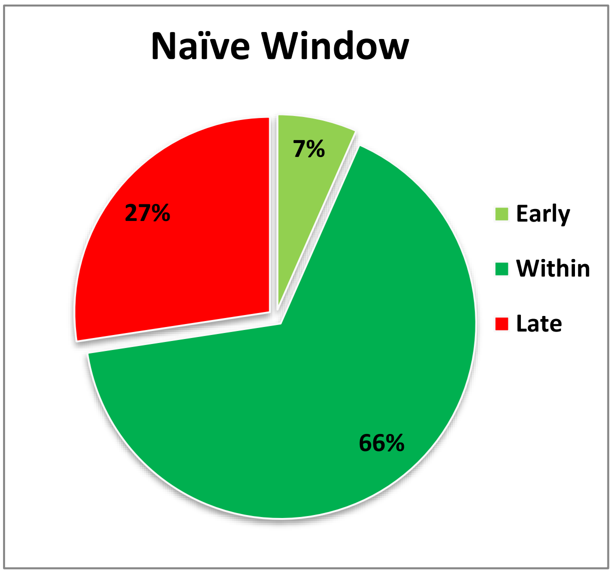
The performance for the different naïve windows are visualized in *Table 10*. It can be concluded that implementing a higher time window (four hour) has less impact on the performance compared to the performance of the adjusted time window. This is due to the fact that adjusting the time window from the lower bound has almost no effect on the performance, since the average error of the partner is higher than 90 minutes. In addition, it can be seen that most of the orders are delivered with an error on the right side of the distribution (91.4%). Therefore, the lower bound of the time window should be close to the given ETA in order to achieve a higher delivery performance.

Performance Measurement	
Four Hour Window	65.7%
Three Hour Window	50.9%
Two Hour Window	33.3%
MAE	102.6 minutes
MAPE	7.13%

Table 10 - Performance measurements current situation Partner B

4.5.3 Partner C

The third partner for analysis is Partner C. This partner is responsible for the deliveries around Noord-Holland. One specific thing about this partner is the fact that more than 30% of the data is filtered out, because these orders were delivered more than 300 minutes too late. This could be the result of negligent signing out orders when actually delivered. Hence, the rest of the sample data is used for analysis. *Figure 13* represents the performance using a naïve time window of three hours, and *Figure 14* using an adjusted time window. In this case, the lower bound is set by subtracting one hour from the ETA, and the upper bound is set by adding two hours to the ETA. It can be concluded that using an adjusted time window has low impact on the average performance.



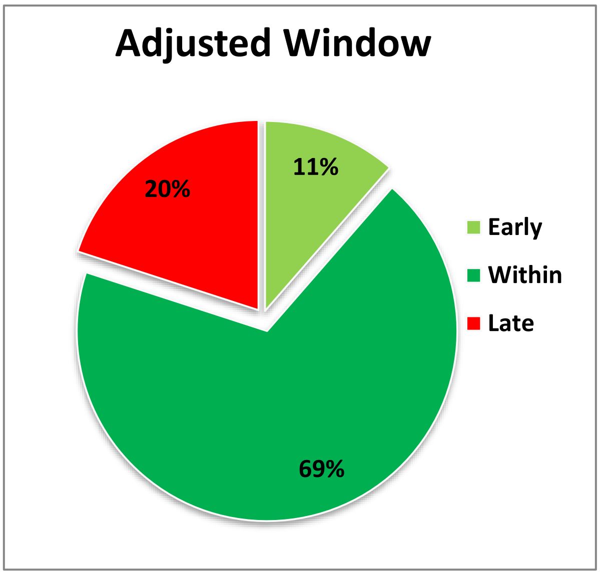


Figure 13 - Performance using naïve time window for Partner C

Figure 14 - Performance using adjusted time window for Partner C

For the naïve window, the absolute deviation per time unit is calculated. *Figure 15* shows the distribution between the difference of arrival time and the actual arrival. The average error of the partner is 47 minutes and the standard deviation is 99.8 minutes. This implies that most of the orders are delivered within the time window. According to *Figure 13* this can be confirmed.

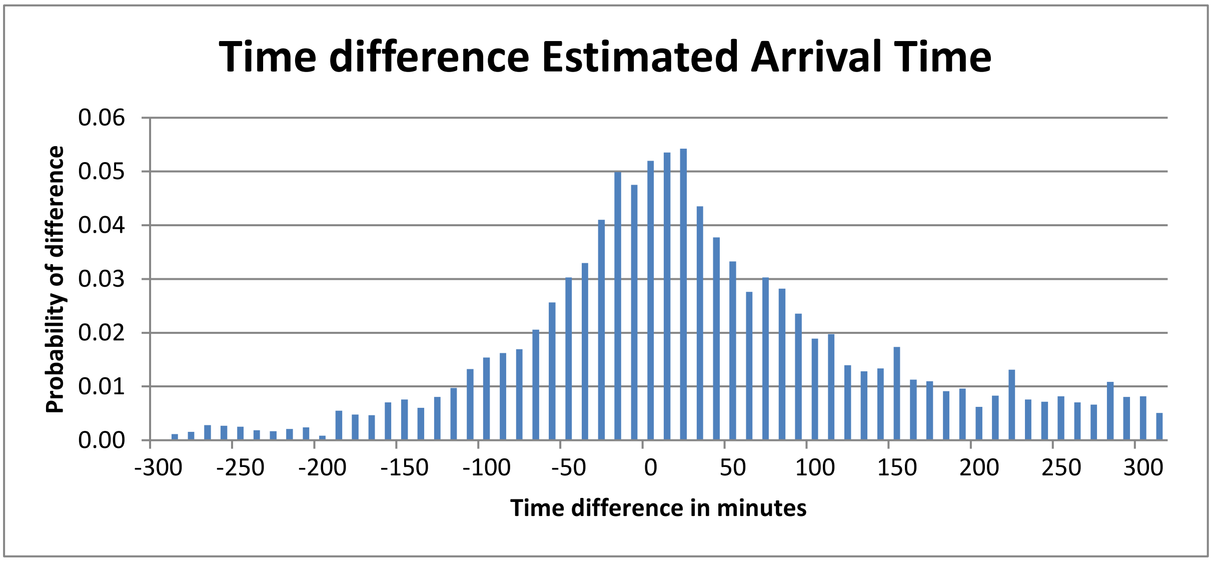


Figure 15 - Difference between estimated arrival time and actual arrival time distribution of Partner C

The performance metrics for the different naïve windows are visualized in *Table 11*. It can be concluded that implementing a time window of four hours has a higher impact on the performance compared to the adjusted time window. This is since the right and left side of the probability distribution (*Figure 15*) are almost evenly spread. Therefore, adjusting the time window only in favour of one side of the bound, has little effect. Instead, by expanding the time window delivery from three to four hours, the performance increases with almost 10%. The same counts for reducing the time window from three- to two hours. It can be seen in *Table 11* that the performance in that case will decline by 15%.

Performance Measurement	
Four Hour Window	75.7%
Three Hour Window	66%
Two Hour Window	51%
MAE	82.1 minutes
MAPE	5.70%

Table 11 - Performance measurements current situation Partner C

4.5.4 Partner D

Partner D is responsible for the deliveries around Zuid-Holland and Zeeland. *Figure 16* represents the performance using a naïve time window of three hours, and *Figure 17* using an adjusted time window. In this case, the lower bound is set by subtracting 0.25 hours from the ETA, and the upper bound is set by adding 2.75 hours to the ETA. It can be concluded that using an adjusted time window has significant impact on the delivery performance.

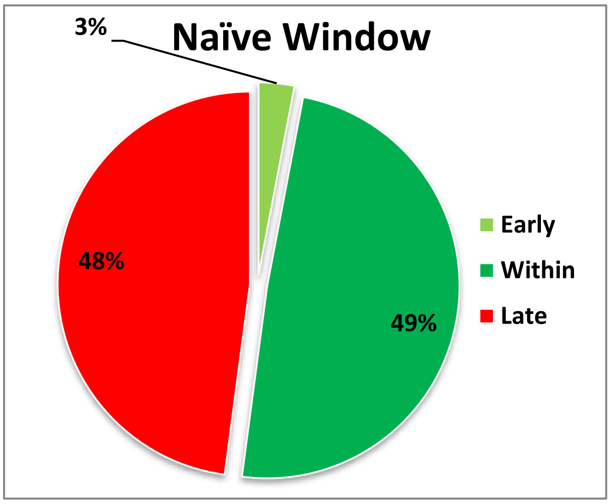


Figure 16 - Performance 3 hour naïve time window Partner D

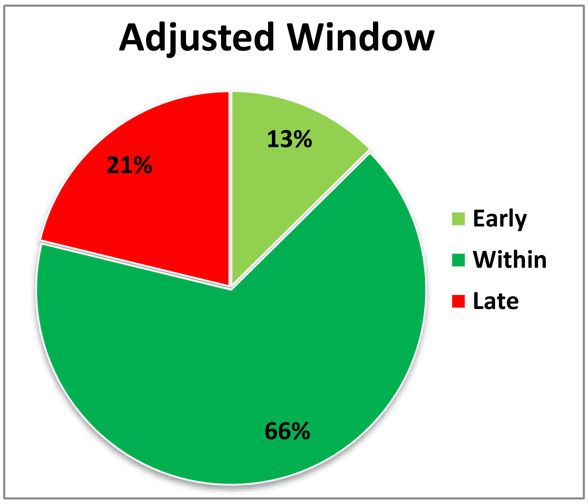


Figure 17 - Performance adjusted time window Partner D

Figure 18 shows the distribution between the difference of arrival time and the actual arrival. The average error of the partner is 90 minutes and the standard deviation is 99.1 minutes. This implies that most of the orders have an error on the right side of the distribution, which can be confirmed by *Figure 16,* where almost half of the total orders are delivered too late.

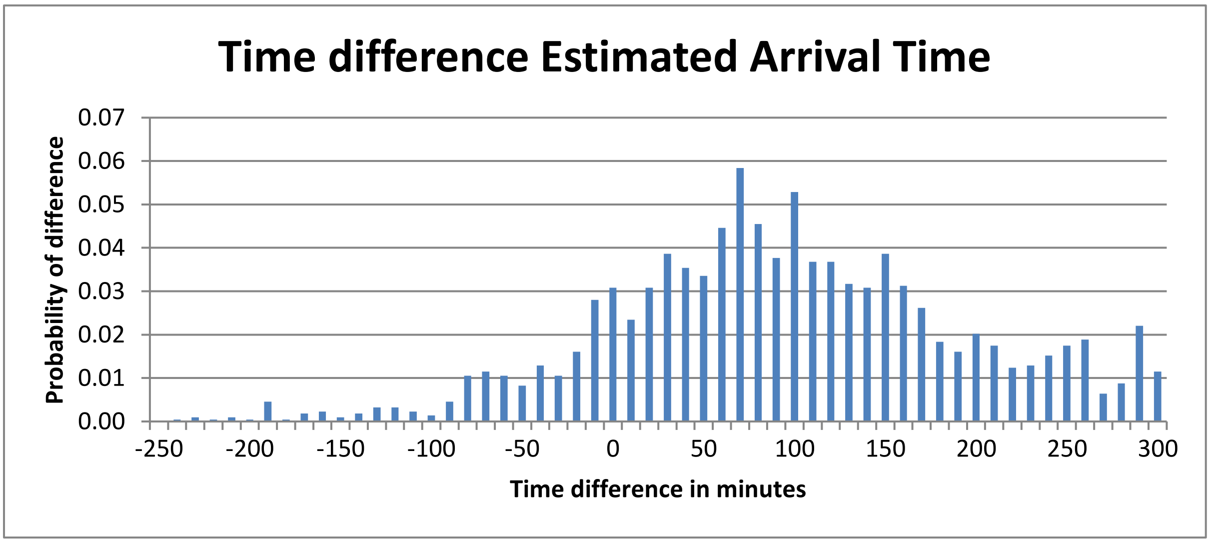


Figure 18 - Difference between estimated arrival time and actual arrival time distribution of Partner D

The performance for the different naïve windows are visualized in *Table 12*. It can be concluded that implementing a time window of four hours has a lower impact on the performance compared to implementing the adjusted time window. This is because the right side of the probability distribution (*Figure 18*) contains more orders than the left side, in practice this means that more orders are delivered too late than too early. This can be confirmed by *Figure 16*, where 48% of the orders is delivered too late. Therefore, adjusting the time window in favour of both sides of the distribution has little effect. Instead, by expanding the upper bound and narrowing the lower bound an adjusted time window of three hours could achieve a higher performance than implementing a four-hour time window. In addition, the MAE is visualized in *Table 12*, which is equal to 108.4 minutes. This implies that implementing a naïve time window will results in a low delivery performance.

Performance Measurement	
Four Hour Window	62.4%
Three Hour Window	49.1%
Two Hour Window	31.4%
MAE	108.4 minutes
MAPE	7.53%

Table 12 - Performance measurements current situation of Partner D

4.5.5 Partner E

Partner E is responsible for the deliveries of healthcare products and pharma transport. This network operates for deliveries in the Benelux, however for the analysis we only consider orders delivered in the Netherlands. *Figure 19* represents the performance using a naïve time window of three hours, and *Figure 20* using an adjusted time window. In this case, the lower bound is set by subtracting 0.5 hours from the ETA, and the upper bound is set by adding 2.5 hours to the ETA. It can be concluded that using an adjusted time window has significant impact on the delivery performance. Most of the orders that were initially too late are now delivered within the three-hour time window.

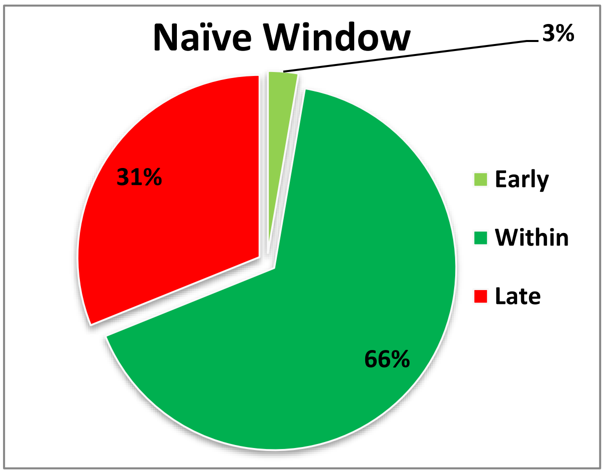


Figure 19 - Performance 3 hour naïve time window for Partner E

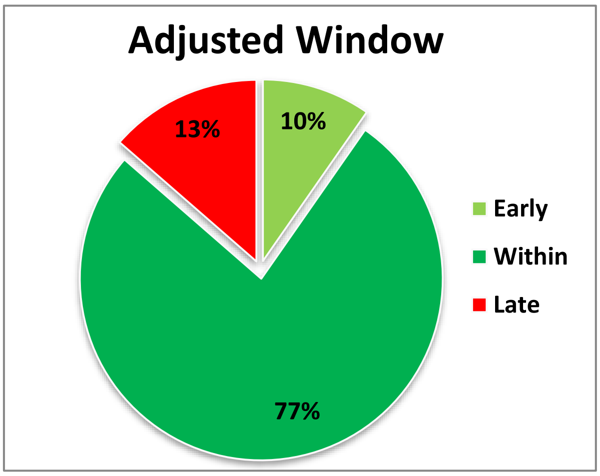


Figure 20 - Performance adjusted time window for Partner E

Figure 21 shows the distribution between the difference of arrival time and the actual arrival. The average error of the partner is 61.2 minutes and the standard deviation is 81.1 minutes. This implies that most of the orders have an error on the right side of the distribution. According to the distribution, 85% of the orders have a positive error. Therefore, by adjusting the time window to the upper bound a higher delivery performance can be achieved.

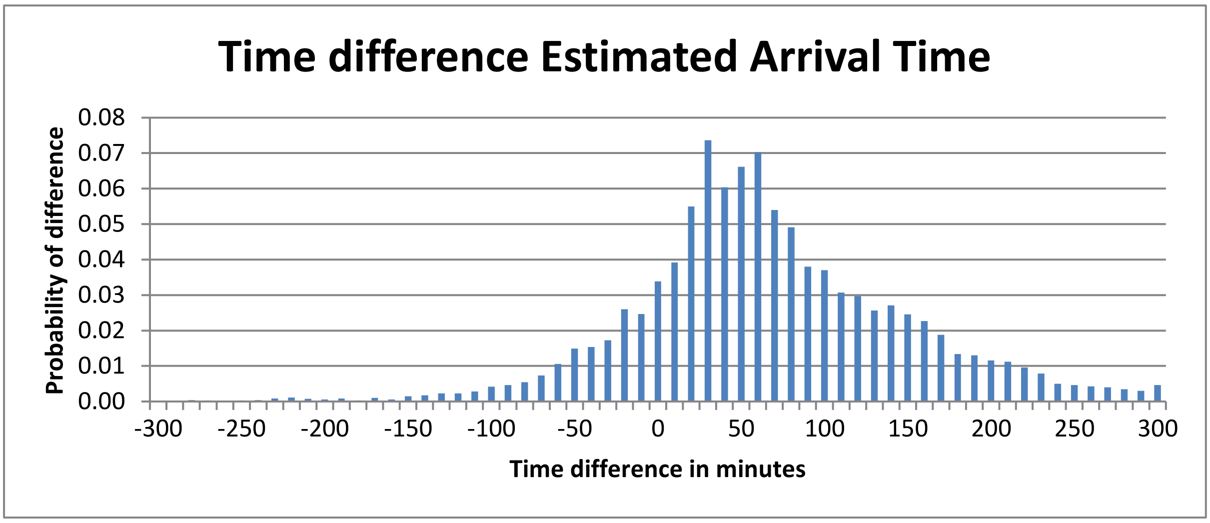


Figure 21 - Difference between estimated arrival time and actual arrival time distribution Partner E

The performance for the different naïve windows are visualized in *Table 13*. It can be concluded that implementing a time window of four hours has almost the same impact on the performance compared to implementing the adjusted time window. This could be clarified by the fact that the current performance is 66%, which implies that most of the orders are already delivered within the time window. By adjusting the time window to the right side of the distribution the orders that are initially too late will be within the time window. However, since the average error is 61 minutes, the adjusted time window will be at his optimum when adding 2.5 hours instead of 1.5 hours to the upper bound. From that point on, the performance will decline. In addition, the MAE is visualized in *Table 13*, which is equal to 79.1 minutes. This is close to the 90 minutes that is added to the lower and upper bound of the naïve time window. Therefore, by expanding the naïve time window to four hours, the upper bound is equal to 120 minutes, resulting in an extra 10% of orders included in the time window performance.

Performance Measurement	
Four Hour Window	77.1%
Three Hour Window	66.2%
Two Hour Window	49.7%
MAE	79.1 minutes
MAPE	5.50%

Table 13 - Performance measurements current situation of Partner E

4.6 Modelling

In order to compare the different techniques with each other and asses the performance, results are obtained based upon the dataset of the case study company. Therefore, the models are built with data from the period specified before. In addition, the techniques are validated using 15% of data compared to the data for building the model. Two techniques for building the prediction model have been selected; Regression and Neural Network. Both techniques have different parameters that have been selected for modelling.

4.6.1 Regression Model

Partner A

The regression technique is based on the parameters used as input. For Partner A six of the eight variables are used for predicting the delivery time window. These parameters are chosen based on the result of the regression model. In addition, the error of the predictions are checked using a normality distribution (Appendix A). These variables and their coefficients can also be found in *Appendix A* under Partner A. The performance of the model for the different time windows can be found in *Table 14*. The validation of the technique results in a slightly higher performance, which is quite unusual. However, this can be the fact since the data for validating the regression model is based on a different time period.

Performance Measurement	Regression	Validation
Four Hour Window	93.2%	95.9%
Three Hour Window	88.3%	91.2%
Two Hour Window	73.8%	78.4%
MAE	45.8 minutes	42.0 minutes
MAPE	3.18%	2.91%

Table 14 - Performance using the regression technique for Partner A

The six variables that are selected for developing the model are tested on their correlation coefficient. As can be seen in Appendix A, the coefficient values show that there is almost no relationship between the parameters. Only the variables 'Product Code' and 'RC' show a weak positive relationship. As can be seen in *Figure 22*, this partner scores a high performance considering the delivery within the time window. If even the orders that are delivered too early and within the time window are summed up, the final performance of this regression model will be 97%. Compared to the initial situation, the performance of the delivery in the naïve time window increases with 4%. The low increase is due to the fact that in the current situation the overall performance was already high for this partner. However, it can be concluded that the developed regression model improves the performance.

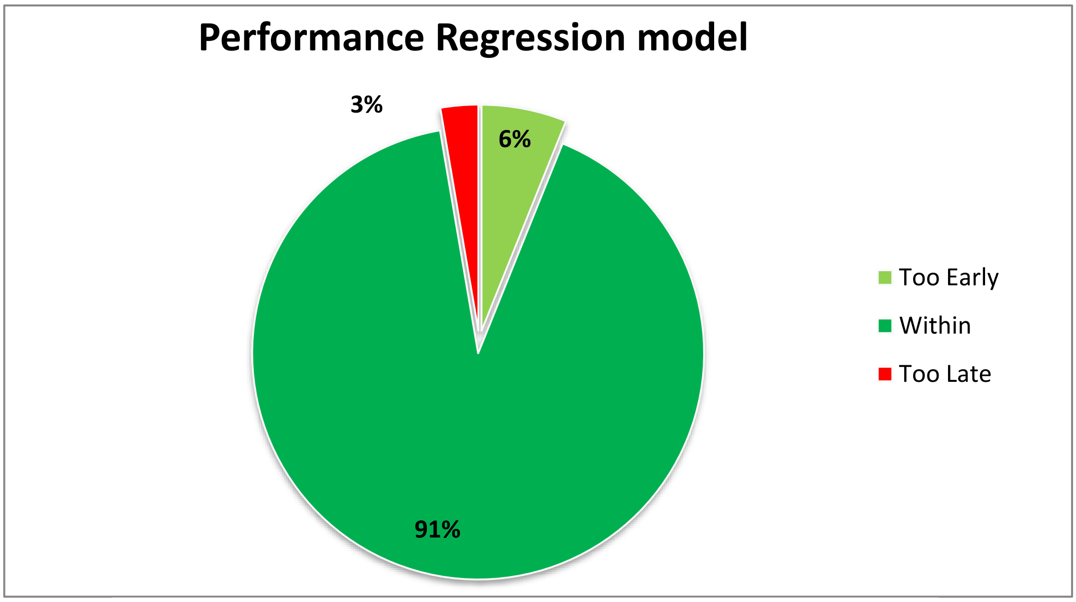


Figure 22 - Performance three-hour time window for regression model Partner A

Partner B

For this partner, only the variables “product code” and “ETA” are used for building the final regression model. This is due to the initial regression output, which can be found in Appendix A. *Table 15* represents the performance of the regression model for the different time windows. It can be concluded that this is a great improvement compared to the initial situation where the performance for delivery within the three-hour time window was only 51%. In addition, the MAE drops with 40 minutes and the performance of the other time windows also increases significantly.

Performance Measurement	Regression	Validation
Four Hour Window	86.9%	88.4%
Three Hour Window	76.1%	76.3%
Two Hour Window	57.3%	57.0%
MAE	62.6 minutes	61.8 minutes
MAPE	4.34%	4.27%

Table 15 - Performance regression model for Partner B

The two variables selected for developing the final regression model are tested on their correlation coefficient, *Appendix A* shows the results. There is no correlation found between the two variables “product code” and “ETA”. The performance of the regression model for the three-hour window is visualized in *Figure 23*. Here it can be seen that the performance improved significantly and the percentage of orders that are too late and too early are evenly spread. The standard deviation is 77.7 minutes, which is almost the same as in the initial situation.

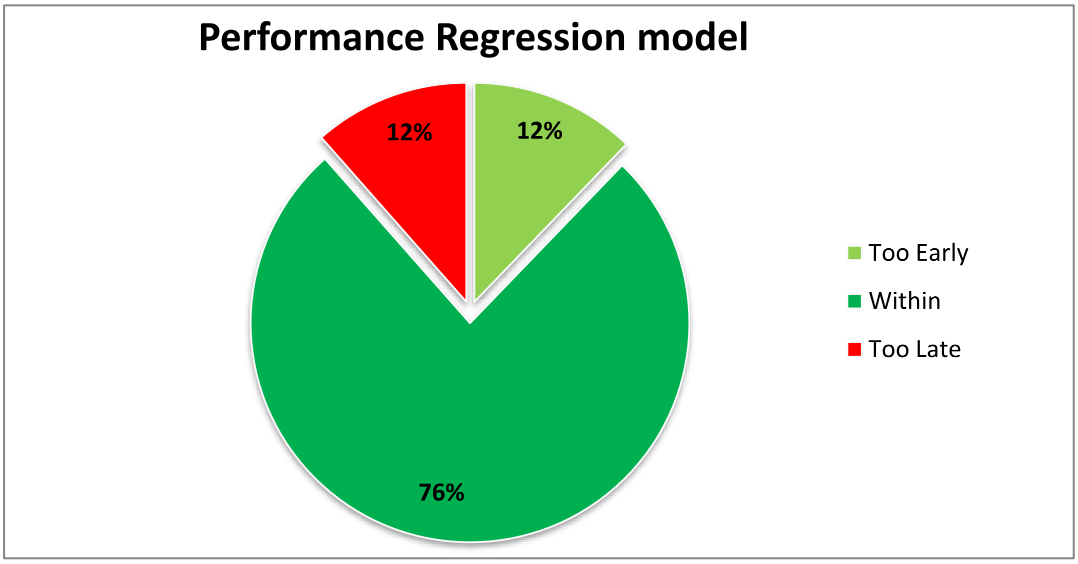


Figure 23 - Performance three-hour time window for regression model Partner B

Partner C

For this partner, the following three variables are used to develop the model; the from zip code, the number of collo delivered, and the ETA. The parameter selection is based on the initial regression output. The performance of this model can be found in *Table 16*. It can be seen that the performance has slightly improved. Compared to the initial situation the performance improved for about 5%, which is less than expected. In addition, the MAE has decreased with about 10 minutes. The small improvement is due to the great variety in time delivery.

Performance Measurement	Regression	Validation
Four Hour Window	79.5%	82.4%
Three Hour Window	70.1%	69.5%
Two Hour Window	52.9%	53.9%
MAE	74.8 minutes	71.4 minutes
MAPE	5.19%	4.96%

Table 16 - Performance of regression model for Partner C

All the variables used for developing the regression model are tested on correlation, the results are specified in Appendix A. According to the correlation coefficient, it can be stated that there is no correlation between the four variables. The overall performance of the regression model is visualized in *Figure 24*. The number of orders that are delivered too early is higher than the number of orders that are delivered too late. This is since most of the orders are delivered on the left side of the probability distribution (Appendix A). The standard deviation has slightly decreased to 98.2 minutes.

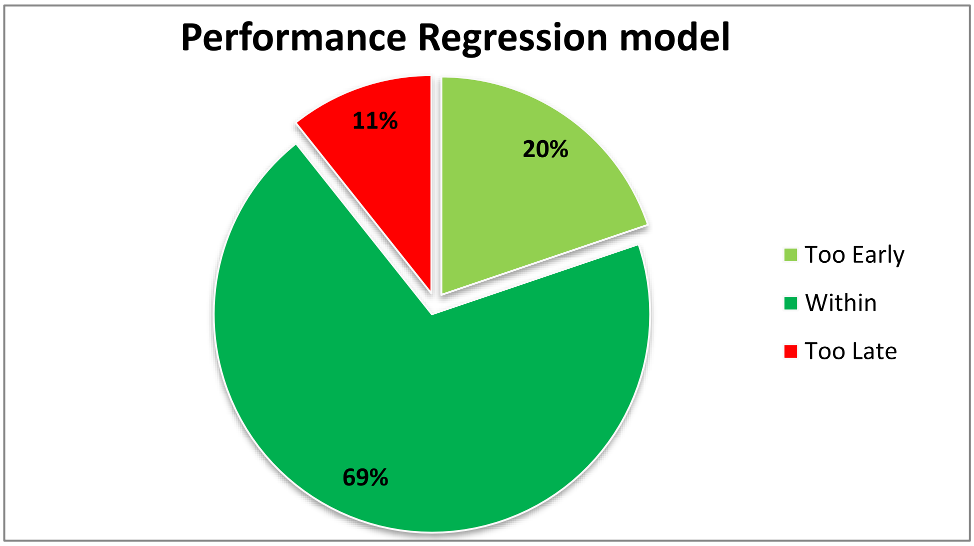


Figure 24 - Performance three-hour time window for regression model Partner C

Partner D

This partner only uses three variables, based on the initial regression output, for developing the final regression model. Both the variables that are associated with address intelligence are used, and the ETA. The performance of the regression model for the different time windows can be seen in *Table 17*. Compared to the initial situation, the performance using the regression model increases significantly. In addition, the MAE reduces with more than half an hour. Therefore, it can be stated that using this regression model for this partner, will improve their time window performance.

Performance Measurement	Regression	Validation
Four Hour Window	79.3%	81.3%
Three Hour Window	65.8%	66.4%
Two Hour Window	48.3%	45.8%
MAE	75.7 minutes	73.1minutes
MAPE	5.26%	5.07%

Table 17 - Performance of the regression model for Partner D

The variables used for building the regression model are tested on correlation, the results are specified in *Appendix A*. Analysing the correlation coefficients resulted in the conclusion that none of the variables used in the model are correlated. The performance of the three-hour time window delivery using the regression model is visualized in *Figure 25*. Most of the orders are delivered within the time window, and the orders delivered outside this window are almost evenly spread. This is a huge difference compared to the initial situation where almost half of the orders was delivered too late. In addition, the standard deviation has slightly decreased to 96.2 minutes.

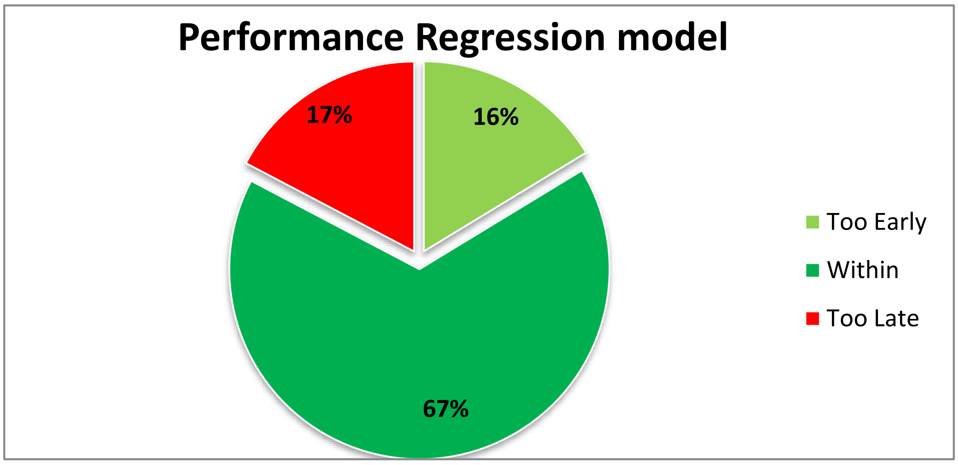


Figure 25 - Performance three-hour time window for regression model Partner D

Partner E

The regression model for this partner is based on the following input parameters: the "to zip code", the "day of the week", the "ETA", and the number of "collo delivered". The performance of the different delivery time windows are presented in *Table 18*. The performance of the three-hour time window comes close to 80%, which is almost 20% higher compared to the initial situation. In addition, the MAE drops with almost 30%. Overall, it can be stated that the regression model for this partner achieves significant improvements, which can be confirmed by the results of the normality distribution visualized in *Appendix A* under Partner E.

Performance Measurement	Regression	Validation
Four Hour Window	88.2%	87.9%
Three Hour Window	79.6%	79.2%
Two Hour Window	64.8%	65.9%
MAE	57.1 minutes	55.6 minutes
MAPE	3.97%	3.86%

Table 18 - Performance for the regression model for Partner E

All the parameters used for developing the regression model are tested on correlation, where *Appendix A* specifies the results of the correlation coefficients. According to the results, it can be concluded that none of the variables is correlated. *Figure 26* presents the performance of the three-hour time window delivery using the regression model. As can be seen, most of the orders are delivered within the time window. The number of orders delivered too late is higher than the number of orders that are delivered too early. This could be expected since more than 30% of the orders were too late in the initial situation. In addition, the standard deviation also slightly decreases to 77.1 minutes.

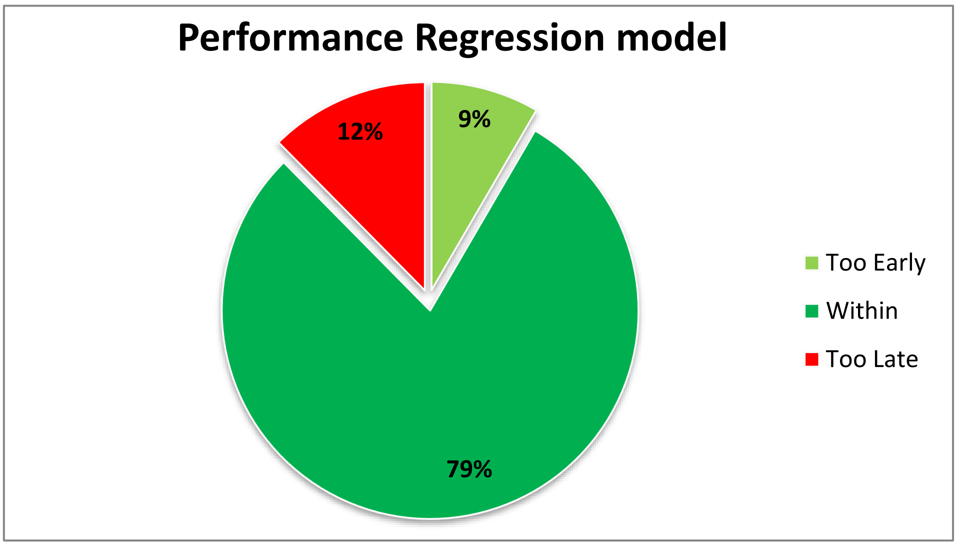


Figure 26 -Performance three-hour time window for regression model Partner E

4.6.2 Neural network

The Neural network model has three input parameters; the number of input nodes, the number of hidden layers, and the number of nodes in each layer. The number of input nodes is determined using the training module of Levensberg-Marquardt using feature selection with the MAE as performance measurement. In order to get a representative output of the performance, the model run 100 times using the training data. After the model has calculated the average performance for each input node, the 'best' variable is chosen based on 'Feature selection'. Thereafter, the model runs again with the specific variable(s) and the model is optimized by testing different number of neurons in each hidden layer. In order to get a representative output the model run 10 times.

Partner A

Figure 27 represents the performance of the network using different number of input nodes. It can be concluded that the optimal number of input nodes is two (variable 2 & 3). This results in a model with the parameters 'ETA' and 'Day of the week'.

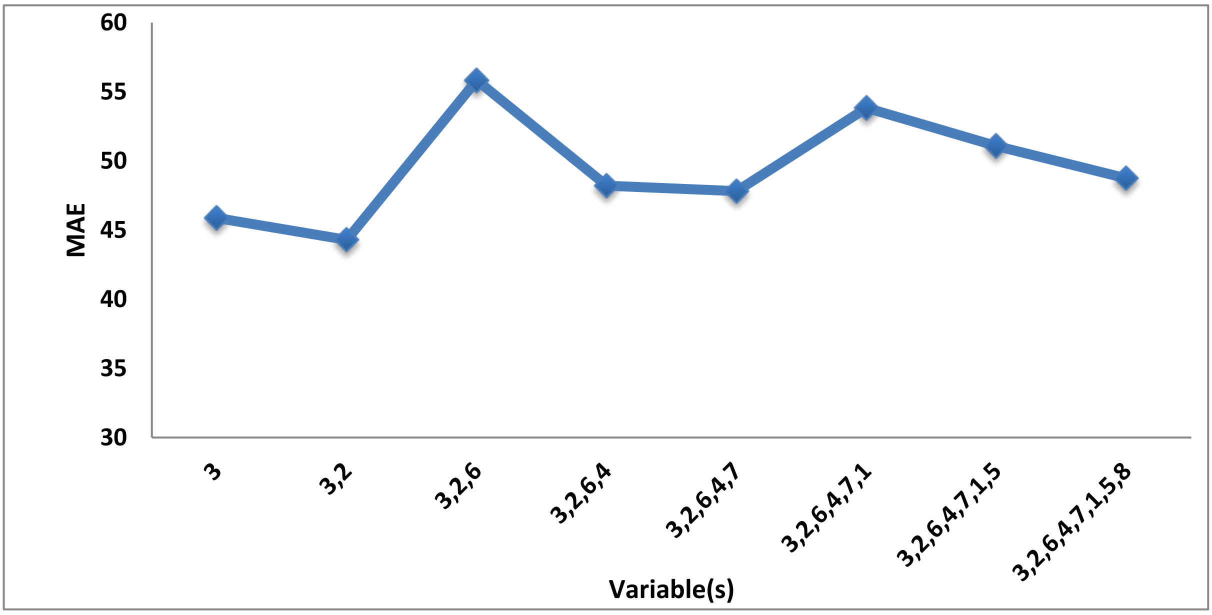


Figure 27 - Performance of the ANN model for Partner A using different variables

In order to optimize the performance of the network, the optimal value of neurons in the hidden layers should be found using parameter tuning. In the previous chapter the choice for the number of hidden layers is explained. The optimal number of nodes will be determined by testing the performance of the model with different combinations of hidden nodes using the testing data. As can be seen in *Table 19* the optimal number of neurons in hidden layer one and two is [15;10].

		Layer size hidden layer 2						
		3	5	7	10	15	20	25
Layer size hidden layer 1	3	44.56	47.52	45.79	46.67	45.35	44.57	44.73
	5	48.62	45.22	46.75	46.23	45.19	45.68	44.42
	7	46.91	46.86	44.60	45.11	45.67	45.78	43.10
	10	47.15	43.90	43.96	43.73	42.98	45.68	44.44
	15	44.49	44.13	45.55	**41.06**	44.01	47.77	41.72
	20	43.42	43.51	44.29	42.65	44.03	41.70	43.85
	25	47.84	42.40	44.01	43.72	44.17	43.35	44.60

Table 19 - Performance of the model with different number of nodes in the hidden layer's for Partner A

The performance, using the validation set, for the different time windows are visualized in *Table 20*. The performance of this partner is very good. When using a four-hour time window almost every order is delivered within this time window. The MAE and the MAPE confirm that this model achieves a good performance.

Performance Measurement	Neural Network
Four Hour Window	95.0%
Three Hour Window	89.7%
Two Hour Window	78.8%
MAE	41.3
MAPE	2.87%

Table 20 - Performance Neural Network model for Partner A

Partner B

Figure 28 represents the performance of the model using feature selection. As can be seen, the optimal number of input nodes is two (parameters 2 & 3). This results in the same model as Partner A, which includes the parameters 'ETA' and 'Day of the Week'. The performance of the model decreases when adding more variables.

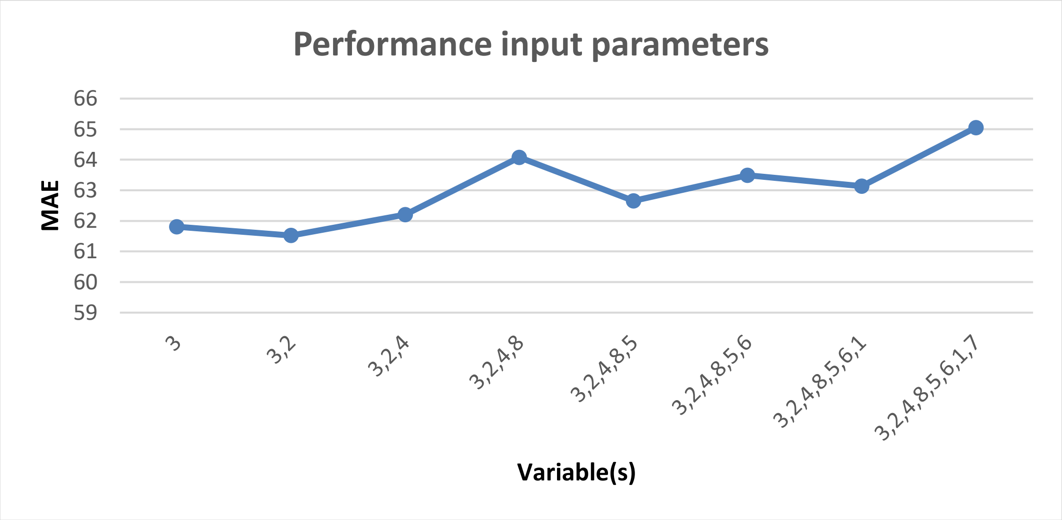


Figure 28 - Performance different input parameters of ANN model Partner B

The optimal number of neurons in the two hidden layers is found using parameter tuning. Since the number of hidden layers is two, different combinations with hidden nodes will be tested using the testing data. *Table 21* presents the results after the model run 10 times for each combination. The optimal number of neurons in hidden layer one is 25, and for hidden layer two 20.

	Layer size hidden layer 2							
		3	5	7	10	15	20	25
Layer size hidden layer 1	3	61.68	64.55	61.51	60.37	63.43	61.52	60.99
	5	63.89	60.86	61.13	63.71	62.35	62.90	62.89
	7	62.73	61.38	62.50	61.90	63.18	62.75	63.42
	10	63.46	60.55	60.94	61.22	63.98	61.26	62.61
	15	62.46	62.01	61.64	62.03	61.43	59.97	60.93
	20	65.05	536.62	60.67	64.43	60.27	63.16	61.47
	25	61.03	62.14	65.21	60.32	63.44	**59.80**	61.42

Table 21 - Performance of the different number of neurons in each hidden layer for Partner B

Table 22 visualizes the performance of the model for this partner. As can be seen, the performance is quite okay. Almost 80% of the orders are delivered within the required three-hour time window. Extending the time window with one hour results in a 10% performance increase, while decreasing the window with one hour results in a performance decrease of more than 15%. The MAE is almost equal to one hour and the MAPE is slightly more than 4%, which are also reasonable values.

Performance Measurement	Neural Network
Four Hour Window	88.3%
Three Hour Window	78.9%
Two Hour Window	62.1%
MAE	58.2
MAPE	4.04%

Table 22 - Performance measurements for the ANN model for Partner B

Partner C

The performance of the ANN model with different number of input nodes is presented in *Figure 29*. Based on feature selection, the variables 'ETA' and 'To Zip code' achieved the highest performance. However, it can be seen that the performance when including more variables does not differ much from the performance of the optimal number of input nodes. The MAE was 74.1 minutes.

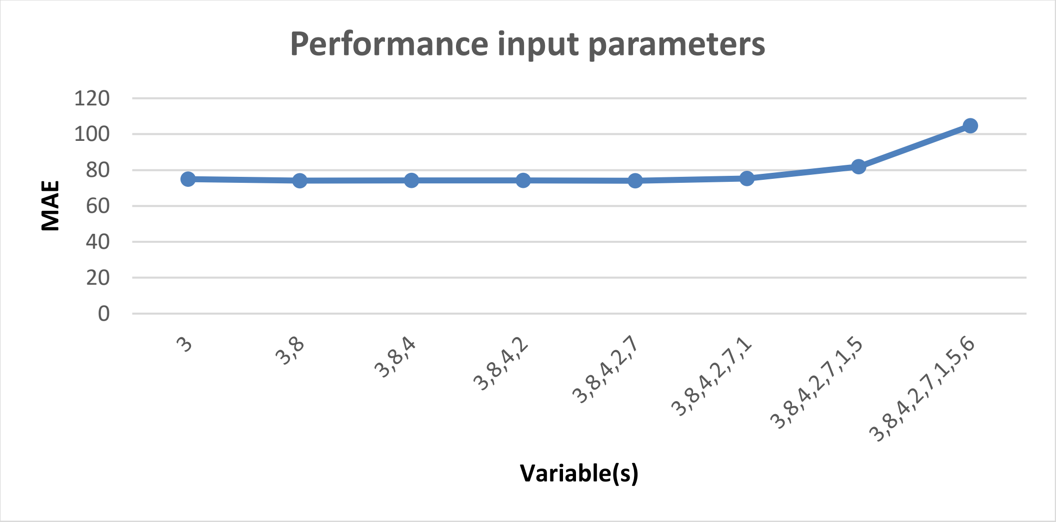


Figure 29 - Performance different input parameters of ANN model Partner C

The performance of the model can be optimized by finding the optimal number of neurons in each hidden layer. This will be done by testing different combinations of hidden nodes in layer one and two. *Table 23* presents the performance for the tested number of nodes. It can be seen that the optimal number of hidden neurons in layer one and two are respectively 25 and 20. This results in a MAE of 71.37 minutes which is a good improvement compared to the initial performance.

		Layer size hidden layer 2						
		3	*5*	*7*	*10*	*15*	*20*	*25*
Layer size hidden layer 1	*3*	72.10	77.08	76.99	79.60	75.91	75.52	73.95
	5	76.62	74.05	77.87	77.33	74.58	77.79	72.70
	7	75.88	73.32	74.28	73.95	78.44	76.24	77.91
	10	74.27	73.95	75.10	73.26	76.24	74.76	75.01
	15	75.02	75.34	71.34	74.16	73.72	73.52	**71.33**
	20	74.53	74.23	74.31	72.15	75.81	72.25	72.46
	25	75.57	75.99	71.53	72.38	76.64	**71.37**	75.48

Table 23 - Performance of the different number of neurons in each hidden layer for Partner C

In *Table 24* the performance for the different time windows can be found. For the three-hour window the performance is reasonable. However, when the time window is reduced to two hours the performance decreases with almost 20%. Increasing the performance results only in a performance increase of 8.2%. This can be explained by the fact that the MAE is almost 70 minutes. Most of the orders will already fall in the three-hour window, but will not if the time window is narrowed to one hour.

Performance Measurement	Neural Network
Four Hour Window	81.9%
Three Hour Window	73.7%
Two Hour Window	54.3%
MAE	69.5
MAPE	4.83%

Table 24 - Performance measurements for the ANN model for Partner C

4.6.2.4 Neural Network Partner D

Figure 16 represents the performance of the model using different input parameters. Including the variables 'Day of the Week', 'ETA', and 'To Zip code' results in a model with a MAE of 70.8 minutes, which is the highest performance achieved using feature selection. Including more variables in the model results in a lower model performance.

The performance of the model is improved by testing different combination of neurons in the two hidden layers. As can be seen in *Table 25*, the optimal number of neurons are respectively 25 for hidden layer one, and 7 for hidden layer 2. This results in a MAE of 61.88, which is 12.5% better compared to the initial model created. Using the two hidden layers and the hidden neurons in these layers, the final prediction model is created.

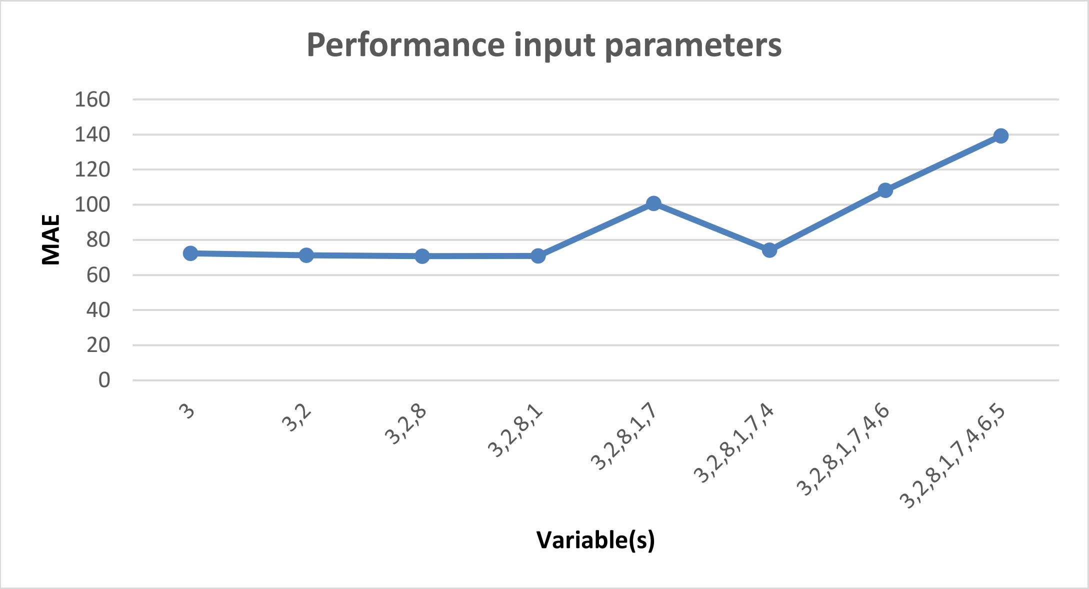


Figure 30 - Performance different input parameters of ANN model Partner D

	Layer size hidden layer 2						
Layer size hidden layer 1	*3*	*5*	*7*	*10*	*15*	*20*	*25*
3	71.89	69.09	72.29	71.08	80.11	70.89	79.17
5	71.80	74.95	71.25	69.92	72.66	77.87	74.09
7	70.82	71.31	76.18	71.92	74.47	70.65	75.87
10	73.29	74.07	70.43	75.75	71.07	69.62	73.38
15	69.66	74.37	76.50	68.96	72.17	67.79	69.66
20	69.98	72.90	69.81	62.83	66.64	73.64	69.39
25	65.97	70.06	**61.88**	69.69	68.78	65.41	62.67

Table 25 - Performance of the different number of neurons in each hidden layer for Partner D

After the final model is built, the performance is tested using the validation set. The performance for this model is visualized in *Table 26*. The number of orders delivered within the three-hour time window is above 70%. This is the lowest performance achieved compared to the other partners. In addition, a notable mark is the fact that the MAE was 61.88 minutes after using parameter tuning. However, after validating the model the MAE was 69.6 minutes. This can be due to the fact that this partner has a high variability between the ETA and the actual time of delivery. This affects the prediction model and therefore the performance.

Performance Measurement	Neural Network
Four Hour Window	83.7%
Three Hour Window	70.9%
Two Hour Window	54.6%
MAE	69.6
MAPE	4.83%

Table 26 - Performance measurements for the ANN model for Partner D

Partner E

The performance of the model achieves the optimal value with four input parameters. Only for this partner there are seven input parameters tested instead of eight, since the variable ‘RolContainer’ is not used in the dataset. The following parameters are included in the model; ‘Day of the Week’, ‘ETA’, ‘To Zip code’, and the ‘Product Code’. This results in a MAE of 56.5 minutes, which can be seen as a reasonable value. The performance including different variables is visualized in *Figure 31* below.

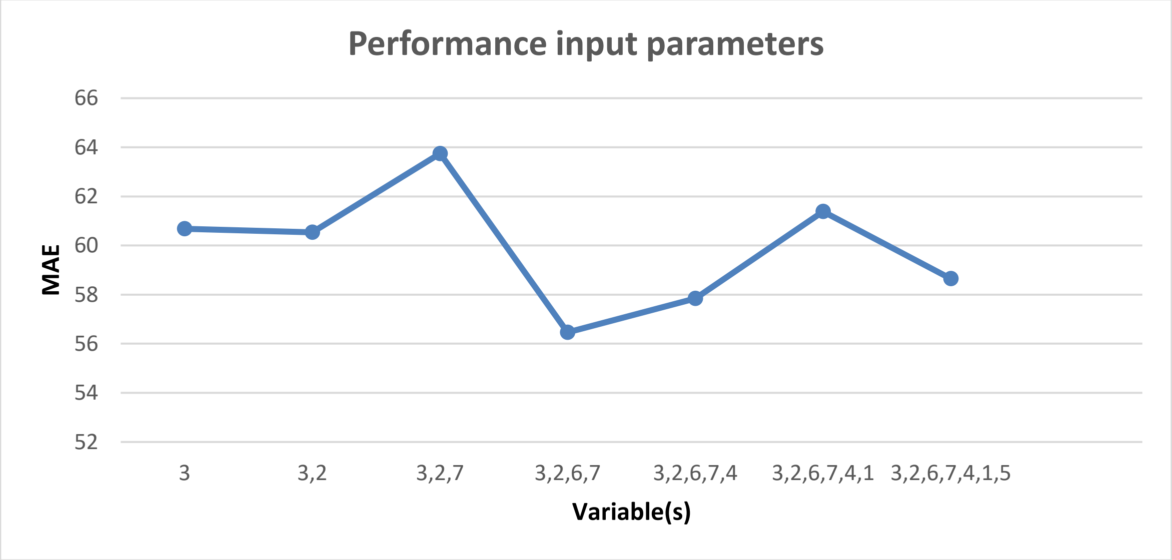


Figure 31 - Performance different input parameters of ANN model Partner E

After the input parameters are set, the optimal number of neurons in the two hidden layers are tested. This is done in the same way as by the other partners. It can be seen in *Table 27* that the optimal value for layer size one is 5 and for layer size two 25. The next step is to run the final model and compare the predicted values with the actual delivery time using the validation set.

Layer size hidden layer 2

Layer size hidden layer 1	*3*	*5*	*7*	*10*	*15*	*20*	*25*
3	58.63	56.93	57.53	56.79	55.80	56.47	58.92
5	57.74	56.25	54.43	55.73	57.94	58.23	**53.99**
7	59.34	55.71	56.92	56.78	55.36	57.40	55.07
10	55.42	55.34	57.41	55.10	59.30	57.92	56.34
15	54.13	56.38	54.95	56.81	54.20	59.77	56.30
20	54.65	54.93	56.30	56.68	55.42	56.72	58.75
25	55.24	56.53	57.47	56.24	55.85	61.43	55.76

Table 27 - Performance of the different number of neurons in each hidden layer for Partner D

The performance metrics for the developed model are visualized in *Table 28*. The number of orders delivered within a three-hour time window is 82.5%, which is a good performance. Decreasing the time window to two hours has a great negative impact on the performance compared to increasing the window to four hours. This can be explained by the fact that the MAE is 54 minutes, which is close to one hour. Overall, the model achieves a good performance.

Performance Measurement	Neural Network
Four Hour Window	90.1%
Three Hour Window	82.5%
Two Hour Window	66.5%
MAE	54.0
MAPE	3.75%

Table 28 - Performance measurements for the ANN model for Partner E

4.7 Evaluation & Discussion

In this section, the methodology implemented is evaluated. The previous chapter handled all the modelling choices. This part evaluates the performance of the two implemented techniques.

The initial results for the current situation were on average far below the required service level. This is because the delivery performance was no measured KPI. Therefore, implementing the two prediction techniques should result in a great improvement in average performance for all the partners.

Currently, the time window is equally divided by adding and extracting 1.5 hours to the estimated time of arrival. This is named as the "naïve time window". However, a simple technique for improving the performance is by implementing an adjusted time window. This technique finds the optimal time window setting for every partner. The average performance compared to the current situation improved with 18%. This is since most of the orders are delivered later than the ETA value, instead of earlier. By setting for example an upper bound of 2 hours instead of 1.5 hours above the ETA, there are more orders that fall in the time period of 1.5 hours to 2 hours.

The first technique that was implemented was a multi-linear regression model. For every partner this model was developed with different parameters, since the initial performance show great variance between the partners. The case study company communicates a three-hour window with the customer for delivery, therefore the most important measurement is the number of orders that are delivered within this time window. The performance improved on average with 20 % compared to the current situation. In addition, the MAE decreased with 23.1% using the regression model.

Next, the ANN technique was implemented in the case study. For every partner the model was developed with different parameters and hidden layer sizes. The number of hidden layers used was in all models equal to two. Compared to the current situation, the developed model performance on average 24% better. The MAE decreased on average with 25.5%.

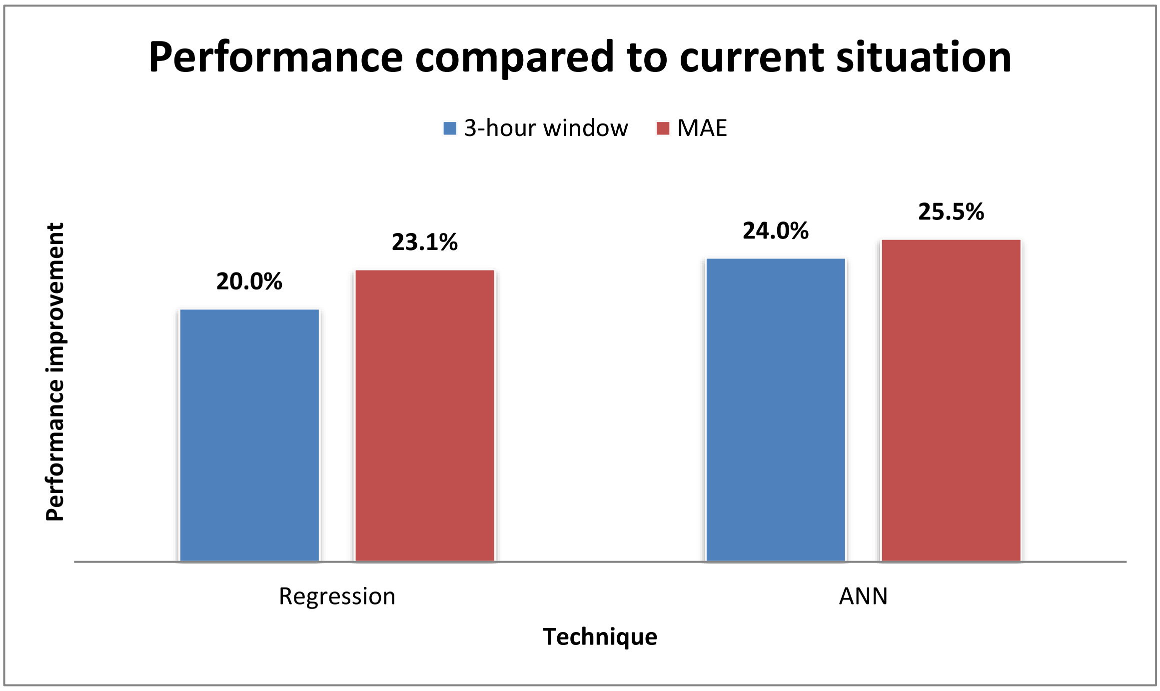


Figure 32 - Performance of the techniques compared to the current situation

Overall, it can be stated the ANN technique outperforms the current situation and the regression model based on the delivery performance in the three-hour window and the MAE. This was expected since neural networks have a major application in forecasting. ANNs are able to learn from a large dataset and recognize patterns easily. The average performance difference between the two implemented techniques is 3.2% for the three-hour delivery performance in favour of the ANN technique. For the MAE, the performance is 2.5% in favour of the ANN. *Figure 32* summarizes the performance improvements compared to the current situation.

Chapter 5

Tooling

In this section, the building of a tool for measuring the delivery performance is elaborated. The tool is intended specifically for the case study company in order to measure and improve their time window delivery performance. This is done by measuring different performance metrics.

5.1 Process building tool

For the case study company a tool for measuring the delivery performance of the partners is built in Excel using the programming language VBA. The choice for this software is because it is the most practical and suitable way, since other people will work with this tool. The first step is to analyse the requirements from the business, by talking with experienced employees from the case study company. After specifying all the relevant measurements and functionalities, the tool is built using reverse engineering. This is the process of extracting knowledge from the business and then building the tool based on the extracted information. This implies that first the interface with all the KPI's are specified, before any calculations or steps are made.

Building the model requires many calculations for the KPI's. In addition, the tool uses different macros to visualize and refresh the data. Based on the ETA, the upper and lower bound for delivery and the orders that are delivered within the time window can be calculated. Using different pivot tables, other KPI's can be specified in the tool. For a representative visualization, a chart is made such that the overall performance can be interpreted quickly.

5.2 Functionalities

The tool has the function to calculate the different KPI's that are specified below:

- *Delivery within 90 minutes*

This KPI stands for the number of orders delivered within a three-hour time window. In the initial situation a naïve time window is used. Therefore, the delivery within 90 minutes stands for 90 minutes above and below the expected time of arrival. In the tool the number of orders delivered with a MAE $\leq$ 90 minutes divided by the total number of orders delivered, between the same time period t, represent the value of this KPI (in percentage).

- *Delivery between 90 and 180 minutes*

The second KPI stands for the number of orders that have a MAE between 90 and 180 minutes. In practise this means that the order is delivered outside the time window. This KPI is calculated by dividing the number of orders delivered within a MAE between 90 and 180 minutes divided by the total number of orders delivered in the same time period t.

- *Delivery after 180 minutes*

This KPI is included because it suggests which orders are scanned out too late. It is very unlikely that a driver delivers his order more than three hours too late. Therefore, including this measurement will result in an overview of orders that are likely to be scanned out too late. The calculation is done by dividing the number of orders with a MAE > 180 minutes by the total number of orders in the same time period t.

- *Orders that are signed out too late and after 18:00*

Orders that are delivered after 18:00 suggest that the driver is not correctly scanning his orders. The planning of the routes is usually made between 07:00-16:00. Therefore, orders delivered outside this interval are mostly unreliable values.

- *Orders without ETA*

The case study company wants insight in the communicated ETA's from their partners. Therefore, this KPI is calculated using the tool. It represent the percentage of orders that are planned without communicating an ETA through. This is one of the most important KPI's, since this affects directly the customer and the case company. The customer cannot follow his order using the Track&Trace application, and the case company is spending much more time on answering questions about the delivery status.

- *Time deliveries*

The case study company provides a service to their customers for time deliveries. This might be orders that have to be delivered before 10:00, 12:00, or 14:00. These time deliveries are not part of this research, however they are included in the developed tool requested by the case study company.

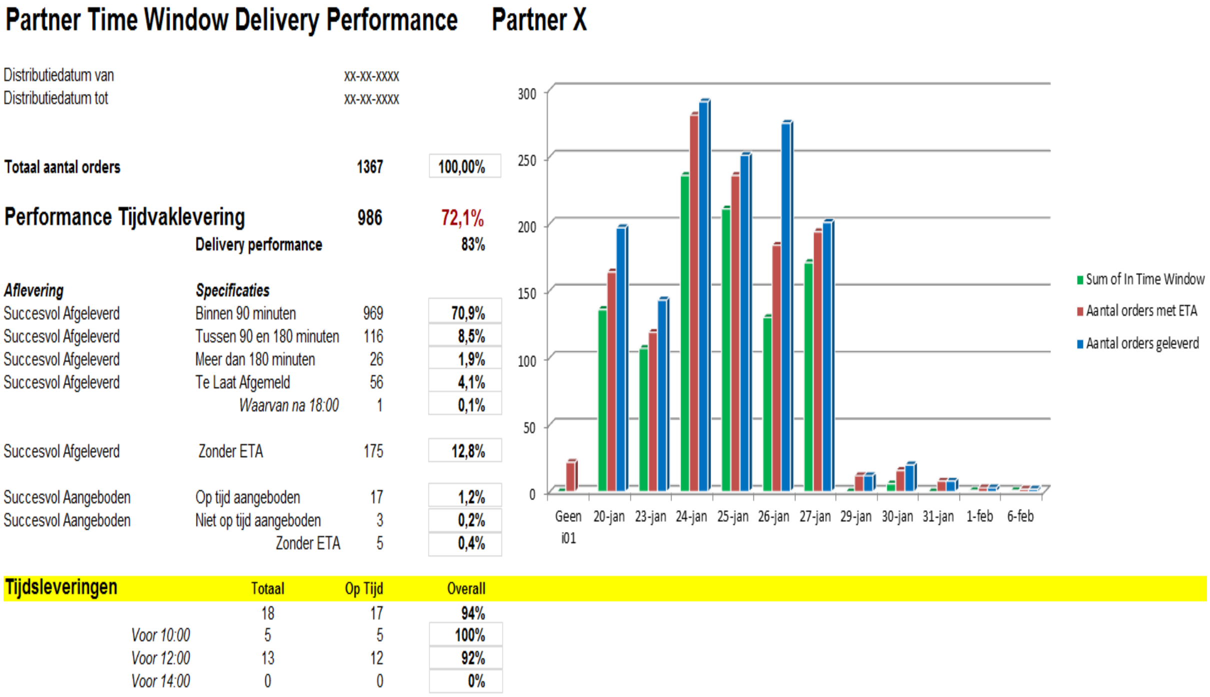
- *Orders without i01 (delivery scan)*

The last performance measurement the tool presents, are the orders that have no physical delivery scan. This means that these orders are not delivered or scanned. Since the case study company wants insight in their delivery performance this KPI is included.

The main function of the tooling is to present what the performance is, and how to improve it. Therefore, every percentage visualized in the tool has a button which can be clicked on and consequently presents the raw data in this percentage. In addition, by using the slicer of the pivot table, every report has the function to filter on the specific partner. By representing the performance in a graph form and visualizing it as daily cluster columns, the user and partner can easily see what the performance behaviour was on daily basis. The tool has also a macro for generating a report with only the raw data selected for the specific partner. This macro copies the excel report to outlook and the user can easily send it to the partner.

5.3 Interface

To give an idea on how the tool looks like, *Figure 33* represents an example interface when running the data for a specific partner. As can be seen, on the left side the real facts are presented in the form of percentages, and on the right side a chart is presented to visualize the orders delivered in the time window, orders with no ETA, and orders that are eventually delivered with a delivery scan (i01).



Partner Time Window Delivery Performance **Partner X**

Distributiedatum van		xx-xx-xxxx	
Distributiedatum tot		xx-xx-xxxx	
Totaal aantal orders		**1367**	**100,00%**
Performance Tijdvaklevering		**986**	**72,1%**
	Delivery performance		**83%**
Aflevering	***Specificaties***		
Succesvol Afgeleverd	Binnen 90 minuten	969	**70,9%**
Succesvol Afgeleverd	Tussen 90 en 180 minuten	116	**8,5%**
Succesvol Afgeleverd	Meer dan 180 minuten	26	**1,9%**
Succesvol Afgeleverd	Te Laat Afgemeld	56	**4,1%**
	Waarvan na 18:00	1	**0,1%**
Succesvol Afgeleverd	Zonder ETA	175	**12,8%**
Succesvol Aangeboden	Op tijd aangeboden	17	**1,2%**
Succesvol Aangeboden	Niet op tijd aangeboden	3	**0,2%**
	Zonder ETA	5	**0,4%**

Tijdsleveringen	Totaal	Op Tijd	Overall
	18	17	**94%**
Voor 10:00	5	5	**100%**
Voor 12:00	13	12	**92%**
Voor 14:00	0	0	**0%**

Figure 33 - Interface of the developed tool

5.4 Testing

To make the functionalities and interface as user friendly as possible, the tool is evaluated with the Chain&Partner Manager. Eventually, he will be the user of the tool and distribute the reports to the specific partners. After discussing the concept version of the tool, a final version was made with valuable input from the manager.

In addition, the tool is used in a pilot case to improve the delivery performance at two partners. This pilot case started on 3 April 2017 and will take until 1 May 2017, therefore the results are not included in this research. The main goal of the tool is to increase the number of orders communicated with ETA and improve the scan performance. Using the tool will help to get a good and reliable overview of the current situation and show the specific orders with irregularities. It can be both used on daily and on weekly basis. In the pilot case the tool is used to generate results on daily basis. However, in the future the Partner Manager will use the tooling to generate summarized performance reports on weekly basis.

Chapter 6

Conclusions

This section concludes this report. First the research questions as presented in *Chapter 1* are answered. Next, the practical implications and recommendations for the case study company are presented. Finally, the limitations of this study are addresses and future research directions are indicated.

6.1 Research Question

The goal of this research is to improve the prediction of reliable time window estimates for the delivery of customer orders. The main research question was therefore:

How can accurate and reliable prediction estimates for delivery time windows be generated?

The goal of the first sub-question was to get an overview of the different methods that are currently available for predicting delivery time windows. This was answered by conducting a literature study to get insights in the different techniques available for predicting delivery time windows. Two techniques were selected; multi-linear regression and neural networks. Using input from two articles, two different models were build based on the techniques.

The second sub-question searches for the factors that influence the actual ETA. To assess which parameters should be included in the developed models, the factors that influence the actual delivery time must be discussed. With input from literature and from experienced employees from the case study company, the second sub-research question is answered. In total eight different parameters were presented as potential variables to include in the models. Two of these variables are based on the concept of address intelligence presented in the article of van Duin et al. (2015). The other six parameters were found by using the input from employees and assessing which data is available in the TMS of the case study company.

The third sub-question searches for ways to make the predicted delivery time windows more accurate. The techniques can be implemented in different ways. The multi-linear regression model tests the relationship between the target variable and the independent variables. Adding or deleting parameters from the model could improve the time window estimates. Therefore, the variables that were not significant are deleted from the final model. In the neural networks model the parameters are selected based on feature selection. The optimal combination of parameters in the model are selected, consequently these variables are used for building the final model. In addition, using different layers and nodes in the ANN model the performance can be improved. This answer the third sub-research question which was related to make the delivery time window predictions more accurate.

The last sub-research question is associated with the case study. In order to evaluate the different techniques a dataset was generated to study the performance. First, the initial situation at the case study company was analysed using a naïve and an adjusted time window, since this will be used as benchmark. This was done by using the following performance measurements; number of orders delivered within the time window, orders delivered outside the time window, MAE, and the MAPE. In addition, different time windows are used which are respectively a two, three, and four-hour window of delivery.

For the current situation, the adjusted time window outperforms the initial situation with 18% in the three-hour time window. The regression model achieves a 20% performance increase, and the ANN model 24%. For the MAE, this stays the same using an adjusted time window, since only the window around the ETA is adjusted. However, the regression model performance 23.1% better than the current situation and the ANN model 25.5% using the MAE as a performance metric. The summarized results are visualized in *Figure 34*.

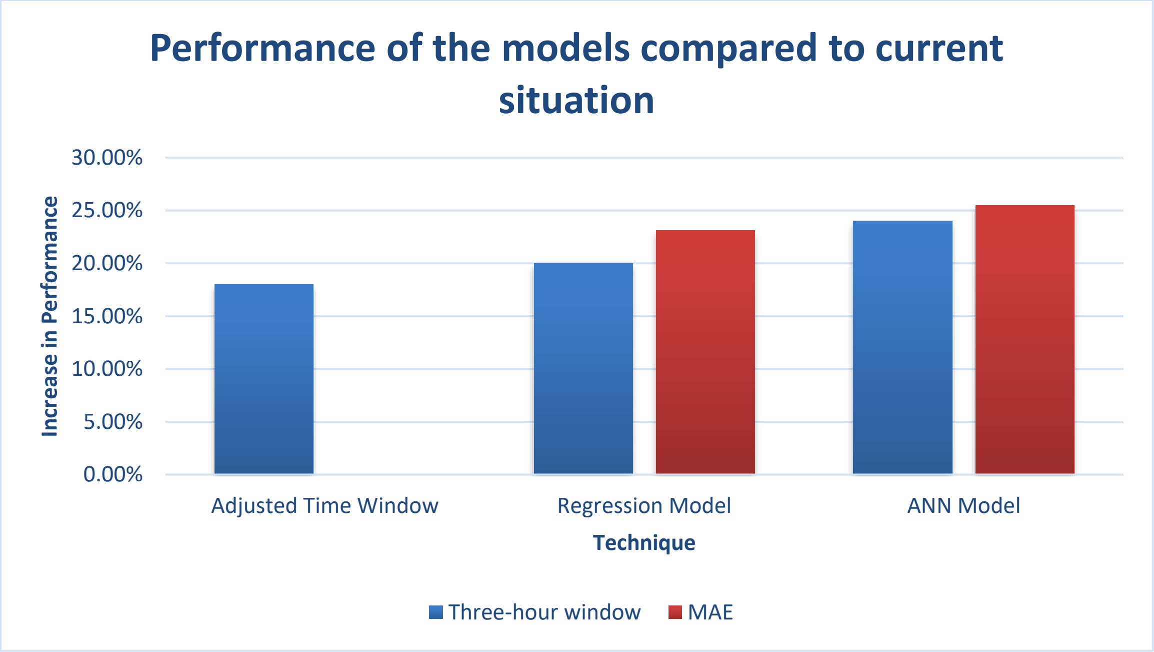


Figure 34 - Performance of the different techniques in a three-hour time window

The two techniques implemented outperform the current situation easily. However, implementing these models into practice will be very difficult for the case study company. The 'ETA's' are predicted using a planning software, and in addition communicated with the case study company using different IT applications. The planning software and the IT applications could be different for every partner. Therefore, it is very complicated to apply these techniques into the daily practice of the partners. However, the adjusted time window can be implemented by applying a different window around the ETA. This is a cheap and easy way to increase the delivery performance at the case study company.

Another practical implication is to use the tool to measure the delivery performance at every partner. This is not only built to measure the performance, but also to improve it by giving practical feedback tot the planners . It is user friendly and gives an overview on a high level but also on the lower level by zooming into detail. It is also designed to combine different reports into one file, since all the report are built using the same dataset extracted from the TMS.

This research was conducted at a freight distribution company. Therefore, most of the results are only applicable for the case study company. However, since this research tested two scientific techniques for predicting delivery time windows, some of the results are generalizable. One important note is the fact that the ANN model outperforms the regression technique and the performance of the current situation. The models are developed using a representative dataset with enough potential parameters. Therefore, it can be concluded that the ANN model can be used for predicting accurate delivery time windows in the transportation industry. As regards to the regression model, the performance was also good. It can be expected that the parameters used to build the models can easily be extracted from a company's TMS. Therefore, the results based upon the developed models are generalizable for predicting delivery time windows.

6.2 Practical implications & Recommendations

The two techniques perform much better compared to the current situation. The neural network model outperforms the regression technique as expected. An interesting insight is the fact that when implementing an adjusted time window instead of using a naïve time window, the performance gets close to the performance of the regression technique. The implementation of an adjusted time window would have two major benefits. The first one is the implementation, since it is very easy to apply this in daily practice. The other benefit is performance related. Given the current situation, adjusting the time window would result in a performance increase of 21.1%. This would be the case when applying different adjusted time windows for every partner. However, this will be too complex to implement into the daily practice of the company. Therefore, an overall adjusted time window will be implemented which is the same for every partner. Instead of using a naïve three-hour time window, the time window is built by extracting one hour and adding two hour to the ETA. This results in an overall increase in delivery performance of 6.7% for the three-hour window.

Figure 35 represent the overall performance for the different adjusted time windows. The optimal time window setting lies by extracting 0.75 hours from the ETA and adding 2.25 hours. However, this research created awareness at the case study company to improve the delivery performance, therefore the performance will improve in the future. To create some flexibility in the adjusted time window, one hour is extracted from the ETA and two hours are added.

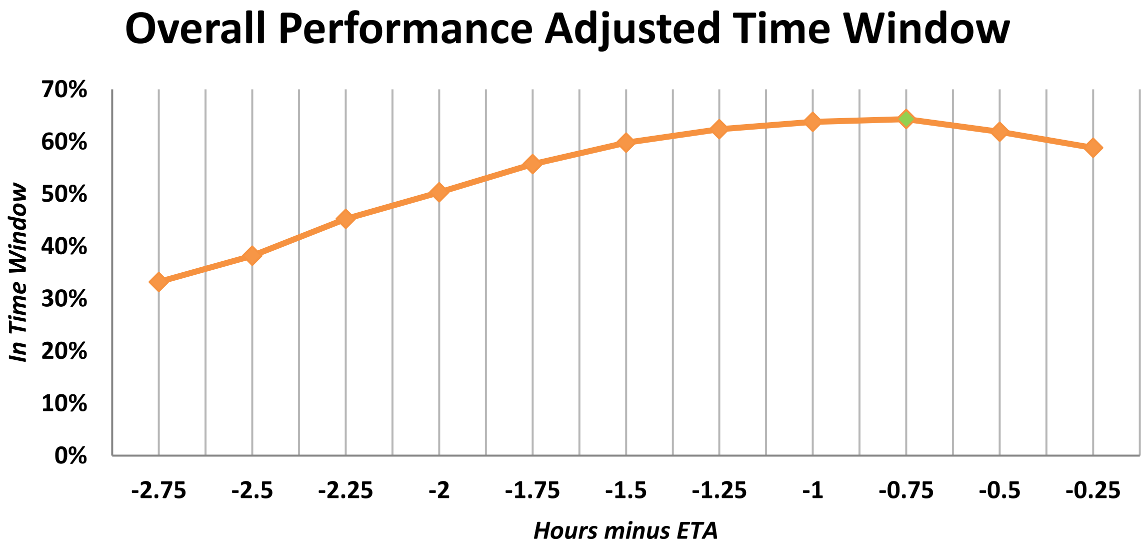


Figure 35 - Overall performance of adjusted time windows

Using and creating reliable data is the key to improve your customer satisfaction. In this research, a lot of time and effort is spent on data preparation. A great part of the initial dataset is not used for analysis, since this contained missing values or unreliable measurement. A recommendation for the case study company is to improve the communication of ETA's and i01 events (Delivery scan). Around 60% of the initial dataset is deleted due to the missing of ETA's or i01 events. Setting KPI's for these two events will result in a greater awareness at the partners for communicating these values with the case study company. Using the developed tool, these two events can be measured for every partner. The implementation of these KPI measurements will result in a performance improvement which directly leads to a decrease in costs. The customer service agents and the planners will focus less on answering questions about planned or delivered orders, instead they have time to focus on other activities. In addition, the customer benefits the most of this since the information communicated with him will be more complete and reliable.

Another practical implementation is to use the developed tool that helps the company to improve their delivery performance. The tool is designed to zoom into the details and therefore can find orders and routes that are delivered far too late. In addition, the tool can find the orders that are communicated without ETA or orders without delivery scan. By creating these insights, the case study company can directly react on these (structural) problems. The tool is designed for two purposes; measuring and improving the delivery time window performance. The interface is created with help of the Chain&Partner manager of the case company and every employee at the case study company could use it. However, it is specifically designed to communicate the performance with the partners. By giving them the possibility to see what orders were delivered too late, they could directly react on it.

6.3 Limitations

One of the main limitations in this research is the fact that the delivery scan is not always a reliable event. For building the two models the ETA and the actual delivery time are needed. However, the actual delivery time is represented as the i01 event. This i01 event is done by the driver of the truck using a PDA or app, which suggest that the human factor here is very important. Scanning the order too late or after the end of the route results in unreliable data. This could influence the performance of the models. Orders that are planned in the morning but delivered in the evening can be classified as unreliable. In addition, orders that get a delivery scan after 18:00 can also be seen as unreliable. The planner plans his orders spread over the day until around 16:00. It is very unlikely that an order is delivered after 18:00. Orders that are delivered more than 5 hours too late can also be classified as unreliable. In this research the data is cleaned in order to overcome this problem. However, it is still difficult to completely overcome this problem, since the delivery scan represent the actual time of delivery of the order. Therefore, the human factor is still applicable and could have influenced more data points that were used for building the prediction models.

Another limitation is the usage of limited information. Most of the variables that were chosen to test in this research, came out of the availability in the initial dataset. There are many other variables that could influence the actual time of arrival, however these cannot be extracted from the TMS. For example, the specifications of the driver or the truck that he drives could have influenced the delivery performance.

6.4 Future Research

Future research can focus on the implementation of other machine learning methods. Since the goal of this research was to measure an improve the delivery time window performance, it is very probable that the developed models can be improved further. However, the gains are expected to be marginal since the used methods achieved very good results. Still, there are most likely optimizations possible. It would be especially helpful if the prediction techniques would first classify the orders in dayparts, as morning and afternoon delivery. There is a difference in the performance considering the orders delivered in the morning and afternoon. This classification could be used to increase the effectiveness of the proposed techniques.

A practical optimization possibility could be to focus on the ETA’s created by the partners. Why are certain orders consequently delivered outside the time window? It was already known that orders that are planned early in the morning achieve a higher performance. Also, the occurrence of certain orders almost always leads to the delivery outside the time window. Further analysis in the prediction of time windows can lead to interesting insights, that allow for improving the delivery performance.

Bibliography

Aranko, J. (2013). *Developing the last mile of a parcel delivery service concept for consumers.*

Barron, A. (1994). A comment on "Neural Networks: A review from a statistical perspective". *Statistical Science 9*, 33-35.

Boyer, K., Prud'homme, A., & Chung, W. (2009). The last mile challenge: evaluating the effects of customer density and delivery window patterns. *Journal of Business Logistics 30(1)*, 185-201.

Bushuev, M., & Guiffrida, A. (2012). Optimal position of supply chain delivery time window: Concepts and general conditions. *International Journal Production Economics*, 226-234.

Corbett, L. (1992). Delivery Windows - A new view on improving manufacturing flexibility and on time delivery performance. *Production and Inventory Management Journal 33(3)*, 74-79.

Du, T., Li, E., & Chou, D. (2005). Dynamic vehicle routing for online B2C delivery. *The International Journal of Management Science*, 33-45.

Elbert, R., Thiel, D., & Reinhardt, D. (2016). Delivery Time Windows for Raod Freight Carriers and Forwarders - Influence of Delivery Time Window on the Costs of Road Transport Services. Darmstadt: Springer International Publishing Switzerland.

Fabri, A., & Recht, P. (2005). On dynamic pickup and delivery vehicle routing with several time windows and waiting times. *Transportation Research*, 335-350.

Genhong, D., Linye, L., & Yao, J. (2009). Multi-strategy grouping genetic algorithm for the pickup and delivery problem with time windows. *Proceedings of the first ACM/SIGEVO summit on Genetic and Evolutionary Computation* (pp. 97-104). New York: ACM.

Ghannadpour, S., Noori, S., Tavakkoli-Moghaddam, T., & Ghoseiri, K. (2014). A multi-objective dynamic vehicle routing problem with fuzzy time windows: Model, solution and application. *Applied Soft Computing, Volume 14, Part C*, 504-527.

Grigori, D., Casati, F., Castellanos, M., Dayal, U., Sayal, M., & Shan, M. (2004). Business Process Intelligence. *Computers in Industry Volume 53, Issue 3*, 321-343.

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques.* Elsevier.

Hosny, I., & Mumford, L. (2007). Single vehicle pickup and delivery with time windows: made to measure genetic encoding and operators. *Proceedings of the 9th annual conference companion on Genetic and evolutionary computation*, 2489-2496.

Jeong, R., & Rilett, L. (2004). Bus Arrival Time Prediction Using Artificial Neural Network Model. *Intelligent Transportation Systems Conference.* Washinton, D.C.: IEEE.

Jung, H., Lee, K., & Chun, W. (2006). Integration of GIS, GPS, and optimization technologies for effective control of parcel delivery service. *Postal technology research center*, 305-350.

Kozlak, J. (2006). Multi-agent Approach to Dynamic Pick-up and Delivery Problem with Uncertain Knowledge about Future Transport Demands. *Fundamenta Informaticae 71*, 27-36.

Lapedes, A., & Farber, R. (1988). How neural nets work. *In: Anderson, D.Z., (Ed.). Neural Information Processing Systems* (pp. 442-456). New York: American Institute of Physics.

Li, B., Riley, M., Lin, B., & Qi, E. (2006). A comparison study of customer satisfaction between the UPS and FedEX: an empirical study among university customers. *Industrial Management & Data Systems 106.2*, 182-199.

Lin, C. (2011). A Vehicle routing problem with pickup and delivery time windows, and coordination of transportable resources . *Computers & Operations Research 38*, 1596-1609.

Lippmann, R. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine*, 4-22.

Morse, P., & Kimball, G. (2012). *Methods for operations research.* Courier Corporation.

Peffers, K., T, T., Gengler, C., Rossi, M., Hui, W., Virtanen, V., & Bragge, J. (2006). The design science research process: a model for producing and presenting information systems research. *In Proceedings of the first international conference on design science research in information systems and technology* (pp. 83-106). Claremont: Desrist.

Reilly, D., & Cooper, L. (1990). An overview of neural networks: early models to real world systems. *An Introduction to Neural and Electronic Networks* (pp. 227-248). New York: Academic Press.

Ropke, S., Cordeau, J., & Laporte, G. (2007). Models and branch-and-cut algorithms for pickup and delivery problems with time windows. *Networks*, 258-272.

Sharda, R. (1994). Neural Networks for the MS/OR analyst: An application bibliography . *Interfaces 24 (2)*, 116-130.

Sprenger, R., & Mönch, L. (2009). An ant colony optimization approach to solve cooperative transportation planning problems. *Winter Simulation Conference* (pp. 2488-2495). Hagen: Winter Simulation COnference.

Swingler, K. (1996). Chapter 1: Introduction. In K. Swingler, *Applying Neural Networks: a Practical Guide* (pp. 3-21). London: Academic Press Ltd.

van Duin, J., de Goffau, W., Wiegmans, B., Tavasszy, L., & Saes, M. (2015). Improving home delivery efficiency by using principle of address interlligence for B2C deliveries. *The 9th International Conference on City Logistics* (pp. 14-25). Tenerife: Transportation Reserach Procedia 12.

Widrow, B., Rummelhart, D., & Lehr, M. (1994). Neural Networks: Application in industry, business and science. *Communication of the ACM 37 (3)*, 93-105.

Zhang, G., Eddy Pattuwo, B., & Hu, M. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting (14)* , 35-62.

Appendix

A Initial Regression & Correlation Output

SUMMARY OUTPUT

PARTNER A

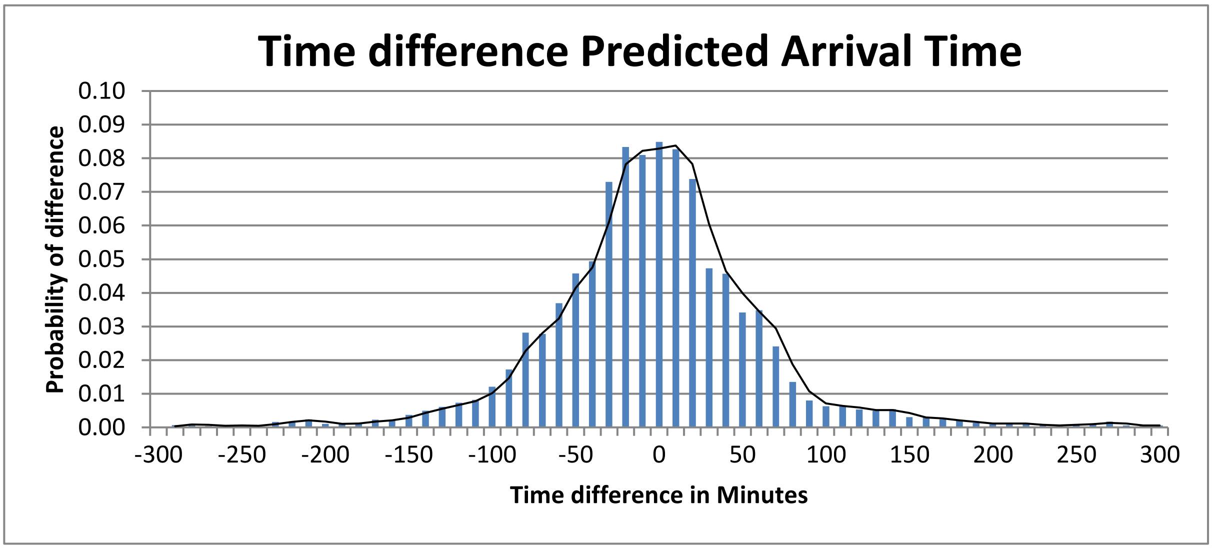
Regression Statistics	
Multiple R	0,99999877
R Square	0,99999754
Adjusted R Square	0,999997538
Standard Error	0,044927411
Observations	xxxxx

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0,151155569	0,774391758	0,195192636	0,845247413
product_code	1,8393E-05	7,35557E-06	2,500548361	0,01242122
To Zipcode	1,00551E-06	4,42182E-07	2,27397919	0,022995274
Day of the Week	0,002564744	0,000388332	6,604516941	4,25953E-11
ETA	0,999994779	1,81298E-05	55157,40127	0
rolcontainer	0,001879266	0,000728154	2,580861681	0,009874165
pal	0,001588381	0,000485489	3,271713917	0,001073793

Regression output

	Product Code	*To Zip code*	*Day of the Week*	*ETA*	*RC*	*Pallet*
Product Code	1					
To Zipcode	0.077	1				
Day of the Week	0.043	0.037	1			
ETA	-0.025	0.034	0.054	1		
RC	0.317	0.019	-0.013	-0.015	1	
Pallet	-0.109	-0.195	-0.020	0.013	-0.155	1

Correlation coefficient



Distribution of regression model

SUMMARY OUTPUT
PARTNER B

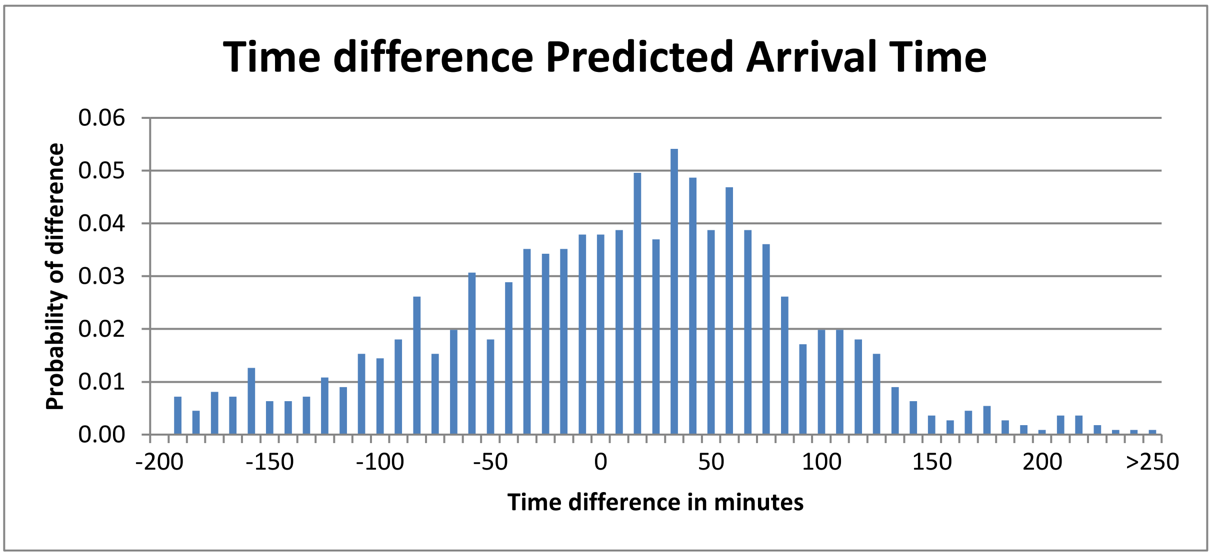
Regression Statistics	
Multiple R	0,999997
R Square	0,999994
Adjusted R Square	0,999994
Standard Error	0,055754
Observations	xxxxx

	Coefficients	Standard Error	t Stat	P-value
Intercept	-5,24851	1,270643	-4,13059	3,66E-05
product_code	-1,9E-05	6,64E-06	-2,79856	0,005148
ETA	1,000126	2,98E-05	33614,06	0

Regression output

	Product Code	*ETA*
Product Code	1	
ETA	-0.044	1

Correlation coefficient



Distribution of regression model

SUMMARY OUTPUT
PARTNER C

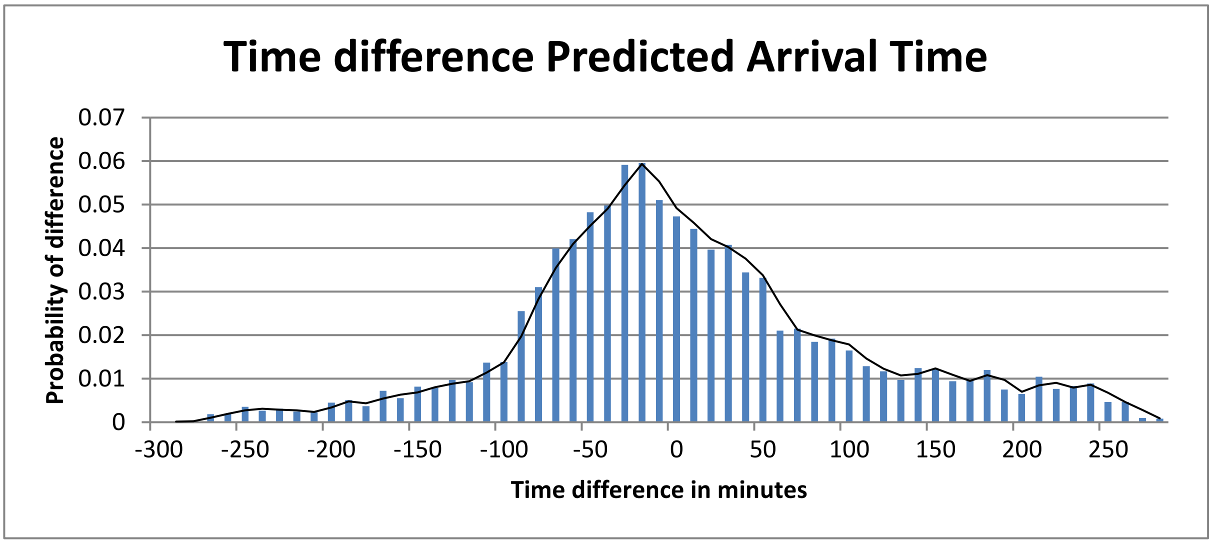
Regression Statistics	
Multiple R	0,999996
R Square	0,999992
Adjusted R Square	0,999992
Standard Error	0,068198
Observations	xxxxx

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-6,72568	1,478944	-4,54762	5,5158E-06
To Zipcode	-1,2E-05	7,96E-07	-14,5156	4,55234E-47
From Zipcode	7,35E-08	2,85E-08	2,582059	0,009841164
collo	-0,00056	0,000167	-3,36359	0,000773435
ETA	1,000159	3,47E-05	28847,63	0

Regression output

	To Zip code	*From Zip code*	*Collo*	*ETA*
To Zip Code	1			
From Zip Code	0.037	1		
Collo	-0.015	-0.013	1	
ETA	0.058	0.019	-0.011	1

Correlation coefficient



Distribution of regression model

SUMMARY OUTPUT
PARTNER D

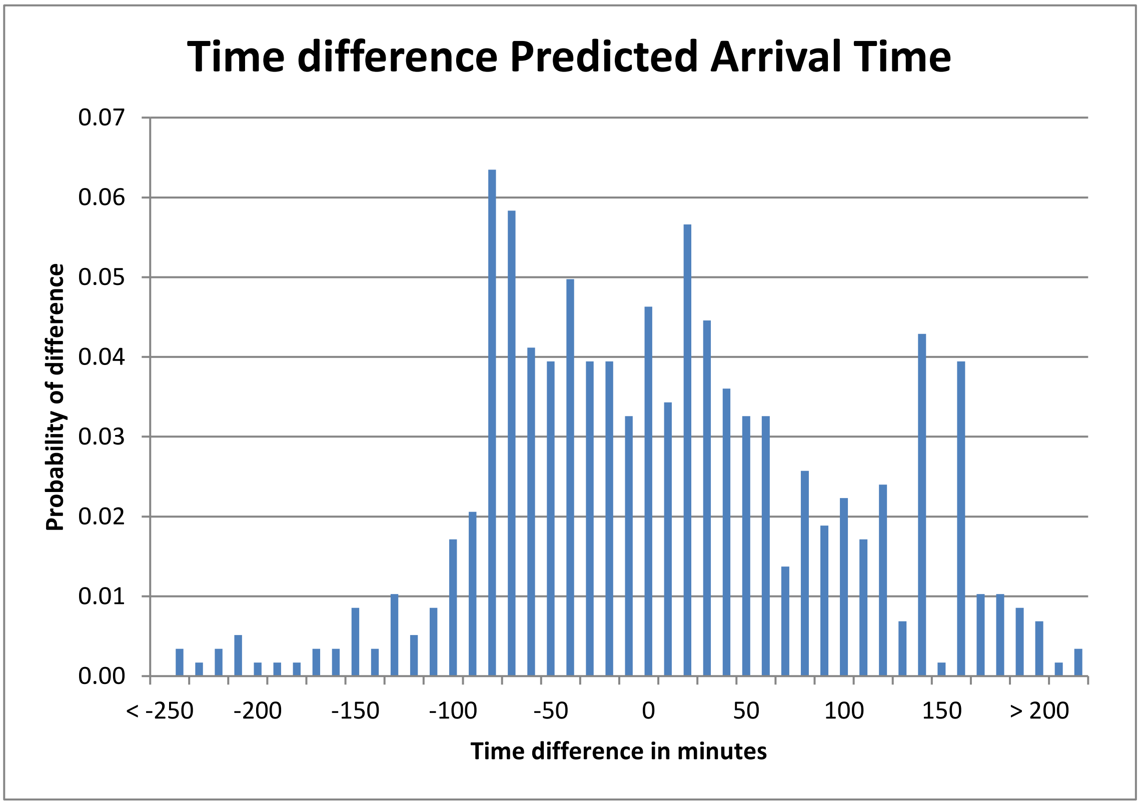
Regression Statistics	
Multiple R	0,999987
R Square	0,999974
Adjusted R Square	0,999974
Standard Error	0,066881
Observations	xxxxx

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	29,20205	4,654374	6,274109	4,23224E-10
To Zipcode	-1E-05	2,06E-06	-4,86762	1,21088E-06
From Zipcode	2,41E-07	5,15E-08	4,676745	3,09388E-06
ETA	0,999318	0,000109	9163,525	0

Regression output

	To Zip code	*From Zip code*	*ETA*
To Zip Code	1		
From Zip Code	-0.123	1	
ETA	0.007	0.110	1

Correlation coefficient



Distribution of regression model

SUMMARY OUTPUT
PARTNER E

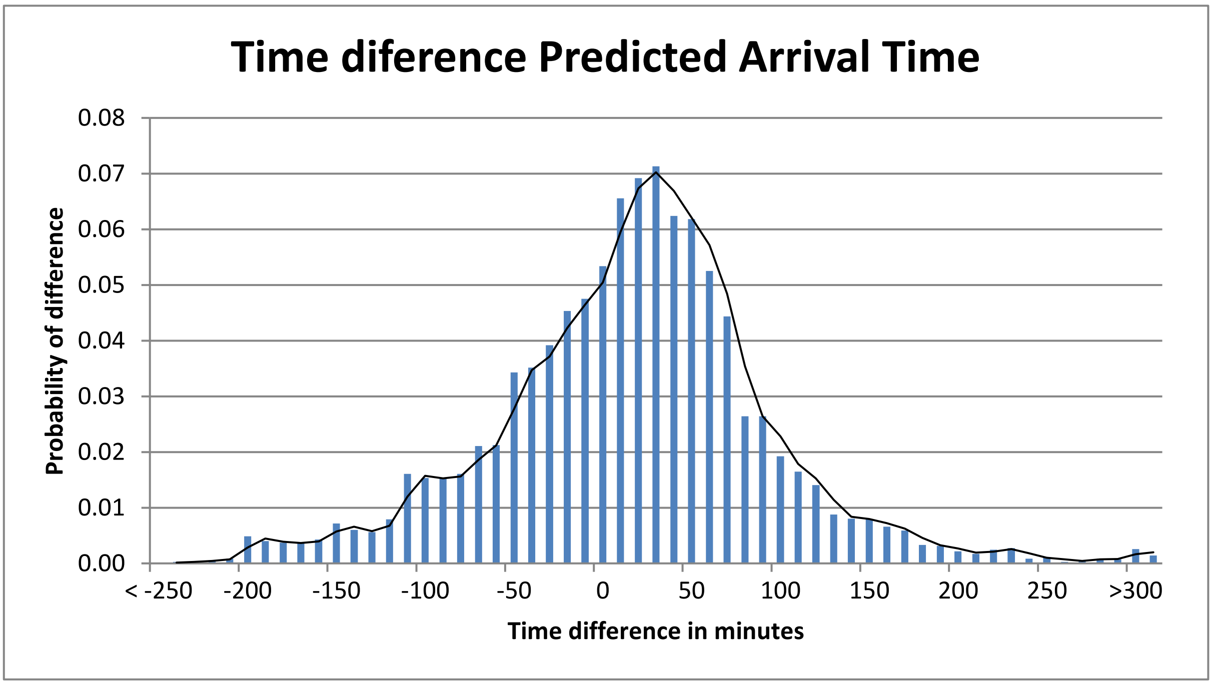
Regression Statistics	
Multiple R	0,999997
R Square	0,999995
Adjusted R Square	0,999995
Standard Error	0,05353
Observations	xxxxx

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-0,09922	1,150653	-0,08623	0,931289025
To Zipcode	-7,1E-06	2,66E-07	-26,7477	4,0957E-150
Day of the Week	-0,00208	0,000468	-4,43997	9,13613E-06
ETA	1,000004	2,7E-05	37076,18	0
collo	-0,00044	0,000191	-2,32052	0,02034176

Regression output

	To Zip code	*Day of the Week*	*ETA*	*Collo*
To Zip Code	1			
Day of the Week	0.002	1		
ETA	0.002	-0.036	1	
Collo	0.001	0.003	-0.005	1

Correlation Coefficient



Distribution of regression model

B Programming Code MATLAB

```
%Dividing data into input and output variable
%Dates need to be imported as a datetime type of variable, not in cell form
   termInput(:,1) = collo(:,1);
   termInput(:,2) = DayoftheWeek(:,1);
   termInput(:,3) = datenum(ETA(:,1));
   termInput(:,4) = FromZipcode(:,1);
   termInput(:,5) = pal(:,1);
   termInput(:,6) = product_code(:,1);
   termInput(:,7) = rolcontainer(:,1);
   termInput(:,8) = ToZipcode(:,1);
   termOutput = datenum(I01Event(:,1));

   Input2 = termInput(:,:);

    %%%%%%%%%%%%%%%%%%%NEURAL NETWORK CREATION%%%%%%%%%%%%%%%%
% The amount of times the model needs to run
for NumberOfRuns=[1:100];
% To be able to test multiple different inputs in one run
for InputVariables=[];
    termInput = Input2(:,[InputVariables 1:8]);
    %The amount of neurons in layer 1
    for LayerSize1 = [10];
        %The amount of neurons in layer 2
            for LayerSize2 = [10];
     %To see the progress while the program is running
    [NumberOfRuns LayerSize1 LayerSize2 InputVariables]

    x = termInput';
    t = termOutput';

% Create a Pattern Recognition Network
hiddenLayerSize = LayerSize1 + LayerSize2;

%For network with 2 layers
net = fitnet([LayerSize1 LayerSize2]);
%%For retraining
%%net = init(net);

%Amount of layers (+input layer)
net.numLayers = 3;

% Connection between layers
net.layerConnect = [0, 0, 0; 1, 0, 0; 0, 1,0];
%If a layer has a bias
net.biasConnect = [1;1;1];

% Setup Division of Data for Training, Validation, Testing
net.divideParam.trainRatio = 70/100;
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;
```

```
% Train the Network
[net,tr] = train(net,x,t);

% Test the Network
y = net(x);
e = gsubtract(t,y);
tind = vec2ind(t);
yind = vec2ind(y);
percentErrors = sum(tind ~= yind)/numel(tind);

%t2 is the real result y2 is the network output
t2 = t(:,tr.testInd);
y2 = y(:,tr.testInd);
error = t2-y2;
performance = mae(error)

%Store performance with 3 forloops
PerformanceAvg(LayerSize1,LayerSize2) = performance;
PerformanceStruct.a{NumberOfRuns} = PerformanceAvg;

%Store performance with 2 forloops
performanceTotal(NumberOfRuns,InputVariables) = performance;
    end
end
end
end
%Averaging the data for 10 iterations with 2 forloops
performanceTotalAvg = sum(performanceTotal)/10;
min(performanceTotalAvg)
%Convert data into excel date
 y = y';
 DateStr = datestr((y),' dd-mm-yyyy HH:MM:SS');

 %Avg with 3 forloops
 %for LayerSize1 = [3 5 7 10 15 20 25 50];
  %  for LayerSize2 = [3 5 7 10 15 20 25 50];
   %      for NumberOfRuns = [1:10];
    %
SumVector(NumberOfRuns,:)=PerformanceStruct.a{1,NumberOfRuns}(LayerSize1,La
yerSize2);
     %    end
%ResultingAvgPerformance(LayerSize1,LayerSize2) =sum(SumVector)/10;
 %    end
  %end

%% Use the Validation data to evaluate the model
 valInd = tr.testInd';
 for z=1:n %% Where n is 15% of the data used
     Outcome(valInd(z,:),:) = I01Event(valInd(z,:),:);
 end
 Outcome = Outcome(valInd,:);
 OutcomeStr = datestr((Outcome),' dd-mm-yyyy HH:MM:SS');
```