ETA prediction for containerships at the Port of Rotterdam using Machine Learning Techniques

Ioannis Parolas – Master Thesis

August, 2016

TU Delft

Delft University of Technology

TNO innovation for life

ETA prediction for containerships at the Port of Rotterdam using Machine Learning Techniques

Master Thesis

Personal Details

Author: Ioannis Parolas

Student Number: 4416597

E-mail: john.parolas@gmail.com

Program: MSc Management of Technology

Graduation Committee:

Prof.dr.ir. L.A. Tavasszy

      Delft University of Technology & TNO

      Faculty of Technology, Policy & Management (TBM)

      Department of Transport & Logistics (T&L)

Dr. J.H.R. van Duin
      Delft University of Technology

      Faculty of Technology, Policy & Management (TBM)

      Department of Transport & Logistics (T&L)

Dr. ing. V.E. Scholten
      Delft University of Technology

      Faculty of Technology, Policy & Management (TBM)

      Department of Technology-Based Entrepreneurship (ETI)

W.M.M. Hopman MSc (TNO)

M.R. Rajagopal MSc   (TNO)

TUDelft

Delft University of Technology

TNO innovation for life

Abstract

The hinterland transportation of incoming containers at container terminals is a complex problem, due to the various actors involved and their often conflicting interests. A promising solution towards the problem for hinterland network operators is that of synchromodality, a concept that refers to on-line network planning for hinterland transportation. However, a hindrance to the efficient planning and execution of hinterland transportation is that there is currently no accurate way of predicting the estimated time of arrivals (ETA) for containerships that are reaching container terminals. This results in huge uncertainty over the types and amounts of cargo that reach the terminals, which in turn hinders the fast and cost efficient distribution of the products to inland destinations through trucks, trains or barges.

The current paper will propose a machine learning approach for predicting the ETA of containerships heading towards the Port of Rotterdam, by combining position data from GPS signals with weather predictions. It was found that significant improvement for the ETA predictions, compared to the current situation could be achieved, especially for the cases of the vessels that are more than 60 hours away from the port. Furthermore, the weather interpretation was not of significant importance for estimating the time of vessel arrivals at the port. The value of such an information tool for the various stakeholders involved was also investigated. The interested parties, for which the importance of ETA predictions of sea vessels was assessed are : terminal operators (European Container Terminals in the case at hand), hinterland transportation companies (e.g. European Gateway Services), the Port of Rotterdam, carriers and importers.

Summary

Uncertainty over vessel arrival times is a major hindrance towards the planning activities of the stakeholders involved in container transport. The current expectations regarding vessel arrivals are based on the estimated time of arrival (ETA) of the ship's agent, which is a representative of the carrier company at the port. However, this ETA is not frequently updated and contains large deviations from reality, especially for long time-horizons. These deviations from the expected time of arrival result in difficulties for the planning activities of the stakeholders involved in container transport, such as hinterland transportation parties, since vessel arrival is the starting point of the inland transportation of goods.

The present thesis investigates the possibility of developing an ETA information tool, that will reduce the prediction errors of vessel arrivals compared to the current situation. The route that was taken under consideration was the Asia-Rotterdam route, which is a very frequent one for container transport. The time horizon was accounting for the last 5 days before arrival to the port, excluding the Port operations time. The value of such an information tool is also investigated from the perspective of carriers, container terminals, hinterland transportation parties, the Port of Rotterdam and importers, in terms of its significance for improving their planning activities. Towards that goal, a literature review on what has been attempted so far for estimating a vessel's ETA, was carried out. Through the literature review, two appropriate machine learning methods for tackling the problem were selected, the neural networks and support vector machines.

The variables that were identified as relevant for addressing the problem of ETA predictions were selected from the big data available from marine traffic providers. Those were position and speed data from past voyages, as given in the AIS data, alongside with some technical characteristics of the vessel such as ship length and breadth. These data were combined with weather data from those voyages, with the aim of forecasting the time of vessel arrivals at the Port of Rotterdam. The historical data were used for training the neural networks and support vector machines and the accuracy of the prediction methods was evaluated based on error metrics. It was found that significant improvement compared to the current ETA estimation based on the ship's agent was achieved through both methods, especially for the medium to long time horizons (more than 50-60 hours away from the port). Moreover, the support vector machines were found to outperform the neural networks in all cases.

The dependence of prediction to input variables was also investigated, the main findings being that the ETA of the ship agent was an important variable for improving the accuracy of the models, when provided as an input, and that the weather variables were not assisting the ETA estimation to a sufficient extent. This can be explained by the fact that the ETA of the ship agent is giving an indication of the captain's behavior for the voyage, while on the other hand, the weather conditions are already partially interpreted in the speed that was used as input for the models, which was the speed over ground. Also, the captains have the ability of adjusting the engine power to counter adverse weather conditions, which would be extremely difficult to realize only from the weather conditions along the route of the vessel.

The value that this ETA information tool would have for the various stakeholders involved in container transport was also investigated, through a literature review and interviews conducted with representatives from the ECT container terminal at the Port of Rotterdam and consultants specialized in the operational activities of the Port of Rotterdam and hinterland transportation parties. The main findings were:

- Hinterland transportation parties can be greatly benefited due to the cost reduction associated with assigning more containers to barges or trains. A better estimation of vessel arrival for the long-time horizon that was achieved in this case enables better estimation of the needs in capacity for barges and rails for the week ahead.
- Container terminals can plan their schedule better for berth allocation and allocation of manpower and equipment to the unloading of the vessels.
- Carriers get a better indication of whether they will be arriving on time for delivering the containers or if there is a need to speed up. The biggest gain though can be achieved by using the ETA information as a competition monitoring tool, for determining the vessel availability in the area for transporting cargo, something which will give them higher bargaining power when making agreements for freight transport.
- The Port of Rotterdam is the main benefactor of the predictions. It can realize both direct and indirect benefits through its usage, the direct being scheduling better the allocation of pilots and its resources due to the reduced uncertainty over vessel arrivals. The indirect benefits is the enhanced competitive advantage that it can achieve by providing more reliable services to the carriers and having a more cost-efficient connection to the hinterland.

Due to its leading position in the transportation chain, the Port of Rotterdam should be the main benefactor of this ETA information tool, sharing the information with the other stakeholders, such as terminals and hinterland transportation parties, through a common platform. This can be achieved through the information provided by Intertransis, an information broker company in the field of logistics. One of the most important findings of this study was that the AIS data alone, are enough for making ETA predictions for the route and time-horizon examined. Therefore, there is currently no additional cost involved for Intertransis to acquire the necessary data. What is needed is for the real-time AIS data received, to be fed forward to the data pre-processing algorithm in order to select the variables that were used for making ETA predictions in this thesis. Then, the input variables can be provided to the already trained SVM algorithm, which will produce the ETA prediction. This procedure requires no additional manpower or cost, other than connecting the data pre-processing algorithm to the AIS real-time data receiver.

Table of Contents

Chapter 1 - Introduction

This chapter introduces the problem of uncertainty of vessel arrivals at the Port of Rotterdam and the research aim of the project. First, a general introduction to the challenges posed to the Port of Rotterdam for its efficient planning of its operating activities is provided in section 1.1. Then, section 1.2 describes the practical problem of estimating ship arrivals at the port, whereas section 1.3 presents the scientific aspects of the problem. The research objective is stated in section 1.4, which provides the motivation for the research questions posed in section 1.5. The research approach to answer these questions is introduced in Section 1.6 and finally, section 1.7 describes the structure of the remainder of the report.

1.1 General Introduction

Maritime container shipping has a central role in today's global economy, since it accounts for a significant part of the world trade. The port of Rotterdam serves as a destination for numerous containerships each year, a fact that places it among the most important terminals in continental Europe. The majority of container terminals within the port are operated by the European Container Terminals (ECT) company. ECT is participating in a number of inland terminals, and through its subsidiary company, European Gateway Services (EGS) provides inland transportation services that connect Port of Rotterdam (PoR) hinterland with ECT maritime terminals. As part of its value-adding strategy, EGS has developed new transportation products based on the concept of synchromodality that will offer differentiated service levels for the intermodal market in the region (van Riesen, 2013). The concept of synchromodality refers to an intermodal transportation network with online planning, able to adapt in real-time to meet delivery requirements (Bakas & Crainic, 2007).With online planning, it is meant that the planned transportation schedule can be adapted during the process, in order to account for the case of unexpected changes.

However, one of the main problems that ECT is facing, is that it currently has no sufficient way of accurately predicting the estimated time of arrival of the containerships at the port of Rotterdam, for instance due to weather conditions. This results in huge uncertainty over the types and amounts of cargo that reach the terminals, which in turn hinders the fast and cost efficient distribution of the products to inland destinations by EGS, through trucks, trains or barges. Therefore, being able to accurately estimate the time of arrival of the different containerships at the port of Rotterdam would have a positive impact on the supply chain operations of EGS.

Despite contractual obligations to notify the Estimated Time of Arrival (ETA) 24 hours before arrival, ship operators often have to revise it due to unexpected events like weather conditions, delay in a previous port and so on (Fancello, Pani, Pisano, Serra, Zuddas, & Fadda, 2011). A container vessel's delay, entails delayed containers unloaded in the terminal, and thus leads to delayed transshipment of containers to the hinterland side. For planners the decision-making processes related to this topic can sometimes be very complex without the support of suitable methodological tools.

The aim of this project is to propose a method for accurately predicting the estimated time of arrival of containerships, given weather predictions, as well as assessing the value that such a decision support system can have for the port operators and the other stakeholders involved in container transport. The research was undertaken as an internship at TNO.

1.2 Practical Problem

More than 80 % of the volume of global merchandize is transported by sea and handled by ports worldwide (UNCTAD, 2013). Containers are responsible for the transportation of more than 70 % of the value of this seaborne trade, facilitated by a seamless transfer of goods within multiple modes of transportation (UNCTAD, 2007). These containers are handled at seaport terminals, something which distinguishes the latter as crucial interfaces between landside and seaside transportation, and between various modes of transport. Therefore, seaports have been identified as critical infrastructures, which are essential elements that affect the economic and social well-being of a country (Mokhtari, Ren, Roberts, & Wang, 2012). This in turn means that ensuring the smooth and efficient operation of deep sea ports is of vital importance for the fast and cost-efficient distribution of goods to their final inland destinations.

However, there are certain disturbances that hinder the planning activities of the sea ports, which also affect the transportation process for all the stakeholders involved in the supply chain of container transport. One major source for these disturbances is stemming from the uncertainty over ship arrivals at the port. A late arrival of a sea vessel at the terminal, results in delays in unloading the ship and assigning it to the modes of hinterland transportation, namely truck, barge or train. This may result in delayed delivery to the final inland destination, thus causing the dissatisfaction of shippers.

Furthermore, uncertainty over the time of ship arrivals results in huge uncertainty over the demand profile, the types and amounts of cargo that reach the port over a specified time period. For the Port of Rotterdam, where numerous ships have to be loaded and unloaded every day, this has important implications. Firstly, allocating berth places for the arriving vessels becomes extremely difficult to plan, as well as the number of working shifts that need to be assigned to serve the incoming vessels on a given day (Fancello, Pani, Pisano, Serra, Zuddas, & Fadda, 2011). Moreover, booking enough capacity for the hinterland transportation side, through EGS, is subject to uncertainty. Since the arrival of vessels deviate significantly from what is expected, EGS cannot optimally decide in advance, the required capacity for barge and rail. Booking more than necessary results in extra costs, while booking less capacity results in increased usage of trucks, which also drives the costs up.

Taking the above into consideration, it becomes evident that the matter of accurately forecasting the estimated time of arrival for sea vessels at a port, is of vital importance for the cost-efficient execution of port operations and the hinterland transportation of goods. Reliable estimation of ship arrival can facilitate more efficient allocation of resources (human, spatial

and mechanical) for the port. Also, through the information that can be obtained about the estimated time of arrivals of containerships, the hinterland transportation process can be updated on-line, achieving faster delivery times and at a lower cost.

The current research aims at incorporating big data, such as GPS signals, current speed, vessel heading and weather predictions from marine traffic providers, in order to accurately forecast the time of arrival of containerships at a port. The value proposition of such an information tool will also be assessed, in terms of the positive impact that it can have for the port operators and the other stakeholders involved in container transport.
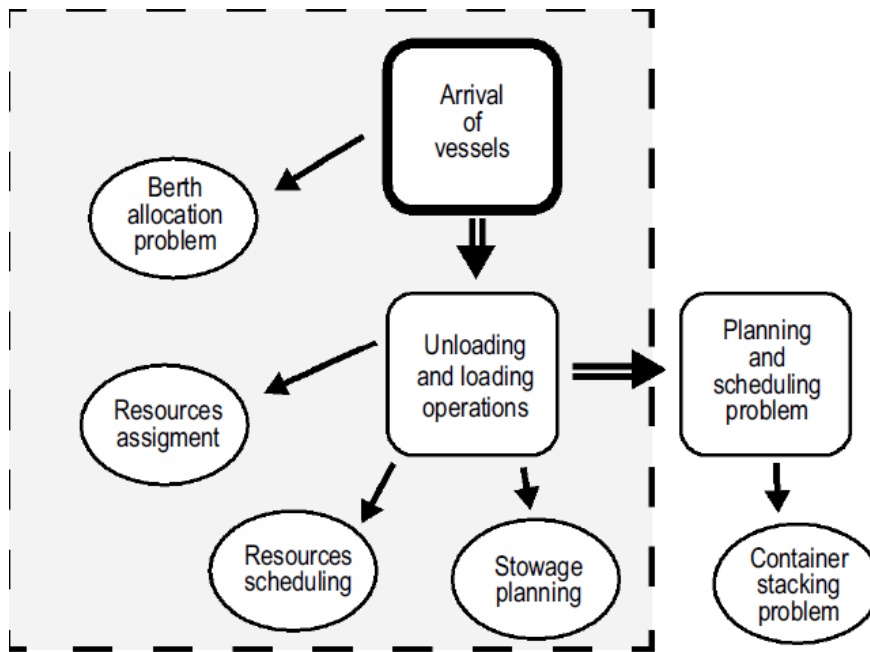
Figure 1: *Main Planning and scheduling problems at container terminals*

source: (Fancello, Pani, Pisano, Serra, Zuddas, & Fadda, 2011)

1.3 Scientific Problem

The problem of predicting the estimated time of arrivals for containerships at a port is related to operations research and is making use of forecasting techniques as its tool for achieving more effective process management for the container terminal operators. Through the information that can be obtained about the estimated time of arrivals of containerships, the hinterland transportation process can be updated on-line, achieving faster delivery times and at a lower cost. From a scientific point of view, the research aims at incorporating big data, such as GPS signals and weather predictions, in order to accurately forecast the time of arrival of containerships at a port, using machine learning techniques, such as neural networks and support vector machines. The value proposition of such a decision support system will also be assessed, in terms of the positive impact that can have for the port operators and the stakeholders involved in container transport.

The current methods developed on the subject are based on the work of Fancello et al (2011). There, a decision support system was proposed using a neural network that could decrease the uncertainty regarding time arrivals to 6 hours for the day ahead. However, the model was accounting for a short time horizon, since it was aiming at reducing the workload at the container terminal for the day ahead, and was not making use of weather data. The current research will aim at developing a model for predicting time of arrivals of ships over a longer time interval, up to 5 days from Port by incorporating weather predictions, in order to enable scheduling of the hinterland transportation activities.

The scope of the research will be to predict the estimated time of arrival of containerships at the port of Rotterdam, excluding the waiting time due to port operations. The relevant time window for the research will be a period of up to 5 days. The containership voyages that will be examined are those following the route from Asia to Port of Rotterdam. Therefore, predictions about the arrival time will be made after the ships pass Tunisia on a rolling time horizon, updated every few hours. To exclude waiting time due to port operations, Noordhinder (in Belgium) will be selected as a point of reference, which is positioned approximately 8 hours south of Rotterdam by ship. Therefore, limitations of the research is that it will not account for tankers or ships that are positioned more than 5 days away from the Port of Rotterdam, since the uncertainty over such a time interval is greater and difficult to estimate. Also, voyages that had a stop in other European ports before the Port of Rotterdam, such as Felixstowe and Antwerp are out of the scope of the current research, since they are also subject to their respective port operations time.

1.4 Research Objective

The research objective of the project is to make use of the data provided by marine traffic providers in order to accurately predict the estimated time of arrivals (ETA) of containerships at the Port of Rotterdam, in order to assist in the planning activities of the stakeholders involved in container transport in terms of improving cost and efficiency. For that reason, the value of such an information tool will be assessed for the stakeholders involved in container shipping. The research can serve as a starting point in demand profiling of the arriving cargo. If the problem of estimating ship arrivals at the Port of Rotterdam is resolved, information about the type of cargo can be incorporated in the future, in order to enable terminal operators to know in advance the expected arrival of the different types of cargo and their quantity. In that way, the latter will be able to schedule the inland distribution activities more efficiently, thus saving costs and time. The model that will be developed will span over a medium time horizon (up to 5 days from the port), accounting for the containerships known to have a matched fixture with the destination being the Port of Rotterdam.

The final deliverable will be a model that integrates the speed and position data of the vessels, as provided by marine traffic providers, with weather predictions, and accurately predicts the estimated time of arrival of containerships at the Port of Rotterdam. As part of the research, the value of such a decision support system for the port operators will be evaluated, in terms of improvement regarding their supply chain operations.

1.5 Research question and Sub-questions

Considering the problem presented above, the following main research question has been formulated:

How can the big data, provided by marine traffic providers, be used in order for the stakeholders involved in container transport to improve their business processes, by addressing the uncertainty regarding expected vessel arrival times at the Port of Rotterdam?

In order to answer the main research question, the following sub-questions need to be addressed:

1. *What is the added value that a more accurate prediction of the estimated time of arrivals for containerships at container terminals would have for the planning activities of the stakeholders involved in container transport?*
2. *Which are the main factors and to what extent are they affecting the average speed of the loaded containerships sailing towards container terminals?*
3. *How can a model be developed in order to accurately predict the estimated times of arrival for the containerships that have the port of Rotterdam as their destination, accounting for a medium-range time horizon?*

Through combining the answers of the sub-questions, the main question can sufficiently be addressed. The first sub-question has the purpose of determining the added value that the decision support system developed for estimating the time of arrivals for containerships, has for the terminal operators and the other stakeholders involved in container transport. In other words, it will be investigated how such a decision support system can generate value for the port operators (e.g. ECT) and what would be the business model for the port to make returns on the value it creates for others.

The second sub-question aims at realizing the main factors that have an impact on the speed with which containerships sail towards the Port of Rotterdam. The analysis undertaken here will serve as input for the model development in the next sub-question. Relevant factors that affect the speed of containerships are technical ship characteristics, weather conditions, such as waves, wind speed and direction, as well as market related factors. For instance, when demand is not high ships can choose to travel at the lowest attainable speed (slow steaming) to reserve on fuel.

The last sub-question aims at developing a model that uses as input the data identified to be relevant towards determining the ship's speed from the previous sub-question, and predicts the time of arrival for the containerships sailing towards the port of Rotterdam. The final deliverable will be a model predicting the ETA of containerships arriving at the Port of Rotterdam as a function of time, accounting for a medium-range time horizon. The machine learning techniques that will be used are neural networks and support vector machines, as it will be explained in chapter 3.

Then, having developed the model and assessed its value for the port operators, the main research question can sufficiently be addressed by proposing the way that the data available can be used for the development of the model.

1.6 Research Methodology

The research undertaken, consists of three different parts, as indicated by the sub-questions stated in section 1.5 . For each of the sub-questions the methodology used will be described below.

1.6.1 Investigating the value case of an ETA information tool

The first sub-question, regarding the value that such a decision support system has for the stakeholders involved in container transport, will be addressed through a qualitative approach. For this part, findings concerning the hindrances towards efficient planning from the side of terminal operators that occur due to uncertainties over the types and amounts of cargo reaching the Port of Rotterdam, can serve as a starting point for formulating a business model that can help ECT achieve more efficient planning for the terminal operations and its hinterland transportation activities. Further insights into the value proposition of the decision support system was gained through three scheduled interviews with representatives from the relevant stakeholders that have an interest in the project.

1.6.2   Literature review for factors affecting vessel speed

The main factors that affect the speed with which the containerships sail towards the Port of Rotterdam can be obtained through a literature review, based on the models that have been developed so far that have attempted to model ship velocity. The findings of this sub-question will serve as input for the model development at the next stage. The purpose of this step is twofold: on one hand to ensure that the most important parameters affecting the speed with which the containerships move towards the port of Rotterdam are taken into account, and on the other hand to understand exactly which information from the big data is relevant for the model development stage.

1.6.3   Machine Learning techniques towards ETA model development

Developing a data analytics approach for accurately predicting the ETA of containerships at the Port of Rotterdam is the backbone of the proposed research. The first step is to collect the relevant data regarding the past voyages of containerships. This information can be obtained from the AIS (Automatic Identification System) data. The AIS data contain information regarding the past voyages of the different types of ships (De Boer, 2010). At frequent time observations the GPS signal of the ships, latitude and longitude, are recorded, as well as the speed, draught and direction towards where the vessel is sailing (Loptien & Axell, 2014). The information obtained through the AIS data can be combined with data regarding weather predictions for the relevant time period. These data will be obtained from the company Hermess, which provides metocean data. It should be mentioned here that Hermess is a partner of TNO towards the realization of the project and therefore, its data are accessible.

After the data mining process, data cleansing and pre-processing is needed to formulate the input vector, which will be used as a predictor for estimating the time of arrival of containerships. The input vector will be using the following elements:

From the **AIS data**:
1) Average speed per voyage – obtained from the average speed in the voyage
2) previous, current speed in voyage
3) Distance to be covered
4) Direction of Vessel
5) Captain's ETA to port
6) Draught of the ship
7) Technical characteristics (length, breadth)

From **Hermess data**:
8) Sea-state, waves at the surface, wave direction
9) Wind, direction of wind
10) Currents' magnitude and direction

Based on the above inputs, the estimated time of arrival will be predicted. The methods that will be used are neural networks and support vector machines, as they will described in the literature review (chapter 3). Both of these methods are based on learning from previous examples on historical data and then generalizing to predict future events. The historical data available contain the voyages of 2015 and the first two months for 2016. Out of those historical data, 60% were used for training the neural network and the support vector machine, another 20% were used for validating the models and the last 20% for testing them. During the training phase examples from past voyages are used for "teaching" the aforementioned algorithms how to recognize patterns in the data and make predictions. Then, the validation set is used for selecting the optimal set of parameters for the neural network or the support vector machine, by selecting the parameters that minimize the error on the validation set. The real error that the algorithms would give in practice is determined in the testing phase, where the algorithms are fed with the inputs regarding position and weather data, and make predictions about the estimated time of arrival. These are examples that the algorithms have not "seen" before during the training or validation phase. This way, the aforementioned algorithms will be able to operate and process the data presented by the input vector to recognize patterns and predict the time that the ships will reach the port. The different methods will be evaluated based on error analysis, according to the predictions they produce and the actual estimated times of arrival of the ships, as given in the historical data. The software package that will be used for the analysis of the data is Matlab. Matlab was used for data mining and data cleansing, due to its efficiency in manipulating big data in structured arrays. It also contains important built-in functions for the purposes of data analysis.

1.7 Structure of the Report

The report is structured as follows: chapter 1 introduced the problem of uncertainty of vessel arrivals at the Port of Rotterdam and the research objective of the project. Chapter 2 follows with an extensive analysis of the usefulness of such an information tool for the stakeholders involved in the container transport process. Then, Chapter 3 will provide an overview of the literature on the topic to understand current forecasting techniques, as well as to assess the impact that better ETA predictions can have for the Port operators. Chapter 4 continues by describing the methodology followed for predicting the Estimated time of arrival for sea vessels, followed by chapter 5 which presents the results obtained using the proposed neural network and support vector machine. The practical implications of the results, from the viewpoint of the different stakeholders are analyzed in chapter 6. The results are further analyzed and interpreted in the discussion section of chapter 7. Finally, the report is concluded by summarizing the main findings of the thesis and proposing areas for further research.

Chapter 2 – Added value of ETA information tool

In this chapter, the value of the ETA prediction tool will be investigated, from the perspective of the different stakeholders involved in the supply chain activities of container transport. For that purpose, a stakeholder analysis is introduced in the beginning of the chapter, followed by ways that the ETA information tool could be of benefit for their activities.

2.1 Stakeholders involved in container transport

There are several stakeholders involved in the process of container transport, from the moment a company needs to transport a container of goods from a starting location until its final delivery at the required destination. The relevant stakeholders that would have an interest in the proposed ETA information tool, addressing vessel arrivals at the Port of Rotterdam are the following:

- Carriers that carry out the sea-going part of container transport
- Terminal operators (e.g. ECT)
- Hinterland transportation parties (e.g. EGS), including barge, rail and truck operators
- The Port of Rotterdam
- Importers that are the receivers of transported goods

For the purpose of understanding the role of the aforementioned parties in the supply chain of container transport, as well as their interests, a stakeholder analysis is undertaken in the following section. This analysis will assist in providing insight on how the proposed ETA information tool can be of value for the relevant stakeholders as it will be discussed in section 2.3.

2.2 Stakeholder Analysis

2.2.1 Carriers

The carriers are chartered by shippers for the transportation of goods between an origin and a destination deep-sea terminal. For carrying out this task, they operate a fleet of sea-going vessels that are suited for transporting specific types of cargo (e.g. containers, tankers, chemicals). For the execution of the sea transport, carriers receive a fare from the shippers. The fare is calculated based on the following formula:

$$Fare = Flat\ Rate * Cargo * Worldscale \qquad (1)$$

,where: *Flat Rate* is a constant that is reflecting the transportation price based on the distance between the two ports.

Cargo is the amount of cargo transported in tons

Worldscale is a negotiated percentage that is agreed between the shipper and the carrier for the transportation of the goods. If it is above 100% the carrier receives more revenue than the basic index indicated by the flat rate. If it is less than 100%, the carrier has agreed to transport the goods on a discount compared to the basic index (Stopford, 2009).

The shipping industry is very competitive, with many participating companies and none of them having significant concentration of the market. Therefore, the sea shipping industry almost follows the rules of perfect competition, where prices (as reflected in the negotiated worldscale with the shippers) depend heavily on supply and demand (Stopford, 2009). When the ships are relatively few for the transportation of cargo, carriers can charge high prices and keep a high profit margin. However, as more ships are built to capitalize on the opportunity for high fares, the market becomes saturated, with an oversupply of ships but not enough goods to transport. Then the bargaining power of carriers becomes significantly reduced and are forced to negotiate on much lower fares. During this downturn of the shipping industry, many ships are scraped until an equilibrium between supply and demand is reached again and the cycle repeats itself.

During the period of the downturn of the shipping cycle, reducing costs is becoming of significant importance for the carriers in order to keep operating profitably and avoid bankruptcy. In such cases, operating the vessels at the lowest attainable speeds to save on fuel consumption is often attempted, since fuel consumption accounts for the majority of a ship's expenses (around 60%) (Stopford, 2009). At the time this thesis is being written, container shipping is experiencing a significant downturn, since supply of ships is much higher than the demand, thus resulting in low transportation fares and slow vessel speeds in an attempt to save on fuel costs (Wright, 2016).

Based on the above, it can be deduced that the interest of carriers lies in transporting the agreed amount of goods reliably and at the agreed time-frame to the destination, while keeping operational costs (fuel) at the minimum.

2.2.2 Container Terminals

For the purpose of the thesis, the ETA information tool was applied for the case of containerships. Therefore, due to their importance for the current research, the container terminals will be analyzed separately from the Port of Rotterdam.

Container terminals can be decomposed in three parts in accordance to their layout, namely seaside operations, storage and landside operations. On the seaside or quayside of the terminal, containers are either loaded onto or unloaded from sea vessels. Then, the containers are stored in stacks in the storage area, which is called the yard. The storage area is therefore facilitating the decoupling of seaside and landside operations (Voss, Stahlbock, & Steenken, 2004). On the landside, containers are loaded onto or unloaded from barges, trains and trucks.

Containers can belong to three distinct categories, namely import, export or transshipment containers. Import containers are brought in by deep-sea vessels, stored in the terminal briefly, and need to be transported to the hinterland via barge, rail or truck. Export containers follow the opposite path. The research undertaken is mainly aimed at examining the imported containers and how this process can be benefited from an ETA information tool.

Once a vessel arrives at the container terminal, it is assigned to a berthing place, according to the berth planning of the terminal. There, the unloading of the ship takes place via a number of quay cranes. There are special transport vehicles that move containers from the quayside to the yard and vice versa. These can be trucks, straddle carriers or automated guided vehicles (AGVs) in (semi) automated ports. After storage, containers are moved to hinterland transportation modes (barge, rail or truck) for final delivery to the inland destination.

The movement of large volumes of goods requires for complex planning processes from the terminal operators. For that reason, it becomes evident that operations' planning is a key component of container terminal management. Four major divisions can be distinguished in the container terminal planning operations:

1. Berth planning calls for deciding the mooring slot and time slot for the ships at the quay (sea side) where they can be served with a planned number of quay cranes.
2. yard planning allocates the storage spots in the yard for import, export and transshipment containers.
3. Vessel planning refers to planning the order of unloading and loading containers from and onto the ship, while ensuring the stability and safety of the vessel.
4. Resource allocation reserves the required manpower and equipment for carrying out the aforementioned planning operations.

The following figure depicts the various planning and operational aspects of a container terminal.
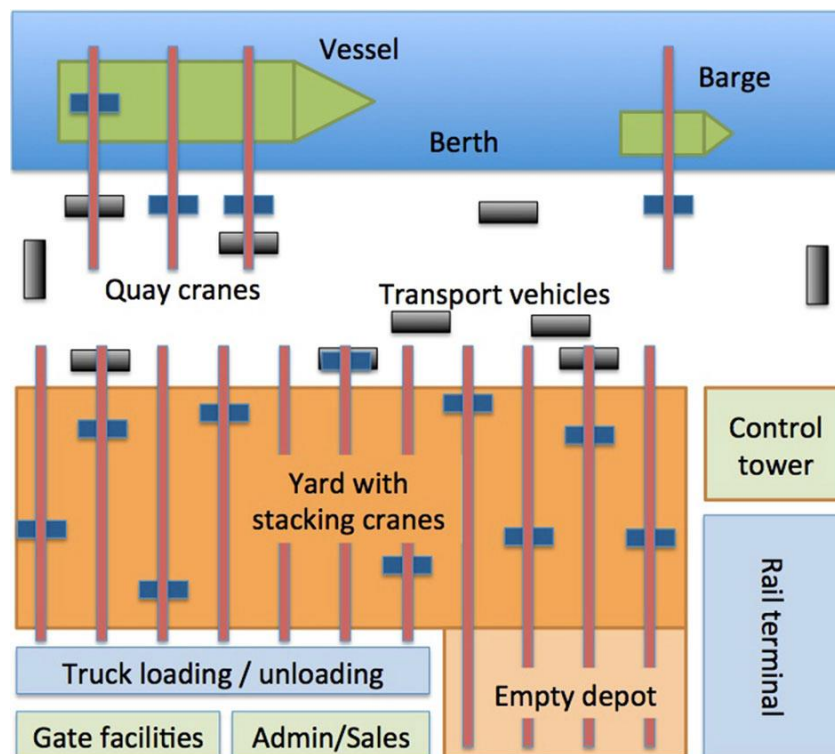
Figure 2: *Top-view of container terminal,* source: (Kemme, 2013)

2.2.3 Hinterland Transportation Parties (e.g. EGS)

The hinterland transportation parties are responsible for connecting the deep sea terminals of the Port to inland terminals. They receive the containers and allocate them to trucks, barges or trains for carrying out the hinterland transportation leg. Due to the increased competition between ports, providing cost-efficient and reliable services that connect the terminals of the port to the hinterland is becoming increasingly important. Their interests lie in ensuring cost-efficient , safe and reliable services for the inland supply chain activities. The role of hinterland transportation will further be elaborated in the following section (2.3)

2.2.4 Port of Rotterdam

Before analyzing the Port of Rotterdam, the following terms should be defined: 'port', 'port authority' and 'terminal'. A port is a geographical area where ships are brought alongside land to load and discharge cargo – usually a sheltered deep-water area such as a bay or river mouth (Stopford, 2009). The port authority is the organization responsible for providing the various maritime services required to bring ships alongside land. Ports may be public bodies, government organizations or private companies. One port authority may control several ports (e.g. Saudi Ports Authority). Finally, a terminal is a section of the port consisting of one or more berths devoted to a particular type of cargo handling (Stopford, 2009). Thus there are coal terminals, container terminals, etc. Terminals may be owned and operated by the port authority, or by a carrier that operates the terminal for its exclusive use.

Ports have several important functions which are crucial to the efficiency of the ships which trade between them. Their main purpose is to provide a secure location where ships can berth. However, this is just the starting point. Improved cargo handling requires investment in shore-based facilities. If bigger ships are to be used, ports must be built with deep water in the approach channels and at the berths. Of equal importance is cargo handling, one of the key elements in system design. A versatile port must be able to handle different cargoes – bulk, containers, wheeled vehicles, general cargo and passengers all require different facilities. There is also the matter of providing storage facilities for inbound and outbound cargoes. Finally, land transport systems must be efficiently integrated into the port operations. Railways, roads and inland waterways converge on ports, and these transport links must be managed efficiently.

The Port of Rotterdam has established itself as the leading port in Europe, in terms of the amount of cargo handling. However, ports are also involved in a high competitive industry for attracting an increasing number of vessels. Furthermore, new concepts such as the physical internet (Hakimi, Montreuil, Sarraj, Ballot, & Pan, 2012) may change the landscape of power between ports, and significant ports may be out of the picture in the near future, without adjustments. Therefore, it lies within the interests of the Port of Rotterdam to provide high-level services to the vessels, such as low waiting and handling time, and value adding hinterland transportation services connecting its terminals to inland destinations.

2.2.5 Importers

The importer is the buyer and receiver of the cargo inside the container, after the execution of the hinterland transportation leg. Importers are interested in the on-time and cost-efficient delivery of the goods at the specified location, convenient for their purposes. The focus of the present thesis is on the import transport chain, since the volume of incoming containers is significantly bigger than that of the export transport chain. The role of the importers and the value of an ETA information tool for their purposes will further be elaborated in the following section (2.3).

2.3 Value proposition of ETA information tool

In this section the value that an ETA information tool can have for the different stakeholders will be assessed, taking into account their specific needs. The ways that they can change their business operating activities are also investigated.

2.3.1 Methodology followed for assessing the value of ETA predictions

To determine the value of improved ETA predictions for the stakeholders involved in container transport, their business operations and interests were identified and analyzed through a literature review. The stakeholder analysis was mainly based on the work of Stopford (2009), for the ocean carriers and the port of Rotterdam, and to Voss et. al (2004) for the container terminals. Then, having formed an initial insight, the added value of the ETA predictions for the different stakeholders, as well as how this would affect their operating activities, were further investigated through interviews that were conducted with relevant stakeholders from the container terminal side, hinterland transportation parties and the Port of Rotterdam (Appendix D).

The interviews were conducted in a semi-structured manner, starting from the initial question regarding the current situation about expected vessel arrivals and how their inaccuracy affects the business process of the stakeholder under examination. Depending on the answer, other questions were asked, with the aim of identifying how improved ETA predictions would assist in improving the existing situation in terms of planning and cost reduction. Through this approach, the conversation was conducted in a more interactive way, without limiting the interviewee to giving only specific answers, but letting them express their thoughts on the topic more broadly. For instance, this approach led to the valuable finding that the bulk of the problem when it comes to the planning activities of terminal operators, is stemming from the uncertainty regarding the expected vessel arrival times when there is a previous stop in a European port, prior to reaching the Port of Rotterdam. At the end-stage of the interviews, a question was posed regarding other parties that would be interested in such an ETA information tool, in order to ensure that all the necessary stakeholders are taken into consideration, without omitting an important actor from the analysis.

Having conducted the interviews and the literature review, a synthesis of the findings was performed, which was later verified through a meeting with representatives from TNO, Intertransis and Hermess, the partners involved into developing this ETA information tool.

The results of this synthesis are presented in the following sections. At the end of the chapter, the role of TNO, Hermess and Intertransis, alongside with the usage of the proposed ETA tool will be presented.

2.3.2 Carriers

As it was mentioned in the previous section, carriers place particular importance in delivering the requested cargo on time to its destination. Therefore, they have a deadline for their arrival, after which the company would have to pay a penalty to the shipper for late delivery. On the other hand, carriers try to sail at low speeds when possible, to reduce fuel consumption which comprises a significant part of the vessel's voyage cost. These contradictory interests, result in a common pattern of speeding up in the initial part of the voyage, and once delivery at the port before the deadline seems guaranteed, the vessels significantly slow down to save on fuel, something which will more closely be examined in the data analytics part (chapter 5). Based on the above observation, three areas of how the ETA information tool could prove of value for the carriers have been identified:

1. Firstly, an ETA information tool can give the carriers a better indication of whether their vessel is on track to achieving on-time delivery. This can serve as an indication of whether the vessel needs to speed up to meet the deadline, or if there is room for slowing down, thus saving on fuel.

2. A useful application that could add value to the carriers' activities would be a speed planning decision support tool. This decision support tool would receive as an input the engine power used currently in rounds per minute and, based on weather predictions for the weather conditions ahead, calculate the positive or negative impact that they have on the ship as a percentage. For instance adverse weather conditions could apply a -5% to the speed of the vessel for the route. This could mean that the captain will sail at a speed 5% lower than the one he has currently attained. Such a support tool for speed planning could help reducing the variances in speed, thus saving on fuel. The captain would have the ability to keep a more steady speed instead of speeding up and slowing down later on. This results in reductions of fuel consumption, since fuel consumption is approximately a function of the 3rd power of speed as indicated in the following formula (Stopford, 2009):

$$Fuel\ Consumption = A + B * speed^3 \qquad (2)$$

, where A, B are constants depending on the vessel.

Therefore in a hypothetical case where a ship sails for the first half of a voyage with 15 knots and for the second half with 10 knots, keeping an average speed of 12.5 knots along the whole route would result in reducing fuel consumption by 6,7% (see Appendix E for the calculation).

This was a major finding during the interviews with Ed van Dort, managing partner of Intertransis, where he mentioned that carriers that cooperate with Intertransis

mentioned that "if there is a guarantee that keeping a steady speed could ensure on time arrival at the port, it would be of great benefit for them in reducing costs". However, the realization of such a decision support tool would need as input the rounds per minute used by the ship's engine, something which was not available for the current research. The information obtained from AIS data that were used, provide the speed over ground, which is already determined from the power of the engine and the currents flowing in the region. Therefore, this speed planning decision support tool is recommended as an area for further research at the end of this paper, building on the ETA information tool.

3. A final application of the ETA information tool for carriers would be as a competition monitoring tool. As it has been presented above, carriers are involved in a very competitive industry, where the rules of perfect competition apply. Therefore, knowing how many of the ships are in the region around the Port of Rotterdam and heading towards it, combined with information regarding their arrival time can give a good indication of the competition for the cargo that is available for transport at the port. For instance if a company knows that its vessel is the first to arrive on a given day and moreover there are few vessels available in the region, it can negotiate a high price with the shipper for the transportation of goods. On the contrary, if there are many vessels arriving before the company's containership, the carrier will know that they would have to negotiate for a low price, since in that case the shippers have higher bargaining power. Such a tool can assist in negotiating for the best attainable price for transporting cargo (as reflected in the negotiated worldscale index).

2.3.3 Container Terminals

Container terminals are among the main benefactors of an ETA information tool. As it was highlighted in (Menger, 2016), having access to accurate ETA information is perceived of vital importance for container terminals in order to plan their operating activities.

Firstly, alleviating the problem of vessel arrival uncertainty can lead to significant improvements in the berthing allocation problem. As it was mentioned in an interview with an ECT logistics manager (Appendix D), the terminal is planning on a berthing schedule depending on expected vessel arrivals. If however, a ship arrives later than expected, there may be shortage of space to allocate it at the quay, thus adding significant waiting time to the process. This causes disruptions to the vessels that arrive in the next hours also, since the whole quay planning has to be adjusted. Also, early arrival of a vessel is to be avoided, since the containers that will have to be loaded on the vessel may not be ready for transport at the yard. This happens because whenever a vessel is expected to arrive, the containers that are to be loaded onto it, are stacked in an accessible place in the yard to ensure quick transport to the vessel. If the vessel arrives earlier though, the containers may not be already at the right place in the yard.

Furthermore, changes in the berthing allocation schedule, due to uncertainty of vessel arrivals, have a negative impact on the yard planning schedule. If a vessel diverges from the

original schedule, it is assigned to another berthing place. The cargo in the stock goods has to be removed first from the initial loading position and then, has to be moved to the new berthing place. This results in increased workload, unnecessary moves and longer waiting time for the vessel.

To combat the uncertainty caused by unforeseen vessel arrivals, the terminal usually assigns more resources than necessary to each working shift, in terms of human resources and equipment (number of cranes). This results in increased costs, since manpower costs are the biggest expense of container terminals (Fancello, Pani, Pisano, Serra, Zuddas, & Fadda, 2011). Therefore, an ETA information tool could resolve the problem of resources allocation at the container terminal, thus effectively reducing operational costs. It is characteristic that in the case of Fancello et al. (2011), the neural network that was developed for predicting late vessel arrivals with an horizon of 24 hours, enabled a decrease in the number of working shifts from 4 to 2 for the port of Cagliari.

During the interview that was conducted with the ECT logistics manager, it was mentioned that currently the terminal is using the information provided by the ship agents regarding the ETA of the vessels. The ship agents are representatives of the carrier at the port and they use information given by the captains to provide an ETA of arrival to the terminal. This is the ETA that the terminal is currently communicating with the hinterland transportation parties, barge, trains and trucks that will be responsible for transporting the containers to inland destinations.

For the purposes of the terminal planning activities, accurate knowledge of the ETA 2-3 days in advance is important. As it was highlighted in the interview, the ETA provided by ship agents in the case when Port of Rotterdam is the first port of call in Europe can be off by some hours, but this is not a major hindrance for the terminal operations. The bulk of the problem was identified to be in the cases that vessels stop in another port in Europe for unloading, such as Antwerp or Felixstowe, before going to the Port of Rotterdam. In these cases, vessel arrivals are subject to huge deviations, since the waiting time in the previous port cannot effectively be estimated. Estimating waiting time due to port operations is out of the scope of the current research, therefore, the cases where there was a preceding port of call in Europe, before arrival to the port of Rotterdam, were not examined and are proposed for further research.

Nevertheless, in this master thesis the case of ships stopping for bunkering at Gibraltar (Spain), where usually they can acquire fuel relatively cheap, was taken into account. More specifically, the neural network and support vector machine that were used, were recognizing from the latitude and longitude of the ship that it had entered the bunkering area at the port of Gibraltar, and adding a waiting time. This can serve as a starting point towards tackling the Port Operations waiting time problem.

As far as the voyages that follow the Asia-Rotterdam route directly are concerned, one of the findings of the interview was that a horizon of prediction for up to one week in advance would be useful. The hinterland transportation parties would be more interested in such

information, since they have to coordinate many activities for their schedule and have to plan even further in advance.

2.3.4 Hinterland Transportation Parties (EGS)

Accurate information regarding the ETA of sea vessels is of vital importance for the planning activities of hinterland transportation parties such as EGS. This is due to the fact that vessel arrival is the starting point for all the supply chain activities of hinterland container transport. As it was highlighted in Menger (2016), delays and estimated time of arrival of deep sea vessels scored very high in the information that barge and truck operators would like to have. Accurate information regarding vessel arrivals would enable hinterland transportation parties, such as EGS, to book the necessary capacity for rail and barges, avoiding over or underestimating the demand. Their current business model is based on two types of decisions:

- The first is booking a fixed capacity of rail and barges for a week ahead. If the arriving containers cannot be assigned to a train or barge, they are transported by truck. In the case of overcapacity, where many containers cannot be assigned to multimodal modes of transport (barge/train), the costs for hinterland transportation increase sharply due to the increase in usage of trucks. On the other hand, in cases of under-capacity, where too much capacity has been booked on train and barges that remains unutilized, costs increase due to the low utilization rate of the transport modes.
- The second type of decision concerns the allocation of an incoming container to a mode of transport for the hinterland transportation leg. For this case, apart from the schedule of barges and rail, information regarding the cargo within the container is important, since shippers have different preferences for the transportation mode according to the cargo (Wanders, 2014).

It therefore becomes evident, that an improved accuracy regarding the ETA of containerships, spanning over a horizon of 7 days, could lead to tackling the problem of how much capacity to pre-allocate for barge and rail. If this information is coupled with information regarding the types and amount of cargo that a ship is carrying in the future, this would also resolve the uncertainty for the second type of decision that hinterland transportation parties are faced with. Therefore, it becomes evident that an ETA information tool could reduce the transportation costs for the leg of inland transport.

Currently, the ETA of sea-vessels as communicated by the terminal to hinterland transportation parties is not good enough for the purpose of their planning and operating activities. This is because the hinterland transportation parties plan on a greater time horizon than the container terminal and the ship's agent ETA becomes relatively updated during the last 40 hours of the voyage, when information becomes crucial for the terminal. Therefore, there is room for significant improvement in ETA prediction over a longer time horizon in order to create value for the hinterland transportation parties. In an attempt to quantify the requirements of such an ETA information tool, it was communicated during the interviews

conducted, that an error of 5 hours on average when the vessel is approximately 4 days away from the Port of Rotterdam, would have a significant impact on improving hinterland transportation activities.

Finally, it should also be mentioned that more accurate information regarding vessel arrival would result in an increase in the utilization rate of intermodal transport, thus saving costs and reducing the environmental impact of inland transport. The split between truck, barge and rail is currently standing at 55-35-10 respectively (Port of Rotterdam, 2014). More accurate ETA predictions could enable the increase of multimodal transport, thus benefiting barge and rail operators significantly on the expense of trucks.

2.3.5 Port of Rotterdam

As it was mentioned in the previous section, Port of Rotterdam has established itself as the leading port in continental Europe, despite the high competition that is taking place among the different ports. In order however to consolidate its position and increase its lead within a highly competitive environment, continuous innovation is needed. Therefore, a differentiation strategy towards its customers (carriers) can prove beneficial in sustaining a competitive advantage. In that respect, the ETA information tool can play a crucial role for the Port. This is in alignment with the Port of Rotterdam strategy statement that explicitly highlights its willingness to invest in further improving the efficiency of maritime, inter-terminal and hinterland transport , as well as play an active role in the development of data and data applications in the logistics chain (Port of Rotterdam, 2014).

The proposed ETA information tool can have significant benefits in reducing the hinterland transportation costs. It will also have a moderate effect in reducing handling time at container terminals, as described above. A low handling time for sea vessels is crucial, since during this idle time no additional revenues are produced. Therefore, a reliable service that could ensure faster handling speeds would be valued by carriers, thus choosing more often the Port of Rotterdam for unloading their containers. Furthermore, the fact that hinterland transportation costs can be reduced through such an application will be perceived as beneficial for the shippers. Real-time on line planning of inland transportation can ensure on-time delivery while lowering the costs. This means that in cases of cut-throat competition in pricing with other Ports, hinterland transportation parties could reduce the fee charged for inland transport, while keeping profitable margins.

It therefore becomes evident that the ETA information tool would provide significant advantages for the Port of Rotterdam in terms of achieving a competitive advantage. Carriers and importers will be more eager to do business with the Port of Rotterdam, due to its differentiated services in efficient container handling at the terminals and the cost-efficient execution of the hinterland transportation activities based on synchromodal planning.

Apart from those indirect benefits, there are also direct benefits associated with the usage of such an information tool from Port Operators. On the direct benefits is the ability to plan better for pilot availability to guide the vessels to the terminals, as well as reduction of traffic congestion around the areas of the Port. The latter is because of the fact, that better estimation of arrivals can result in better planning from the truck operators who would therefore arrive closer to the actual arrival time, instead of waiting around the Port.

2.3.6 Importers

The importer is the buyer and receiver of the cargo inside the container. By taking into account the benefits that the ETA information tool has for the hinterland transportation parties, it becomes evident that it could assist in synchromodal planning, due to the information provided about vessel arrival uncertainty. A synchromodal network provides additional value for the importer in various ways. Firstly, synchromodal networks offer a higher diversity of possible services that allow customization of transport to importer needs in terms of the speed/tariff trade off. Secondly, the network will increase its flexibility to respond to changes in volume and demand by easy exchange between alternative solutions, something which results in savings of rescheduling costs. Thirdly, the robustness of the network will be higher, to maintain service quality under changing circumstances, thus providing lower vulnerability of supply chains. Eventually, these networks may allow importers to save costs, while benefiting from high quality services by different routes (Tavasszy & de Jong, 2013).

2.4 Real time ETA Predictor and beneficiaries

In this section it will be described how an ETA predictor based on the models developed can work in a real-time environment and who will be getting the predictions and communicating them to the other interested parties. For that reason firstly, the role of TNO, Intertransis and Hermess in developing the tool will be highlighted. These parties have agreed on a partnership for developing an ETA information tool.

Intertransis is a company that acts as an information broker, providing advice and intelligence to its customers in the logistics sector, with the aim of improving their business processes (Intertransis, 2013). Their interest in developing an ETA information tool lies in the fact that they can expand their portfolio, providing services to their customers involved in container transport. Their asset is that they have access to the AIS data, that enclose information about the position, speed and heading of the vessel, as well as some technical characteristics. These AIS data are crucial for making predictions regarding the ETA of the vessels.

Hermess is a company that provides expertise and data-related services to the marine and coastal environment to support operations, engineering, and management of natural resources (Hermess, 2013). The ETA information tool would thus be a welcome addition to their portfolio. Towards the realization of the project, they provide weather data, information about the currents and waves.

TNO is a research organization which aims at creating innovations that boost the competitive strength of the industry in a sustainable way (TNO, 2015). The role of TNO within this project is to develop the algorithm, that given the AIS and weather data, can predict the ETA of the sea-vessels until the Port of Rotterdam. The present thesis was undertaken within TNO.

Having explained the role for the three partners involved in developing the ETA tool for the containerships heading towards the Port of Rotterdam, the business interrelationships can now be analyzed. The following figure shows the flow of data from the source, until delivery to the customers, namely the stakeholders in container transport.
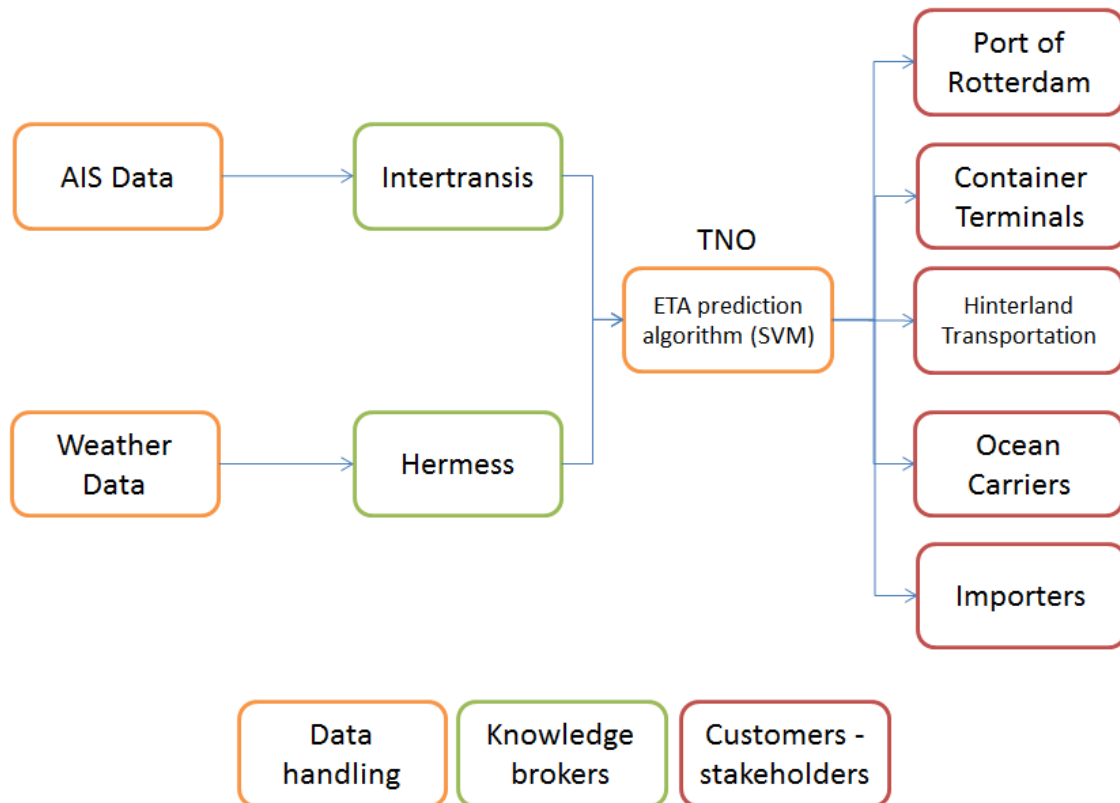
Figure 3: *Diagram presenting the flow of information for the ETA information tool*

The AIS data that are provided from marine traffic providers and the weather data, as collected from the models ECMWF and NOAA (see chapter 4.1 for details), are collected and aggregated from Intertransis and Hermess respectively. Then, these data can be fed forward to the models developed in the current thesis and produce ETA forecasts. The models developed can function by obtaining real time position and speed data, from AIS signals, and produce ETA forecasts based on their training parameters. When the models are on-line for making predictions on actual data, the training phase that is the most computationally expensive will have already taken place. Then, producing forecasts based on real-time new inputs is executed immediately. The weather data have not been found to be of crucial importance for estimating the ETA in the route examined, as explained in chapter 5. Therefore, acquiring these kind of data and the costs involved into doing that, is a step that can be omitted for the purpose of ETA prediction. However, weather data can be important

when estimating a vessel's ship fuel consumption, as suggested in the areas for further research (chapter 7.3).

The implication of this is that by acquiring real-time AIS data from marine traffic providers and feeding them forward to the trained model proposed in this study (SVM), sufficient predictions regarding the ETA of sea-vessels can be made. Then, through a common platform, this information can be communicated to the interested parties, such as the Port of Rotterdam, container terminals and hinterland transportation companies. The role of acquiring the data and using the predictive algorithms lies with Intertransis, since they are the information broker and they have access to the necessary AIS data. The role of TNO is to develop the algorithm for predicting the ETA of sea-vessels for a fee, and also to bring the relevant stakeholders together, making them realize the importance of ETA predictions for the fast and cost-efficient execution of container transport. Its purpose is also to ensure that during the process of the ETA model development, the product is designed according to the needs of the stakeholders.

Overview of Chapter 2

Along this chapter, the value of the ETA prediction tool was investigated, from the perspective of the different stakeholders involved in container transport. In section 2.1 the benefited stakeholders were briefly introduced, while section 2.2 explained their interests. Section 2.3 investigated the value that an ETA information tool would have for each one of them, based on the literature and the interviews conducted. That way, section 2.3 served at answering the first sub-question posed for the research, regarding the added value that an ETA information tool can have for the stakeholders involved in container transport. The benefits were analyzed from the view of the ocean carriers, terminals, the Port of Rotterdam hinterland transportation parties and importers. The chapter ended by explaining how this tool can be realized and which requirements have to be fulfilled.

Chapter 3 – Literature Review

In order to develop the required methods to answer the research question posed in section 1.5, a literature review is carried out. Section 3.1 provides a literature review on current methods used to forecast vessel arrivals at a port. The main variables affecting vessel arrivals will also be presented. Section 3.2 provides a brief overview of the different machine learning techniques and a comparison between them in order to select the ones most promising for the current research. The chapter ends with the literature gap identification that the research aims to bridge, by creating value for the stakeholders in container transport.

3.1 Consequences of ship delayed arrivals at port

By analyzing the scientific literature on the topic, it was identified that numerous problems of complicated nature co-exist at container terminals most of which need to be addressed by integrated solutions. Therefore, a solution to a problem at a container terminal often becomes of significant value for solving other related problems (Murty, Wan, & Linn, 2005).

An overview of the classification of decision problems at container terminals is provided by Vis and De Koster (2003), where five logistic processes are identified as areas of decision problems. Those logistic processes are: the arrival of the vessel, loading/unloading operations, moving the containers from quayside to yard and vice versa, stacking containers in the yard and transport of containers outside the terminal with other vehicles. As it can be deduced from the above, the arrival of the vessel is of vital importance, since the whole process starts from that point. Therefore, providing an analytical solution to the problem of the uncertainty of ship arrivals is essential for improving the availability and functionality of the handling system as a whole. It is characteristic that in his master thesis, Menger (2016) highlights, through a survey-based approach, that information regarding the ETA of containerships is perceived of high importance for the terminal and port operators, as well as hinterland transportation parties. The latter need this information in order to plan adequately for barge or rail capacity and schedules.

Furthermore, van Riesen (2013), states that early identification of disturbances in vessel arrivals is key to the synchromodal planning activities of hinterland transportation parties, such as EGS. This is because the whole process has to be updated depending on early/late arrival of containers, and vessel arrival is the first step for inland transportation. A late arrival of a vessel may result in unavailability of barges or trains to carry out the inland transportation, thus assigning more containers to trucks. However, this is increasing substantially the hinterland transportation costs.

Moreover, a major implication of late vessel arrivals at the port is that the process of assigning manpower and equipment becomes significantly more complicated. Arriving at a later time than the one expected increases the workload at the new arrival time, since extra human resources have to be allocated and equipment for the unloading of more vessels. This creates workload peaks, which result in peaks in energy consumption for the terminal,

something which has a large impact on the annual costs of the terminal (Heij, 2015). The reason for this happening, is that more cranes are put to use at the same time and also, extra movements for transporting the containers in the yard take place, in order to get them to the new berthing place of the delayed ship. If arrival of the vessel is predicted more accurately, then there will be less needs for changing the planning schedule from the side of the terminal operators, thus distributing more evenly the workload during the working shifts. That way, costs in manpower and electricity consumption can be reduced significantly. It is characteristic, that in the case of Fancello et al. (2011) the neural network model that was proposed for reducing uncertainty over arrival times at the Port of Cagliari, achieved a reduction of the working shifts from 4 to 2.

Also, changes in the schedule of the container terminal can result in the increase of demurrage costs. Demurrage costs occur when a vessel is available for discharging its cargo at the terminal, however its cargo takes longer to discharge than what is agreed. In container haulage, customers are given a fixed period in their contract to tip (unload) their container delivery. Acceptable times for tipping are usually between 3 and 4 days time spent on site and after that, it is considered "demurrage". Haulers will usually charge an hourly rate for each hour after the allowed time (Stopford, 2009). Due to late vessel arrivals, the planning activities of the terminal may be interrupted, causing shortage of equipment or manpower at peak hours to handle the vessels. If the terminal is unable to handle the unloading of the vessel over the specified tipping period, this will increase the demurrage costs that the terminal will have to pay to the carrier (Stopford, 2009).

Having addressed some of the consequences of late arrivals for the vessels at the Port of Rotterdam, the need for accurate ETA predictions becomes evident. The next section will present the current techniques that have been used in an attempt to address the problem and which are the most relevant for the case at hand.

3.2 Current methods on ETA prediction of sea-vessels

The state-of-the-art study revealed that despite the rapid technological innovation taking place in recent years, the uncertainty and variation in daily demand forecasting still remain a challenge for port operators. Furthermore, the specific applications are strongly limited, since most research undertaken thus far, concerns container flow prediction in and out of container terminals over only a daily time horizon.

Gambardella *et al.* (1996) proposed a forecasting module for estimating the daily container flow in and out of a terminal, combining two different estimators. The first predicts the number of containers to be loaded onto a ship due to arrive in port, based on past data. The second calculates the percentage of the total number of containers that should be transported by truck to the terminal, as a function of the ship's ETA. The only model capable of predicting ship arrival times has been calibrated by Fancello *et al.* (2011). The decision support system presented, reduces the uncertainty of port arrival time to approximately 6 hours by employing a neural network model. However, it only accounts for a 24-hour time

horizon, without including information for weather data. Therefore, it addresses the ship delays due to port operations. Through that model, the terminal in question is able to plan resources around just two work shifts, instead of 3 or 4, thus effectively applying lean management practices and reducing operating costs.

From the literature review conducted, it has also been identified that developing advanced vessel arrival time prediction tools for transshipment container terminals is closely related with the problem of managing and elaborating large amount of data. To enable the extraction of useful information therefore, it is required to refer to data mining techniques within the Knowledge Discovery in Databases (KDD) process, which is defined as the extraction of useful and not known information from data (Frawley, Piatetsky-Shapiro, & Matheus, 1992). Furthermore, specific attention should be given to data preparation, data cleansing and correct interpretation of results, in order to be able to extract information from the data, according to the KDD process.

There have also been attempts to model the weather effect on vessel speed using deterministic models, based on physics (Szelangiewicz, Wiśniewski, & Żelazny, 2014). In particular, the variables that are used for short term speed prediction are: the nominal speed of the vessel (based on the power of the ship's engine in rounds per minute), the impact of the wind on the sailing speed, as well as the impact from currents and waves. By taking into account the direction of the vessel, equations are built on which the actual speed of the vessel at that particular point is determined. However, their application is limited to short term speed prediction of the vessel, something which is not in alignment with the scope of this research. Furthermore, during the voyage the captain can change the speed of the vessel to arrive before the deadline for delivery at the port or to save on fuel, something which is not addressed in this research. The objective of the current thesis is, based on the speed and position of the vessel, and a weather pattern ahead, to determine the ETA of the vessel. This approach takes into account possible changes in the ship's engine power to counter weather conditions ahead. Nevertheless, the work of Szelangiewicz et al. 2014, serves as the base for addressing the first sub-question regarding the factors that affect the speed of a vessel sailing towards the Port of Rotterdam.

3.3 Machine learning techniques for prediction

At this point, it would be useful to provide an overview of the data analytics techniques that will serve as the basis for addressing the problem of predicting the estimated time of arrivals of containerships. The description given is based on the work of Freeman (2005) and Haykin (2008)**.**

3.3.1 Multilinear regression with gradient descent

Multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to the observed data. Every variable of the independent variable x is associated with a value of the dependent

variable y. A population model for a multiple regression model that relates a y-variable to n predictor variables is written as:

$$y^{(i)} = \theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \cdots + \theta_n x_n^{(i)} + \varepsilon_i \qquad (3)$$

The model is trying to fit the best possible line given the training examples x^i to predict the target value of y^i, with the difference being attributed to the residual ε_i , which is the random error. The superscript (i) refers to the i_{th} training example from the training set. The way to identify the best possible line fitting the observed data is by choosing the values of the θ coefficients, so as to minimize the sum of the squared errors for the sample (Freeman, 2005).

Two possible methods have been identified for this optimization problem, one is by using the normal equation method and the other is by using gradient descent. Due to the large amount of data that will be used for the case at hand, gradient descent has been chosen for the purpose of analysis. This is because the gradient descent converges faster to the optimal solution when there is large amount of data to be processed, while the normal equation method is computationally expensive.

Let us define the following hypothesis function:

$$h_\theta(x) = \sum_{j=0}^{n} \theta_j x_j \qquad (4)$$

For the parameter vector θ (of type R^{n+1}) the cost function to be minimized is:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 \qquad (5)$$

Gradient descent minimizes the objective function (4) by updating the elements of vector θ, until convergence, as follows:

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)} \qquad \text{,for j=0,1,...,n} \qquad (6)$$

,where α is the learning rate parameter.

3.3.2 Logistic regression with gradient descent

Logistic regression is an approach for solving classification problems, namely from a set of explanatory variables $x_1,\ldots,x_n$ to classify the dependent variable y in one of the various distinct classes (for the simple case there are two distinct classes y=0 and y=1). This approach is not suitable for estimating a continuous function, such as the ETA predictions and is only used as an introduction to understand the function of the neurons that are presented in the neural network approach. In this section, an overview of the logistic regression model will be presented.

Instead of a linear function, logistic regression is making use of the "sigmoid" function (also called logistic):

$$g(z) = \frac{1}{1+e^{-z}} \qquad (7)$$

The probability that a training example with independent variables x belongs to the class y=1 is given by the following probability estimation function:

$$Z_\theta(x) = \frac{1}{1+e^{-\sum_{i=0}^{n}\theta_i x_i}} \qquad (8)$$

Therefore, the probability that for a given input vector x, the training example belongs to class y=1 is denoted as:

$$Z_\theta(x) = P(y = 1|x; \theta) = 1 - P(y = 0|x; \theta) \qquad (9)$$

In order to get a discrete classification in one of the two classes y=0 or y=1, the output of the hypothesis function for the logistic regression can be determined as follows:

$$Z_\theta(x) \geq 0,5 \rightarrow y = 1 \qquad (10)$$

$$Z_\theta(x) < 0,5 \rightarrow y = 0 \qquad (11)$$

The way the logistic function g behaves is that when its input is greater than or equal to zero, its output is greater than or equal to 0.5:

$$g(z) \geq 0,5 \; when \; z \geq 0$$

Substituting z with $\sum_{i=0}^{n} \theta_i x_i$ to obtain the function $Z_\theta(x)$ the solution to (8) yields:

$$\sum_{i=0}^{n} \theta_i x_i \geq 0 \rightarrow y = 1 \qquad (12)$$

In order to choose the optimal parameters θ for logistic regression, the cost function that has to be minimized is the following:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)} log(Z_\theta(x^{(i)})) + (1 - y^{(i)})log(1 - h_\theta(x^{(i)}))] \qquad (13)$$

Gradient descent can be used again for finding the optimal parameters θ for the case of logistic regression by updating the values of vector θ, until convergence, as follows:

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) := \theta_j - \frac{a}{m}\sum_{i=1}^{m}(Z_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)} \qquad (14)$$

, with α being the learning parameter specifying the rate at which the vector θ is updated when new training examples are presented to the system.
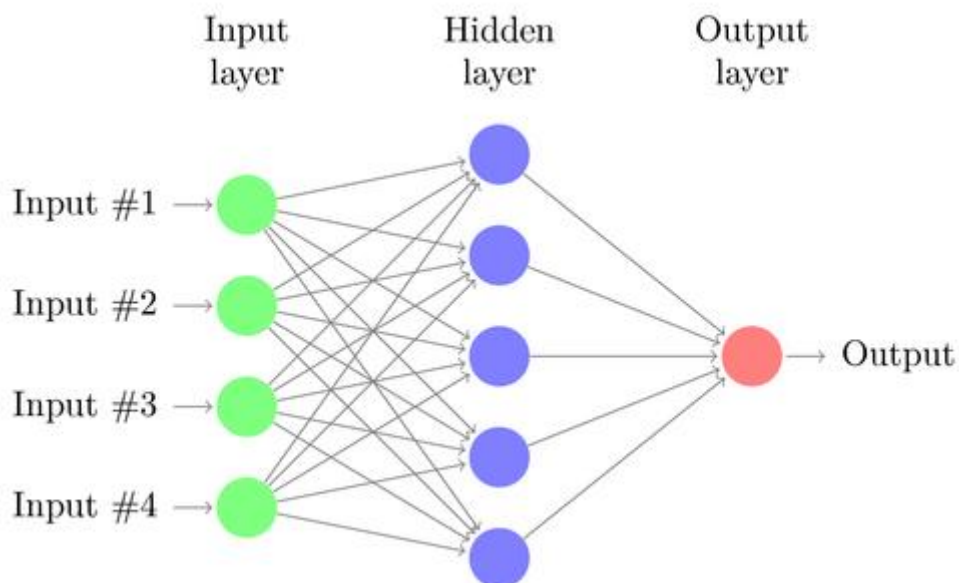
3.3.3 Neural Networks

Neural networks are limited imitations of how human brains work (Haykin, 2008). In an artificial neural network, a neuron is a logistic unit that receives inputs through its input wires, uses the logistic unit computation function (such as $Z_\theta(x)$ described above), and depending on the outcome of the computation sends an output (signal) to the output wires. The following scheme presents a single neuron:
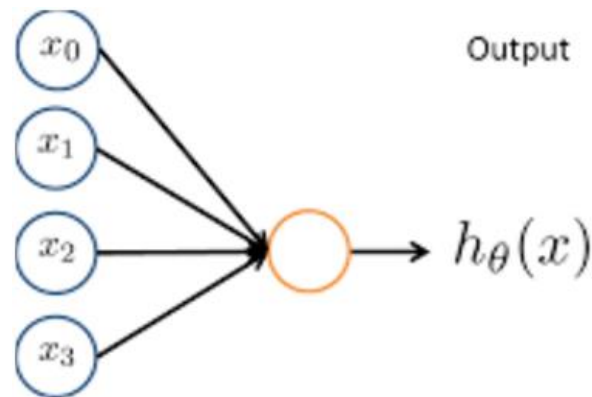
Figure 2*: Inputs and output of a single Neuron. Source: (Haykin, 2008)*

A neural network consists of multiple neurons, organized in layers and interconnected with each other via input wires characterized by their respective weights. Those weights are represented by the 'theta' (θ) parameters in the neural network model. Virtually, a simplistic representation looks like $[x_o, x_1, ..., x_n]^T \rightarrow [\ ] \rightarrow Z_\theta(x)$. The following figure shows such a neural network:

Figure 4: *Representation of a Neural network with 3 layers. Source:* (Haykin, 2008)

The above neural network consists of three distinct layers, namely the input layer, one hidden layer and an output layer consisting of one node. The network architecture is 4X5X1, meaning that the input layer receives 4 inputs which are then forwarded to the hidden layer. After being processed at the hidden layer nodes, their outputs are forwarded to the output layer which then makes a prediction, according to its activation function. This feedforward process is characteristic of the neural network and is used for making predictions, given an input vector. The training phase of a neural network comprises of selecting the optimal weights for each of the connections between the neurons. More specifically, given the input vector at the input layer, and the known output that actually occurred, the problem is defined as choosing the weights for the connections between neurons so as to minimize the error between the prediction and actual observation. An algorithm used for training the network until the optimal weights are selected, is the backpropagation algorithm (Haykin, 2008).

Let us label the intermediate or "hidden" layer nodes of the networks $a_{20} \dots a_{2n}$ and label them as "activation units." The following notation will be applied:

$a_i^{(j)}$: "activation" of unit *I* in layer *j*
$\Theta^{(j)}$: matrix of weights controlling function mapping from layer *j* to layer *j*+1

The formula that computes $a_i^{(j)}$ is the following:

$$a_i^{(j)} = g(z_i^{(j)}) \,,$$

where $z_k^{(j)} = \Theta_{k,0}^{(j-1)} \, \mathrm{x}_1 + \Theta_{k,1}^{(j-1)} \, \mathrm{x}_1 + \cdots + \Theta_{k,n}^{(j-1)} \, \mathrm{x}_n$, for j=2,…,n   (15)

What the set of equations (15) describe is that the inputs in the first layer of the network are forwarded to the second layer, after being multiplied with the weights "theta", and then each neuron calculates its activation function.

The prediction error of the neural network can be minimized with the backpropagation training algorithm, as it was mentioned before. For the training set $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$, forward propagation can be implemented to compute $\alpha^{(l)}$, for l=2,3,…,L , with L representing the total number of network layers.

Then, using the actual output for period t, the prediction error can be computed as:

$$\delta^{(L)} = \delta^{(L)} - y^{(t)}, \quad \text{for t=1,…,m} \qquad (16)$$

 So the "error values" for the last layer are the differences of the actual results in the last layer and the correct outputs in y.

To get the error values of the layers before the last layer, we can use an equation that steps us back from right to left:

$$\delta^{(l)} = ((\Theta^{(l)})^T \delta^{(l+1)}) \, * \, g'(z^{(l)}) = ((\Theta^{(l)})^T \delta^{(l+1)}) * (a^{(L)}) * (1 - a^{(L)}) \qquad (17)$$

The aim then is to minimize the error function $\delta^{(l)}$, by choosing the optimal parameters Θ. It is worth noting that due to their ability to compute complex functions, by organizing the neurons in multiple layers, neural networks have proven to be powerful predicting tools. However, one of their drawbacks is that they have a relatively slow training algorithm.

3.3.4 Support Vector Machines

Support vector machines are among the best "off-the-shelf" supervised learning algorithms. They are based on the concept of large margin intuition to divide the input vectors into classes, based on their similarity. They are characterized by usage of kernels, absence of local minima, sparseness of the solution and capacity control obtained by acting on the margin. They relied on defining the loss function that ignores errors, which are situated within the certain distance of the true value. This type of function is often called – epsilon intensive – loss function (Chapelle & Vapnik, 1999). The figure below shows an example of one-dimensional linear regression function with – epsilon intensive – band. The variables measure the cost of the errors on the training points. These are zero for all points that are inside the band.
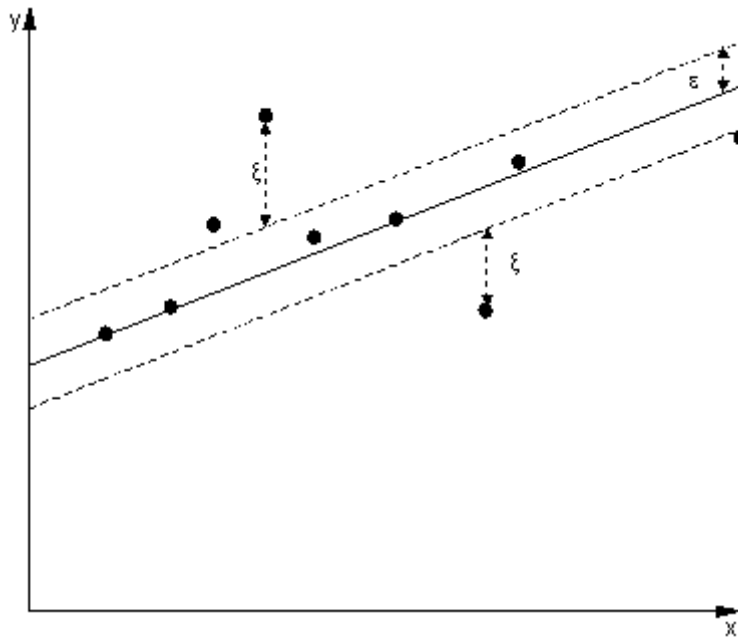
Figure 5: One-dimensional linear regression with epsilon intensive band.

In linear ε-insensitive support vector regression (SVR), training consists of solving the following constrained optimization problem:

$$min_{w,\xi} \frac{1}{2}||w||^2 + C \sum_{i=1}^{n}(\xi_i + \xi_i^*) \quad (18)$$

subject to constraints:

$$y_i - w * x_i - b \leq \varepsilon + \xi_i,$$

$$w * x_i + b - y_i \leq \varepsilon + \xi_i^*,$$

$$\text{and } \xi_i, \xi_i^* \geq 0 \qquad (19)$$

, where w is a weight vector, b is a bias value, (x_i, y_i) is a training sample and its target value, ξ_i and ξ_i^* are so-called "slack variables" enabling the model to allow deviations between the model output and the target value of training examples larger than ε, C is a parameter controlling the extent to which such deviations are allowed and n is the total number of training samples. Equation (18) is called the primal objective, and its variables primal variables. Introduction of Lagrange multipliers α and α^* and solving for the coordinates of a saddle point allow us to reformulate the primal objective and its constraints in the following way :

$$max_{\alpha^*} \left[-\varepsilon \sum_{i=1}^{n}(\alpha_i^* + \alpha_i) + \sum_{i=1}^{n}(\alpha_i^* - \alpha_i)y_i - \frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)(x_i \cdot x_j) \right] \text{ (20)}$$

subject to constraints:

$$0 \leq \alpha_i^{(*)} \leq C$$

$$\sum_{i=1}^{n}(\alpha_i^* - \alpha_i) = 0 \qquad (21)$$

Here, $\alpha_i^{(*)}$ is used to denote both α_i and α_i^*, and α^* to denote the vectors containing all α_i^* values. Equation (20) is called the dual objective. The second constraint in (21) is called the bias constraint. Once the α and α∗ maximizing the dual objective are found, a linear regression SVM determines its output using:

$$f(x) = \sum_{i=1}^{n}(\alpha_i^* - \alpha_i)(x_i \cdot x) + b \qquad (22)$$

The presented SVR model assumes that the relation between x_i and y_i is a linear one. It is also possible to make the SVR model nonlinear. This could be achieved by preprocessing the training patterns x_i by a map $\Psi : X \rightarrow F$ into some higher-dimensional feature space F and then applying the standard SVR algorithm (Schölkopf & Smola, 2002). However, such an approach can become computationally infeasible. Since both the dual objective (20) and the regression estimate (22) only depend on inner products between patterns x_i, it suffices to know $K(x_i, x) = \Psi(x_i) * \Psi(x)$, rather than Ψ explicitly. It is this kernel function K(·, ·) that is often used in SVR to make the algorithm nonlinear. A number of kernel functions are widely used, including polynomial functions, radial basis functions and sigmoidal functions (Wiering, et al., 2013).

For this application, the radial basis function was used as kernel, due to its enhanced ability of mapping feature vectors according to their similarity, as determined by their Euclidean distance. Other Kernels such as the linear were producing large errors, while the polynomial Kernels were computationally much more expensive.

$$K(x, x_i) = \exp\left(-\frac{||x-x_i||^2}{2\sigma^2}\right) \qquad (23)$$

, where $\left|\left|x - x_i\right|\right|^2$ is the Euclidean distance between the two feature vectors and σ is a free parameter.

The SVM generalization performance (estimation accuracy) depends on a good setting of meta-parameters parameters C, ε and the kernel parameters. The problem of optimal parameter selection is further complicated by the fact that SVM model complexity (and hence its generalization performance) depends on all three parameters. Existing software implementations of SVM regression usually treat SVM meta-parameters as user-defined inputs. Selecting a particular kernel type and kernel function parameters is usually based on application-domain knowledge and also should reflect distribution of input (**x**) values of the training data.

3.3.5 Comparison between the different techniques

In order to compare the different techniques presented, the following criteria will be taken into consideration: prediction accuracy, availability of data and computation time. Regarding the accuracy of predictions, multilinear regression can give very good results when the relationship between input variables and output is linear. However, with the weather conditions and the possible changes in speed along the route of a voyage, the relationship between input variables and output is not linear for the case at hand. On the other hand, neural networks can depict non-linear functions between input and output variables, due to their ability to use multiple neurons for their calculations. Their drawback, is that they require abundance of data availability in order to train and generalize sufficiently (Haykin, 2008). Also, their computational time is significantly increased compared to multilinear regression. For the case at hand, enough data was available for training a neural network (600 voyages with 4000 observations per voyage were used) and therefore this is the method that was implemented for the model-developing part. Support vector machines still need a lot of data to make generalizations, but the classification of data according to similarity enables them to generalize better than neural networks, given a limited training dataset (Wiering, et al., 2013). Also, their training time is significantly lower than neural networks. Their drawback is that they may not be able to perform well in really complex problems, such as image and video processing. There, given again abundance of data, neural networks can perform better by increasing the number of layers. However, for this particular case such limitation does not apply, as it will be shown in the results section (chapter 5).

Area of comparison	Multilinear Regression	Neural Networks	Support Vector Machines
Data Needed	Low	Very high	high
Training Time	Low	Very high	medium
Ability to depict complex (non-linear) functions	Low	Very high	high

Table 1: *Comparison between different machine learning techniques. Support vector machines and Neural Networks qualify for addressing the case of ETA prediction*

3.4 Literature gap and scientific contribution

The scientific contribution of the research undertaken here will be to address the knowledge gap regarding containership arrival uncertainty at the Port of Rotterdam. As it has been concluded in the literature review, there is a lot of room for improvement in the estimation of time and cargo arrivals of ships at container terminals, thus hindering planning activities for port operators and the stakeholders involved in the process of container transport. More specifically, the model that has been developed so far accounts for a 24-hour horizon for prediction of a ship arrival, without taking into account weather predictions (Fancello, Pani, Pisano, Serra, Zuddas, & Fadda, 2011). Therefore, it only accounts for the port-related activities regarding the delay of a ship's arrival. The scope of the current research will be to expand the time horizon to middle range predictions (5 days before arrival to port), by manipulating big data and creating value for container terminals, in order to improve their planning activities.

Overview of Chapter 3

This chapter presented the relevant literature for the research undertaken. In chapter 3.1 the previous work on predicting vessel arrivals at a port was presented, alongside with practical implications that such predictions can have. Through this approach, the second sub-question that was posed in the beginning, regarding the factors affecting the sailing speed of the vessels was addressed. It was found that the nominal speed of the vessel (based on the power of the ship's engine in rounds per minute), the impact of the wind on the sailing speed, as well as the impact from currents and waves are important variables affecting the speed of the vessel. However, changes in the engine power during a voyage, render it rather difficult to assess the ETA of the vessel through a deterministic approach based on physics. Then, chapter 3.2 presented the available machine learning techniques, from which neural networks and support vector machines were chosen as capable of addressing the problem of ETA prediction. The chapter ends with the literature gap identification and scientific contribution of the research.

Chapter 4 – Methodology of ETA Model Development

Having analyzed the benefits of an ETA information tool, this chapter proceeds to describe in detail the methodology followed for developing an ETA prediction model, based on neural networks and support vector machines. In the beginning the data that were selected for the purpose of the research will be presented, followed by the choices made in implementing the prediction models.

4.1 Description of data used

The data that were used for the purpose of the thesis were coming from two sources: the AIS (Automatic Identification System) data and data regarding numerical weather predictions. AIS is a mandatory system for ships above 300 gross tonnage that is sent between ships and between a ship and a shore-based station (TNO, 2015). The information sent from the AIS messages is updated frequently, every few minutes, and for every ship provides voyage related information including the ETA of the captain and the destination, as well as dynamic information regarding the vessel's speed, course and position. The dynamic information is obtained from the technical equipment of the ship (e.g. GPS) and is quite reliable. Voyage related information on the other hand, has to be input manually for every voyage, something that is not always done properly (TNO, 2015).

Numerical weather predictions and currents were acquired from the ECMWF (European Centre for Medium-Range Weather Forecasts) and NOAA (National Oceanic and Atmospheric Administration) models respectively. From the ECMWF model the weather magnitude and direction was used, whereas the currents and waves , both in magnitude and direction, were obtained from NOAA. These data were provided to TNO by the company Hermess, which is specialized in handling metocean and weather data. The weather and currents data were mapped, based on time and location, to the voyages under examination.

The data that were acquired for the present case were accounting for the voyages following the Asia – Rotterdam route, without previously unloading in other western European ports, spanning over 2015 and the first two months of 2016. In total, 600 voyages were used, with each one of them sending observations every few minutes. These include the total number of ships that followed the Asia - Rotterdam route directly during the specified time period. The total number of containership port calls in Rotterdam, including empty vessels or vessels for maintenance in 2015, was 7398 (Port of Rotterdam, 2016). Therefore, the voyages that were examined are representative of the population, in order to generalize. However, although there were a lot of observations per voyage, the total number of voyages is relatively low for applying machine learning techniques, since the latter require abundance of data. In order to cope with this fact, the size of the neural networks and support vector machines had to be kept to a minimum, by including a low number of variables for the purpose of estimating the ETA of the vessels.

4.2 Data pre-processing

From the data available that were briefly presented in the first chapter of the report, of significant importance was the pre-processing stage in order to construct the input variables that would be used for predicting the estimated time of arrival for the vessels at the Port of Rotterdam. Due to its significance, the data pre-processing step will be analyzed here.

Two sources of data were used, one coming from the AIS (Automatic Identification System) data and the other from data about weather conditions. The AIS data contain information regarding the past voyages of the different types of ships (De Boer, 2010). At frequent time observations the GPS signal of the ships, latitude and longitude, are recorded, as well as the speed, draught and direction towards where the vessel is sailing (Loptien & Axell, 2014). The weather data on the other hand were obtained from Hermess, a company that provides metocean data, and contained information about the wind, current and waves in magnitude and direction. These two data sources were combined in order to form the input vector, a set of explanatory variables, based on which predictions regarding the ETA of containerships were performed.

The following table presents a list of the input variables that were used for predicting the ETA of container vessels:

AIS Data	Weather Data
Latitude (degrees)	Current U-Component (m/s)
Longitude (degrees)	Current V-Component (m/s)
Distance to be covered (km)	Wind U component (m/s)
Current Speed of the vessel (km/h)	Wind V component (m/s)
Change in speed over the last 3 hours (km/h)	Peak wave period (s)
Average speed based on last 12 hours (km/h)	Peak wave direction (degrees).
Time used for calculating the average speed (hours)	Significant wave height (m)
Length of the ship (meters)	
Breadth of the ship (meters)	
ETA of the ship's agent (number of days)	

Table 2: *Data used as input variables for predicting the ETA of sea-vessels*

The choice of variables was based on a selection of variables identified to be relevant in estimating the ETA of containerships to a port. Most of these variables were identified as relevant in the literature review (chapter 2). Regarding the variable "Time used for calculating the average speed", this served as an indication of how much the neural network or the support vector machine can "trust" the value of the average speed, since in some cases the data used for calculating this average speed were much less than 12 hours. For instance, for the ships giving their first signal at Gibraltar, the time used for calculating their average

speed was very close to zero. In these cases, the neural network or the support vector machine would realize that less weight should be given to the value of the average speed that was presented.

It should be mentioned here that the speeds that were used, were the observed speeds over ground. This means that the observed speed was a function of the ship's engine power at the moment and the currents/weather conditions in the area. Therefore an interpretation of the weather conditions is already included in the speed information. The rest of the weather variables for the route ahead of the vessel until the Port of Rotterdam were used to serve as an indication regarding the captain's driving behavior along the route.

4.2.1 Challenges presented and solutions

The first challenge towards the ETA predictions of sea-vessels was the large volatility presented in the arrival times of the different voyages. The following figure shows the distribution of the arrival time of voyages in 2015, from Gibraltar to the Port of Rotterdam.
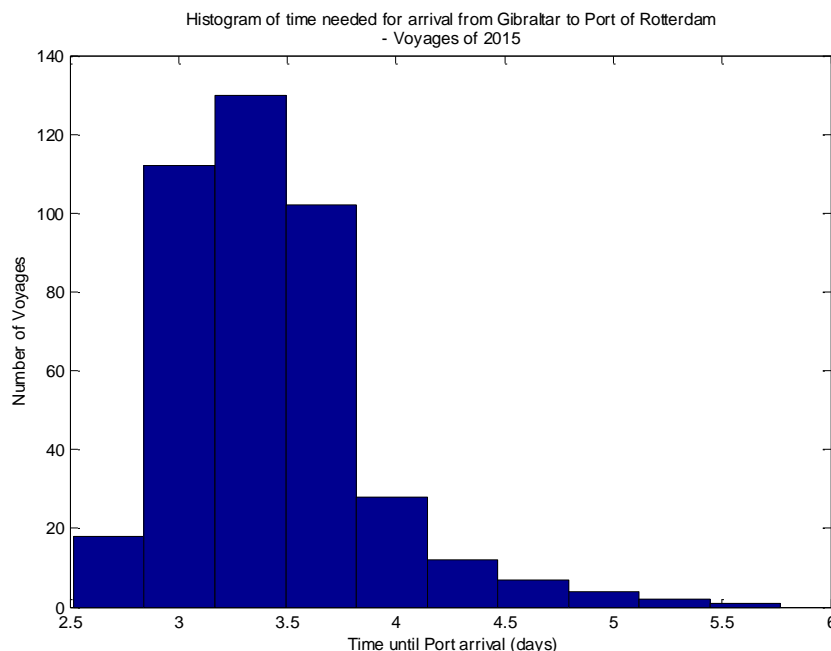
Figure 6: *Distribution of vessel arrival times from Gibraltar to the Port of Rotterdam for the historical data of 2015.*

As it can be noticed, there is significant variation in vessel arrival times for the different voyages, ranging from 2,5 days until 5,8 days. The majority of the voyages is positioned within a time-range of 3 to 4 days, however this is also a big interval in itself. Therefore, the ETA prediction methods that were developed had to include a set of variables that can capture the specific circumstances under which each voyage was taking place, in order to produce an accurate ETA prediction.

During the data pre-processing stage, in order to build the input vector that would serve as the base for making ETA predictions, several challenges were presented. The majority of those

were stemming from the fact that, although there were a lot of observations per voyage (every few minutes per voyage), there was a limited number of voyages available. The dataset was spanning over 2015 and the first two months of 2016, accounting for 600 voyages in total following the route Asia – Rotterdam.

Due to the aforementioned reason, the number of variables used for making ETA predictions had to be kept to a minimum, otherwise there was high risk of overfitting the data. Overfitting refers to the case when there is low error during the training phase of a neural network or a SVM, but high generalization error (Heaton, 2008). This can happen because during the training phase random patterns of the data are recognized and not the ones that actually have high importance for estimating the ETA.

In order to keep the size of the neural network and support vector machine relatively small, a careful interpretation of the weather variables for the route ahead of the vessel had to be determined. The following section (4.2.2) will present the way in which the weather was interpreted in such a way to account for the limitation of keeping a small size for support vector machines and neural networks.

4.2.2 Weather Interpretation – Clustering Approach

As mentioned above, in order to overcome the limitation of data availability, the following weather interpretation was undertaken: 4 weather checkpoints were defined along the route, positioned approximately 12-16 hours apart from each other and at key positions where significant changes of the speed of the vessels were observed. The weather variables were selected for each voyage in those regions, as an indication of what weather conditions are to be expected, with the aim of determining the captain's response and driving behavior. That way, 28 weather variables were selected (7 weather variables for each of the 4 regions). However, since they are still too many compared to the position data that are more important for estimating the ETA of the vessel, the data were clustered in 5 different weather clusters. Therefore, an indication of weather cluster 1 would mean favorable weather conditions for the journey ahead, while a weather cluster of 5 would mean adverse weather conditions. The following figure presents the areas that were chosen as weather checkpoints.
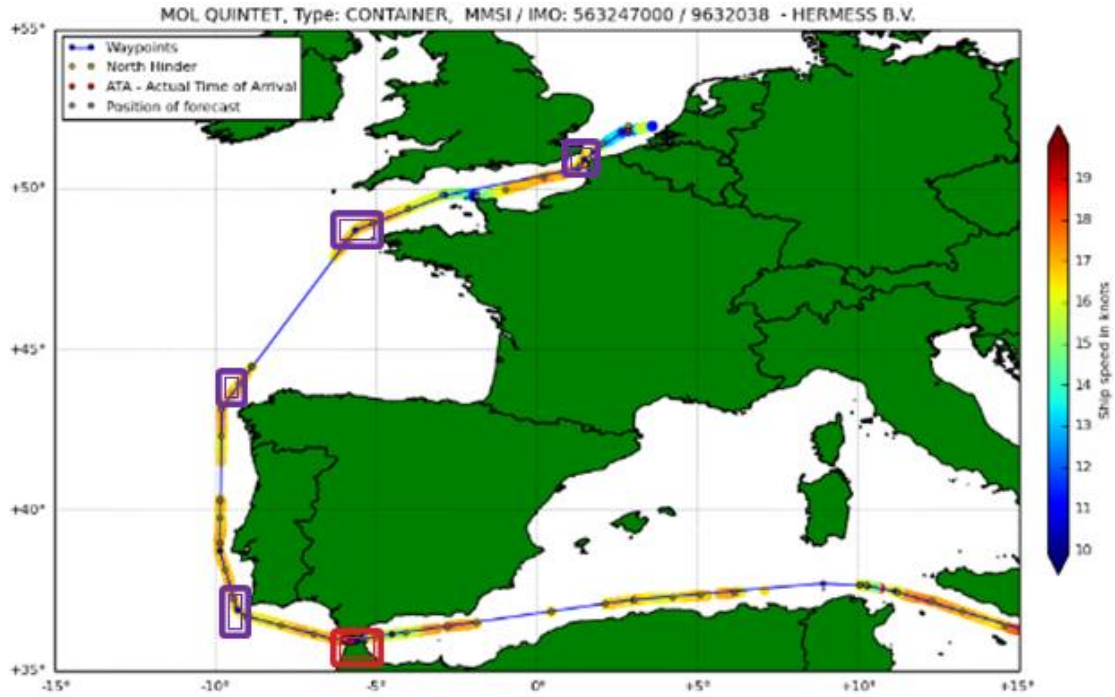
Figure 7: *Representation of a voyage in the examined route. With purple are the areas that were selected as weather checkpoints. (source: Hermess website)*

For clustering the weather variables into 5 distinct classes, the k-means algorithm was used. The k-means is an unsupervised learning algorithm for classifying data types in k different classes, so as similar inputs are classified in the same class. As a metric for the similarity between inputs, the Euclidean distance is used $||X - X_{centroid}||$. The algorithm is initialized with k vectors from the dataset serving as centroids. At each iteration, every input from the dataset is categorized in one of the k classes, based on which centroid has the lowest Euclidean distance from the presented example. Then the centroids are updated to be the average of all the examples that belong to the specific class. The process repeats until convergence of all the examples in their respective classes. The following pseudocode presents how the algorithm works:

<u>Input:</u>
K (number of clusters in the data)
Training set $\{x^1, x^2, x^3 ..., x^m\}$
Randomly initialize K cluster centroids as $\{\mu_1, \mu_2, \mu_3 ... \mu_K\}$
 Repeat {

       For i=1 to m

$$c^{(i)} = min_k ||x^{(i)} - \mu_k||^2 \qquad (24)$$

       For k=1 to K

              $\mu_k = average\ of\ points\ assigned\ to\ cluster\ k$

       }

,where K is the number of weather classes chosen (5 in the case at hand)

$\mu_1, \mu_2, \mu_3 \dots \mu_K$: the centroids of the clusters

m: the total number of examples (number of voyages)

$x^{(i)}$: the training example i (input vector for voyage i)

$c^{(i)}$: the cluster in which the training example i is assigned at each iteration

Following the approach described, the weather data selected from the 4 weather checkpoints along the vessel route were clustered based on similarity in 5 different weather classes. The number of classes (5 in this case) was chosen based on heuristics, since an automatic method for choosing the number of clusters does not exist and it usually depends on the data or the reason for carrying out the clustering (Coates & NG, 2012). A heuristic approach is based on computing the cost

$$J = \frac{1}{m}\sum_{i=1}^{m} ||x^{(i)} - \mu_{c^{(i)}}||^2 \qquad (25)$$

for different number of classes and observe after which number of classes its rate of decrease lowers significantly. The graph for this this case is presented in the appendix (Appendix A).

4.3 Neural Network Approach

In chapter three a brief overview of the mathematic representation of neural networks was presented alongside with the reasons that neural networks were chosen for the purpose of estimating the ETA of sea-vessels. In this section the network architecture selected for the case at hand will be presented, highlighting the choices regarding the input variables and network parameters. Results of the neural network approach for estimating the ETA of containerships are presented in section 5.

There are four key decisions that have to be made regarding the implementation of neural networks:

- The first has to do with the inputs to the neural network.
- The number of hidden layers of the neural network
- The number of neurons in the hidden layer(s)
- The number of outputs of the neural network

The output of the neural network is determined by the forecasting problem, which in this case is predicting the ETA of containerships. Therefore, the output layer of the neural network consists of a single node, which gives as output the ETA of the vessel. The goal is for the output to be as close as possible to the Actual Time of Arrival (ATA).

The inputs to the neural network are the AIS data for each voyage, as presented in Table 1, alongside with the weather variable after clustering. Therefore, there are 12 input variables for the neural network, based on which the predictions regarding the ETA are made.

Regarding the number of hidden layers, for the majority of the problems, one hidden layer is sufficient for achieving good results (Heaton, 2008). The situations in which performance improves with deeper neural networks (2nd, 3rd layer and so on) are very small. However, deeper neural networks can represent functions of any kind and shape, provided there is a sufficiently large dataset on which they can be trained (Heaton, 2008). For the specific kind of problem, the most crucial variables are clearly distinguishable, the set of speeds selected for the vessel, its distance to the destination and the Agent's ETA. Therefore, it can be classified as a problem where one hidden layer is sufficient for finding the relationship between the input and output variables. This, coupled with the fact that the total number of voyages on which the neural network can be trained are limited (total of 600 voyages), led to the choice of using one hidden layer for the case at hand.

Finally, the number of neurons in the hidden layer is crucial. Using too few neurons can result in underfitting, a case when there are too few neurons in the hidden layer for recognizing the patterns hidden in the dataset. On the other hand, using too many neurons for the problem at hand, may result in overfitting, a case when the neural network has more information capacity than that contained in the training set and therefore, the neurons are not trained sufficiently. This results in low training errors, but high errors when trying to generalize in cases that the neural network has not encountered before. There are few heuristics available for selecting the optimal number of neurons, stating that usually that number should be between the input and the output size of the network (Heaton, 2008), so between 12 and 1 for the case at hand. Different network sizes were simulated with the aim of selecting the optimal number of neurons. The one that achieved the lowest error in the validation set contained 7 neurons in the hidden layer. The simulations for selecting the optimal number of neurons are included in the appendix (Appendix B).

Based on the above set of choices, the following figure shows the neural network architecture that was used for tackling the problem of ETA predictions for sea vessels:

Neural Network for ETA prediction
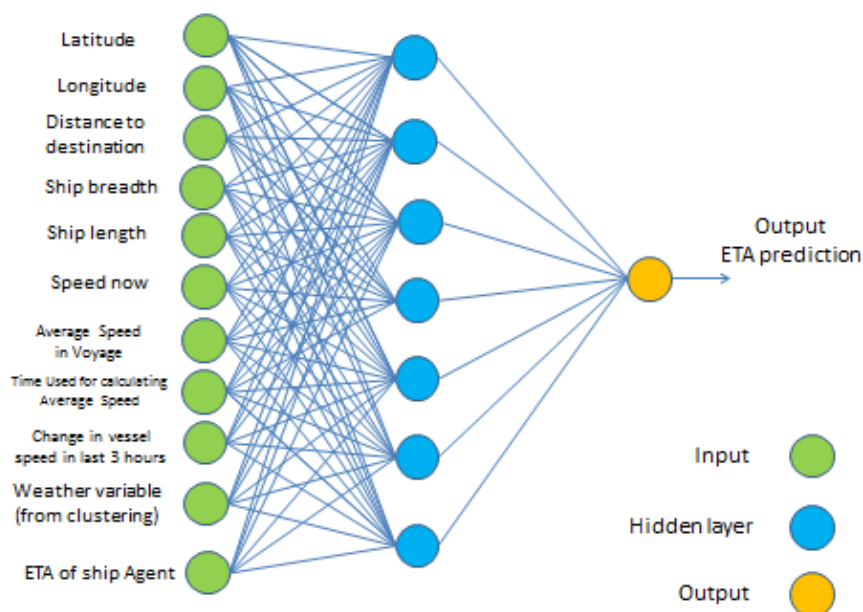
Figure 8:_Neural Network architecture for ETA prediction (created by author)_

The data that were used for training and validating the neural network span over 2015 and the first two months of 2016. In total, the dataset consists of 600 voyages, with each one having sent approximately 4000 observations (AIS data) along its route. Thirteen positions within frequent time intervals of 6-8 hours were chosen along the route of each voyage. From each of those positions, a prediction regarding the ETA of the containership was performed. Therefore, predictions regarding the ETA of sea-vessels were performed on a rolling time horizon along the route from west of Sicily until the Port of Rotterdam. In the beginning of the route relatively few ships were available in the dataset and more were being added when moving closer to the Port of Rotterdam. The total number of examples on which the network was trained and validated was 5380. There are three distinct datasets that are used when using a neural network for making predictions:

- The training set (65% of the total data)
- The validation set (15% of the total data)
- The testing set (20% of the total data)

The selection of how the voyages were distributed among the different sets was random, to avoid having a bias due to seasonality effects. For instance, during the winter, ships may attain in general lower speeds due to weather conditions. Assigning the voyages in random order to the training, validation and testing set, such biases were eliminated.

The Training phase

During the training phase, examples of the input vector are presented to the neural network with the aim of training it in order to make predictions. The aim of this supervised learning phase is to present the set of inputs (position and weather data) alongside with the respective outputs (actual times of arrival) in order to find the optimal weights connecting the neural network layers (Haykin, 2008). This happens through the process of back-propagation that was presented in chapter 2.

The validation phase

The validation set is used for tuning the network parameters. It is separate from the training set and is making use of an error metric, usually mean absolute error in numerical predictions, to determine which set of parameters is optimal (Haykin, 2008). For each set of parameters (number of neurons in hidden layer and regularization parameter) a neural network is trained on the training set and then validated on the validation set. There, the neural network or SVM is presented with new inputs it has not encountered before and predictions are made. The neural network or SVM that achieves minimization of the error metric (mean absolute error) is the one that is selected.

The testing phase

During the testing phase the real error of the neural network or the SVM is determined. A new set of inputs is presented to the model and predictions are made. The mean absolute error from the actual time of arrivals is the real error that the model has. The difference between the testing and the validation set is that the validation error is biased, since effort has been put into minimizing it by tuning the model parameters (Heaton, 2008). On the other hand, no optimization is carried out on the test set, which is only presented once to the neural network or SVM to determine the real accuracy of predictions. Therefore, the testing error is usually a bit higher than that of the validation set.

4.4 Support Vector Machines Approach

The second approach followed for making ETA predictions was through the usage of support vector machines. Their mathematical representation was mainly described in chapter 2. By using the radial basis function kernel the equations are as follows:

$$max_{\alpha^*} \left[-\varepsilon \sum_{i=1}^{n} (\alpha_i^* + \alpha_i) + \sum_{i=1}^{n} (\alpha_i^* - \alpha_i) y_i - \frac{1}{2} \sum_{i,j=1}^{n} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) K(x, x_i) \right] \quad (26)$$

subject to constraints:

$$0 \leq \alpha_i^{(*)} \leq C$$

$$\sum_{i=1}^{n} (\alpha_i^* - \alpha_i) = 0 \qquad (27)$$

Here, $\alpha_i^{(*)}$ is used to denote both α_i and α_i^*, and α^* to denote the vectors containing all α_i^* values. Once the α and $\alpha*$ maximizing the dual objective are found, a regression SVM determines its output using:

$$f(x) = \sum_{i=1}^n (\alpha_i^* - \alpha_i) K(x, x_i) \qquad (28)$$

Where $K(x, x_i)$ is the radial basis function kernel given by the formula:

$$K(x, x_i) = \exp(-\frac{||x-x_i||^2}{2\sigma^2}) \qquad (29)$$

For the specific case at hand x_i is the input vector for each voyage, containing the same inputs as the ones used for the neural network training. The actual time of arrival is y_i, and it is the value that the model is trying to forecast

Of significant importance when training SVMs is the parameter selection process. Parameter C determines the trade-off between the model complexity (flatness) and the degree to which deviations larger than ε are tolerated in optimization formulation for example, if C is too large (infinity), then the objective is to minimize the empirical risk only, without regard to model complexity part in the optimization formulation (Chapelle & Vapnik, 1999).

Parameter ε controls the width of the ε-insensitive zone, used to fit the training data. The value of ε can affect the number of support vectors used to construct the regression function. The bigger ε, the fewer support vectors are selected. On the other hand, bigger ε-values results in more 'flat' estimates. Hence, both C and ε-values affect model complexity, but in a different way (Chapelle & Vapnik, 1999).

The same testing, validation and training sets were used in the case of SVMs, as described in the Neural Network section. The parameter selection process takes place during the validation phase, where the parameters that minimize the error over the validation set are selected as optimal (Appendix A). The real-world error of the model is determined on the testing set.

4.5 Error Metrics used for evaluating ETA predictions

In this section the two error metrics that were used for comparing the performance of neural networks and support vector machines , both between them and to the error of the ship's agent ETA, will be presented.

The error metrics used for evaluating the performance of the predictions were the Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE). The reason for choosing these error metrics are that they can give an indication of the average error in hours (through MAE) and the variance of the prediction errors (through RMSE), therefore enabling a quick evaluation of the prediction results. The error metrics were expressed in hours, something which gives a more accurate picture than when compared to percentage errors (such as MAPE). That is because for the long time horizon (120 hours away from the port) prediction errors of 5 hours correspond to a 4% deviation, whereas for the short time horizon (20 hours away from the port) an error of 3 hours corresponds to a 15% percentage error. Therefore, if

percentage errors were chosen, they could lead to a misleading picture regarding the ETA predictions, due to the different time horizons examined.

4.5.1 Mean Absolute Error (MAE)

The mean absolute error (MAE) is an error metric, commonly used in statistics, to measure how close forecasts or predictions are to the eventual outcomes. As its name suggests, it is the average of all the average forecasting errors and is on the same scale as the data being measured. This implies, that for this case it indicates the number of hours that on average the predictions were wrong from the actual time of arrival of the vessels.

The mean absolute error is given by the following formula:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| = \frac{1}{n}\sum_{i=1}^{n}|ETA_i - ATA_i| \qquad (30)$$

, where f_i: the forecasted value (in this case the Estimated time of Arrival ETA)

y_i : the value that we want to predict (in this case the Actual Time of Arrival ATA)

n: The total number of observations (in this case the number of voyages)

4.5.2 Root Mean Squared Error (RMSE)

The root mean squared error (RMSE) is a very common error metric in statistics when it comes to evaluating numerical predictions. Compared to the similar Mean Absolute Error, RMSE amplifies and severely punishes large errors. It therefore serves as an indicator for the variance of the prediction errors. The formula used for calculating the RMSE is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - y_i)^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(ETA_i - ATA_i)^2} \quad (31)$$

, where f_i: the forecasted value (in this case the Estimated time of Arrival ETA)

y_i : the value that we want to predict (in this case the Actual Time of Arrival ATA)

n: The total number of observations (in this case the number of voyages)

The aforementioned error metrics were used as a guideline for comparing the different methods used for predicting vessel arrivals at the Port of Rotterdam. In general, lower errors signify a better performance of the technique used.

Overview of Chapter 4

In this section the methodology for tackling the problem of making ETA predictions was presented. The following diagram summarizes the steps undertaken:
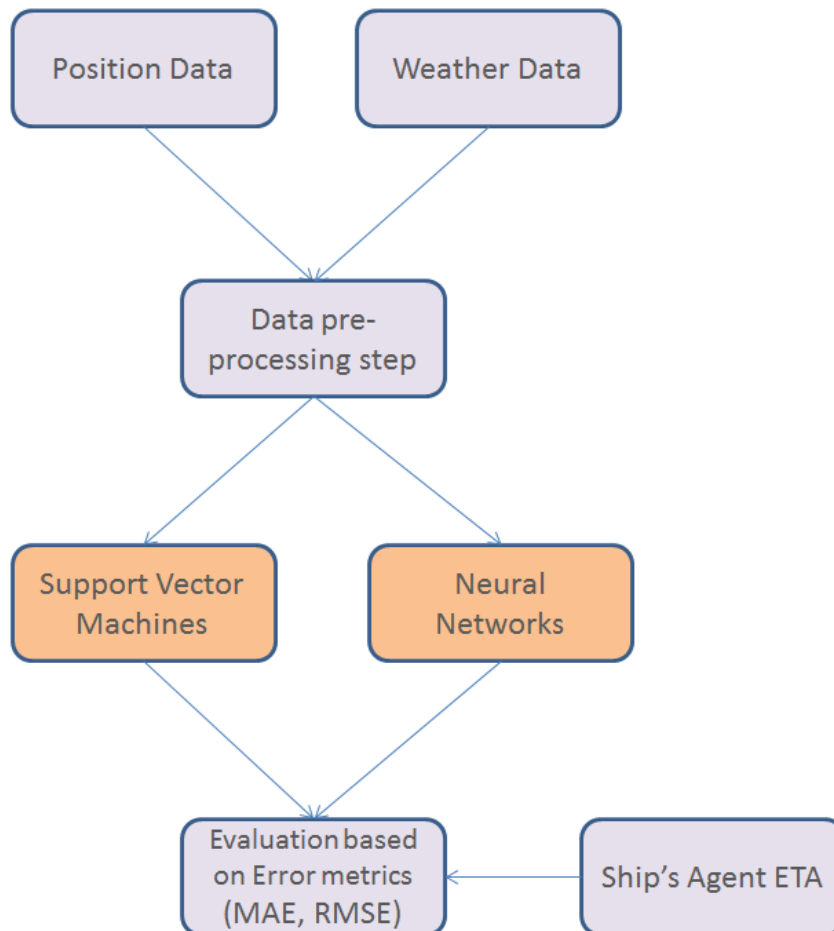
Figure 9: *Schematic overview of the methodology followed for predicting the ETA of containerships*

After collecting the positional and weather data of the voyages of the past year, a pre-processing step took place to select the relevant information and build the input vector, a set of variables relevant for predicting the ETA. Then, the pre-processed data were provided as input for the training of two different machine learning algorithms, the support vector machines and the neural networks. The results that were obtained during the testing phase were compared to those of the current situation (based on the ship's agent ETA) and between themselves, according to the error metrics defined in chapter 4.4, the mean absolute error and the root mean squared error.

Chapter 5 – Results of simulation

The previous chapter described the methodology followed for developing neural network models and support vector machines in order to predict the ETA of containerships at the Port of Rotterdam, based on historical data. This chapter will present the results of the two techniques, under different scenarios. The performance of the algorithms will be assessed on two error metrics, the mean absolute error of predictions and the root mean squared error.

5.1 ETA predictions using Neural Networks and SVMs

Having presented the methodology followed for making ETA predictions on a rolling time horizon, the respective results are presented. The following figures show the mean absolute error (MAE) and the root mean squared error (RMSE) in hours for the Ship's Agent predictions, the neural network and the support vector machine. The error is evaluated in frequent time intervals (every 6-8 hours) from Tunisia, which on average is positioned 120 hours away from the Port of Rotterdam until approximately 20 hours from the port. Another point of reference is Gibraltar which is positioned on average 80 hours from the Port of Rotterdam.
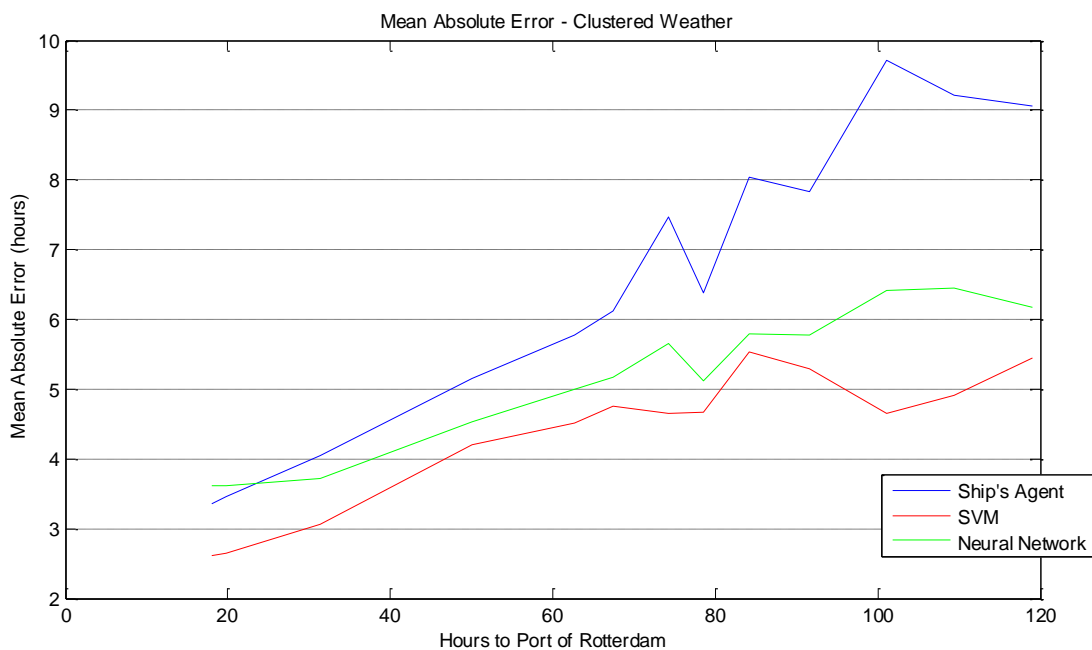
Figure 9: *Mean absolute error on ETA predictions for Ship Agent, SVMs and Neural Networks – clustered weather variable approach*
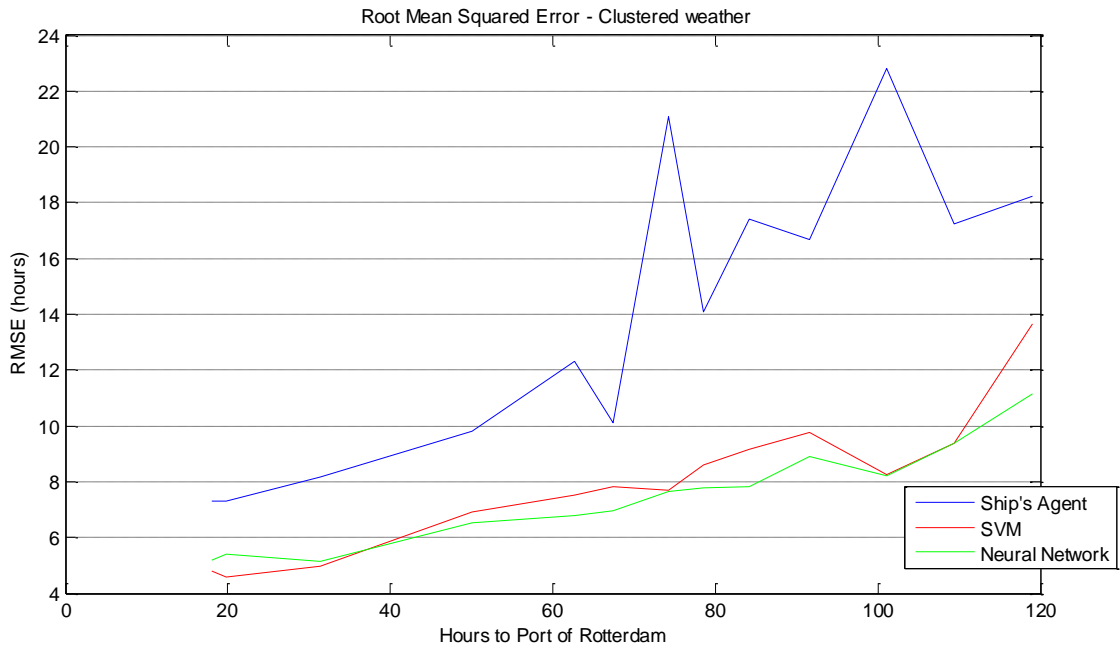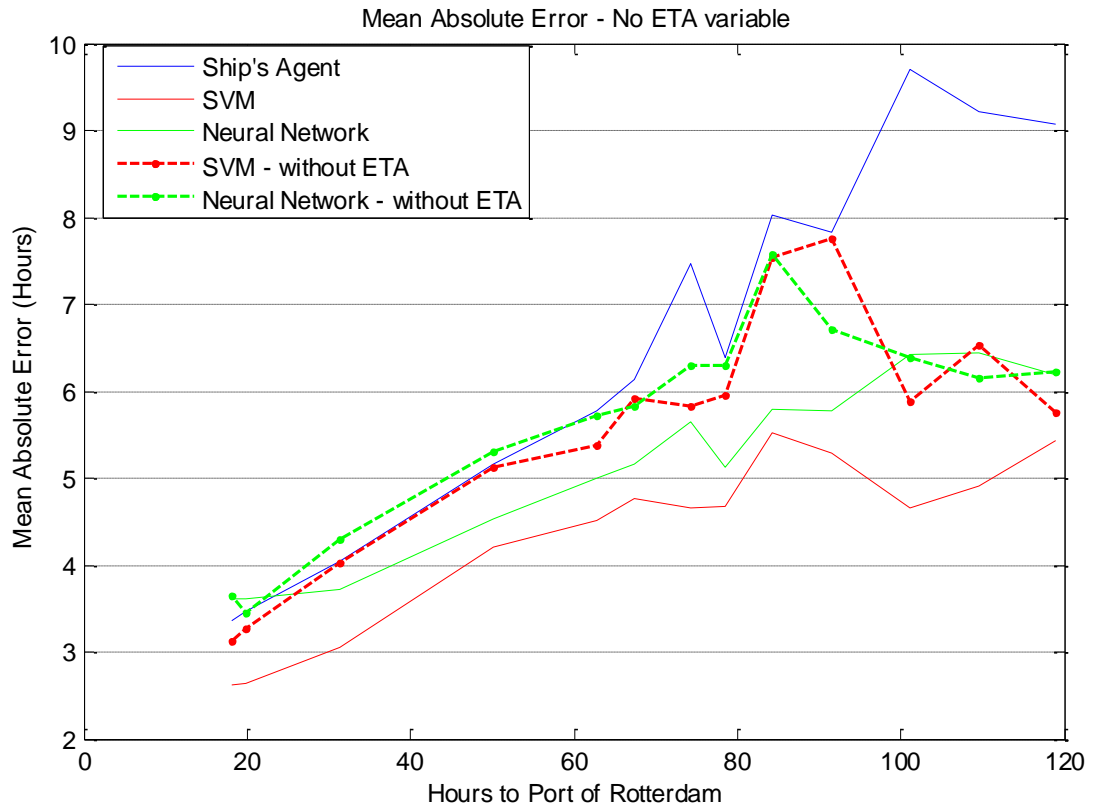
Figure 10: *RMSE on ETA predictions for Ship Agent, SVMs and Neural Networks – clustered weather variable approach*

As it can be noticed in the previous figures, both the SVMs and the Neural Network give more accurate predictions compared to the current situation, that is based on the ETA of the ship's agent. Furthermore, the SVM outperforms the Neural Network for every point in the time-horizon examined. The area where the predictions are significantly improved is between 80 and 120 hours away from the port of Rotterdam. For instance in the region 100-120 hours from Rotterdam the mean absolute error is around 5 hours for the SVM, while the ship's agent error is off by more than 9 hours. There is also significant improvement in the variance, as depicted by the rmse error metric. The regions between 80 and 40 hours away from the Port are characterized by medium improvement on the MAE, but still large improvements in the variance. The areas closest to the port of Rotterdam provide smaller improvement in the accuracy of predictions. That is because uncertainty in those areas is reduced and also, the ship agents update their ETA, since at that point information is becoming important for container terminals and the Port of Rotterdam.

5.2 Sensitivity analysis on Input variables

In order to assess the impact of the ETA of the ship's agent variable and the weather impact on the models, the neural network and SVMs were also trained and validated without those variables. The results can be seen in the figures below, where the first one presents the results when the ship's agent ETA was not given as an input to the Neural Network and SVM and the second one presents the effect when not using weather variables.
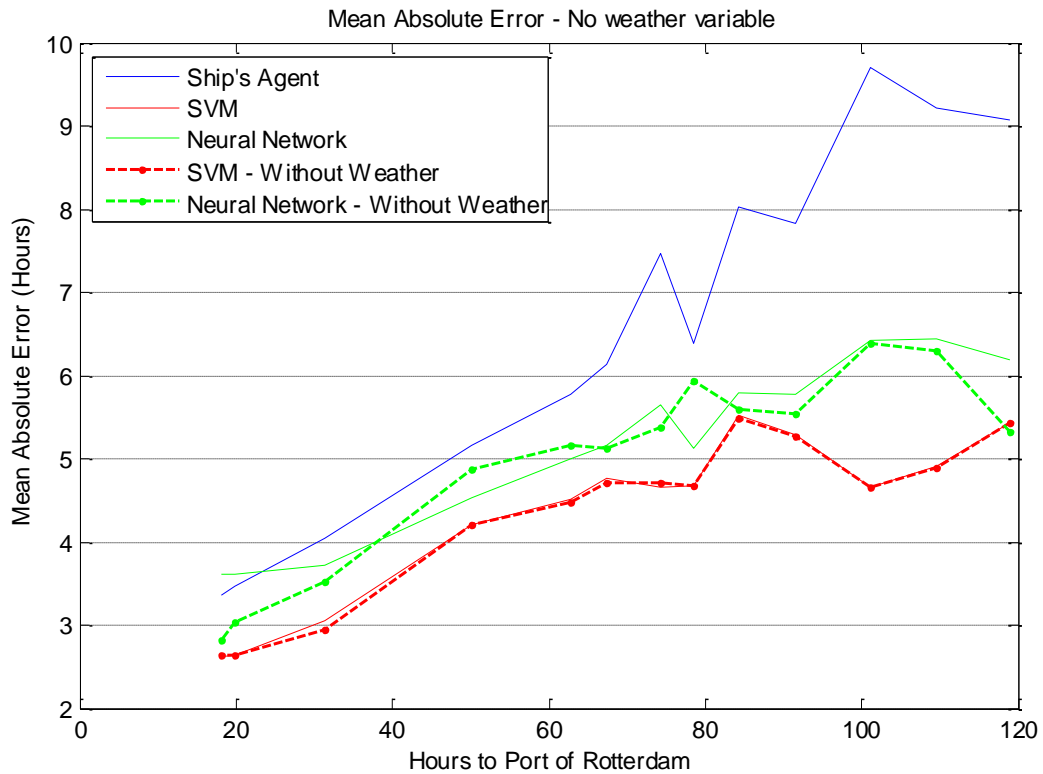
Figure 11: *Mean absolute errors on ETA predictions for Ship Agent, SVMs and Neural Networks – comparison between initial case and model without ETA variable*
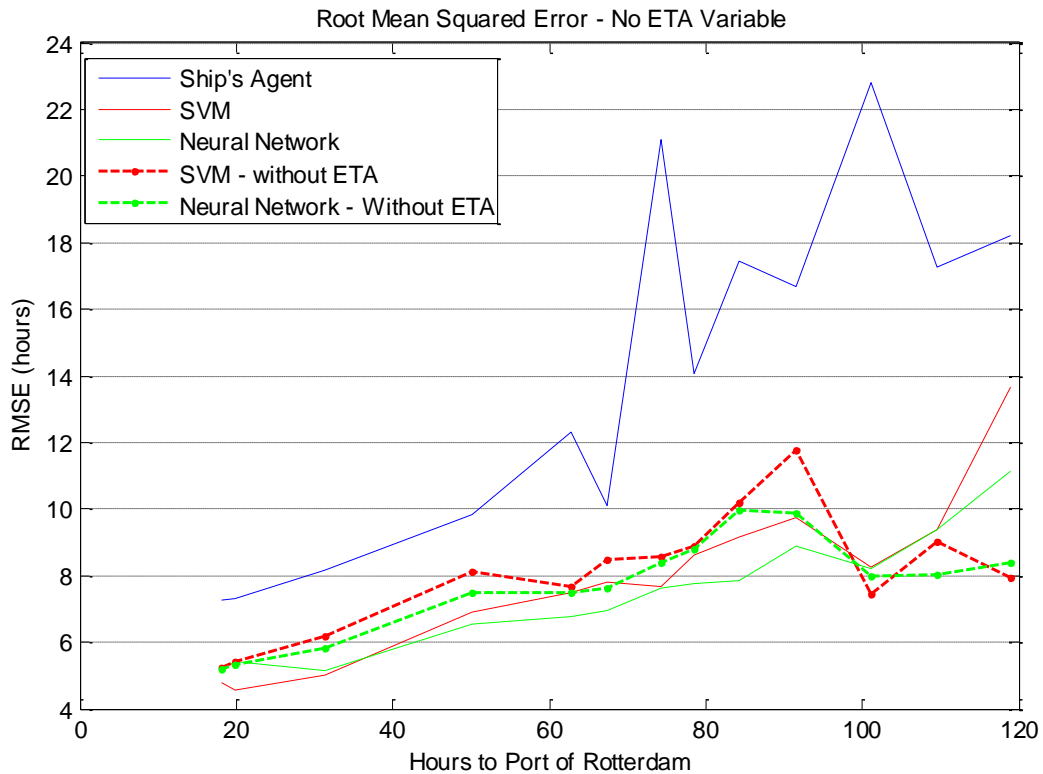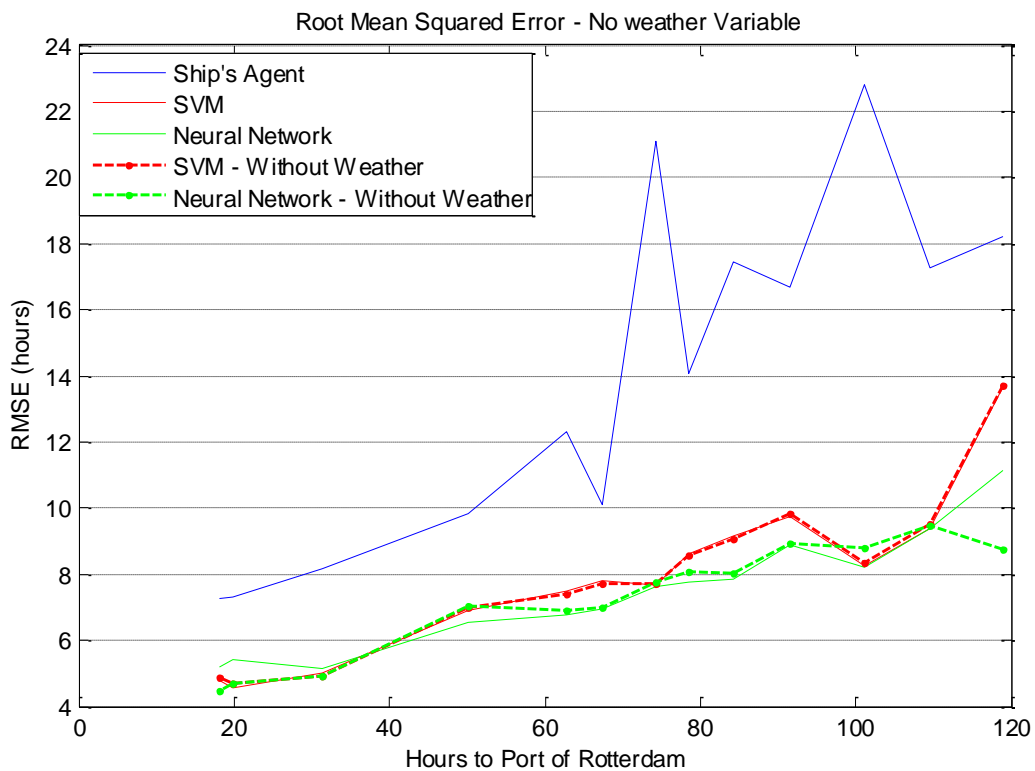
Figure 12: *RMSE on ETA predictions for Ship Agent, SVMs and Neural Networks – comparison between initial case and model without ETA variable*

As it can be observed in the graphs above, the Mean absolute error increases significantly both for the SVM and neural network when the ETA of the ship's agent is not provided as an input to the models. This is to be expected, since the ETA of the ship's agent gives an indication regarding the captain's intentions of how fast he is planning to travel the remaining distance until the Port of Rotterdam. This indication cannot be extracted from the other features of the input vector. However, the error for the regions positioned above 100 hours from the Port of Rotterdam does not increase by the same margin. This is because the ETA of the ship's agent for such distances presents significant errors and thus the models give less weight to it. One implication of this is that it is possible to give ETA predictions, through SVMs or Neural Networks, even when the ship agent has not provided an Estimated Time of Arrival. This can be the case of when trying to expand the time horizon beyond 5 days from arrival at the port. This deduction is reinforced by the fact that the root mean squared error is not affected significantly and remains much better than the respective error of the ship's agent.

Figure 13: *Mean absolute errors on ETA predictions for Ship Agent, SVMs and Neural Networks – comparison between initial case and model without weather variables*
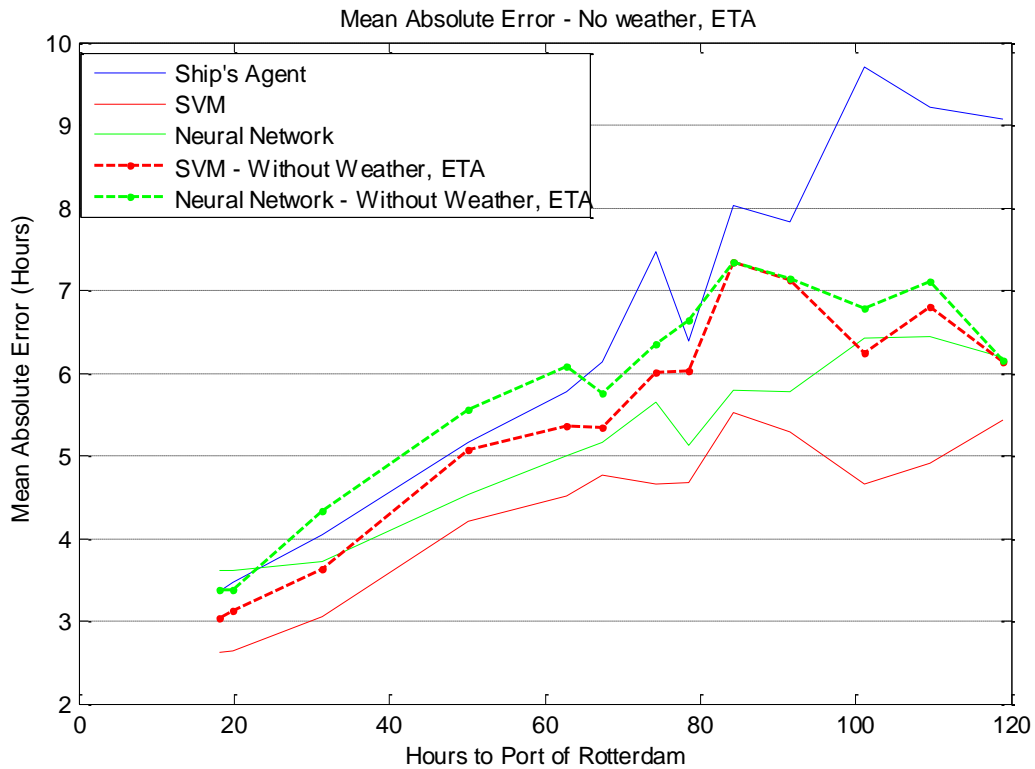

Figure 14: *RMSE on ETA predictions for Ship Agent, SVMs and Neural Networks – comparison between initial case and model without weather variables*

From the figures above it can be deducted that the proposed interpretation of weather has not an impact on predicting the Estimated Time of Arrival of sea-vessels, since the errors do not change in the case of not having a weather feature, compared to the initial case. This can be attributed to two factors. Firstly, the speeds that have been used, as obtained from the AIS data, are speeds over ground. This means that they already contain an interpretation of the currents and weather conditions in the area. More specifically, the speed over ground is a function of the engine power with which the ship is currently operating (in rounds per minute) and the currents/weather conditions in the area. Moreover, the captains have a deadline for final arrival at the Port and failing to meet the deadline will result in penalty costs. They also know that if their time of arrival is different than the ETA they have provided to the port, this may result in increased waiting time, because their berthing place may be occupied by another vessel. Therefore, their driving behavior is influenced mainly by the deadline for transporting the goods and not by the weather conditions. This means that even if the weather conditions are unfavorable, they will increase the engine power of the vessel's speed in order to arrive in time. That is why, a very common pattern that was observed in the data was speeding up in the initial leg, and then, if the deadline could easily be met, slowing down for the remainder of the route. This means that weather conditions definitely have a huge impact on the ship's fuel consumption, however, their impact on the time of arrival is bounded due to the ability of the captains to change the engine power of the vessel.

Last but not least, the interpretation of the weather conditions, as done for the neural network and SVM model, is applicable to the areas before Gibraltar, so more than 80 hours away from the Port of Rotterdam. As a ship moves ahead of a selected weather checkpoint along the route, moving closer to the port, the influence of the weather conditions becomes of lesser importance. Since the number of observations increases as the ships move closer to the port, this creates a lot of noise for the neural network and SVM, so they tend to disregard the weather variable. In order to overcome this limitation, different neural networks and SVMs have to be used for different areas. That way, for the areas that are more than 80 hours away from the Port of Rotterdam, a different model will be used than for the areas that are closer. This approach yields the best results and is presented at the end of the current chapter (section 5.3).

The next simulation was without both the ETA of the ship's agent and the weather variables as input to the model. That way, it can be assessed whether the captain's ETA takes into account the weather conditions ahead of the route. The results are presented below:
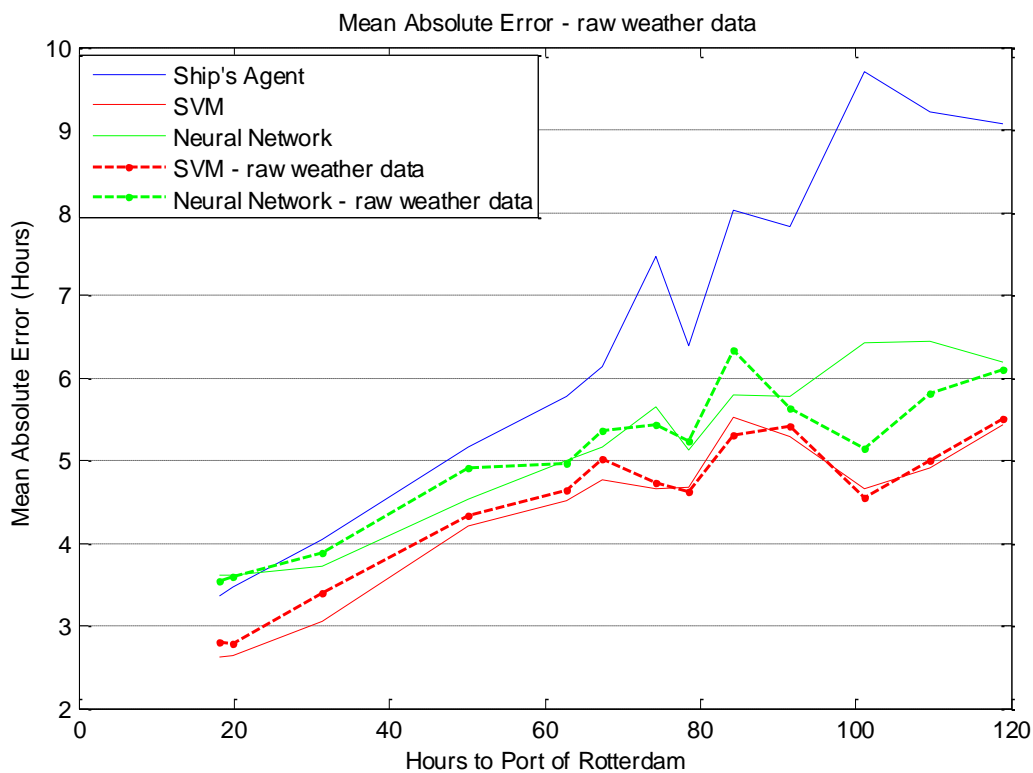
Figure 15: *Mean absolute errors on ETA predictions for Ship Agent, SVMs and Neural Networks – comparison between initial case and model without weather and ETA variables*
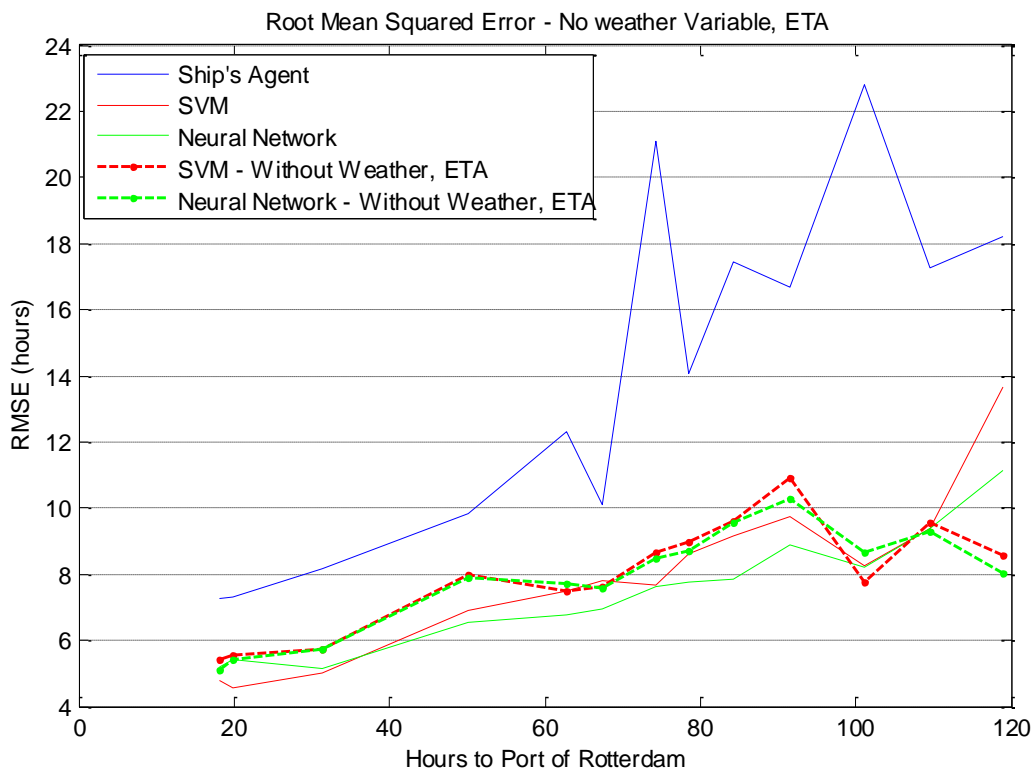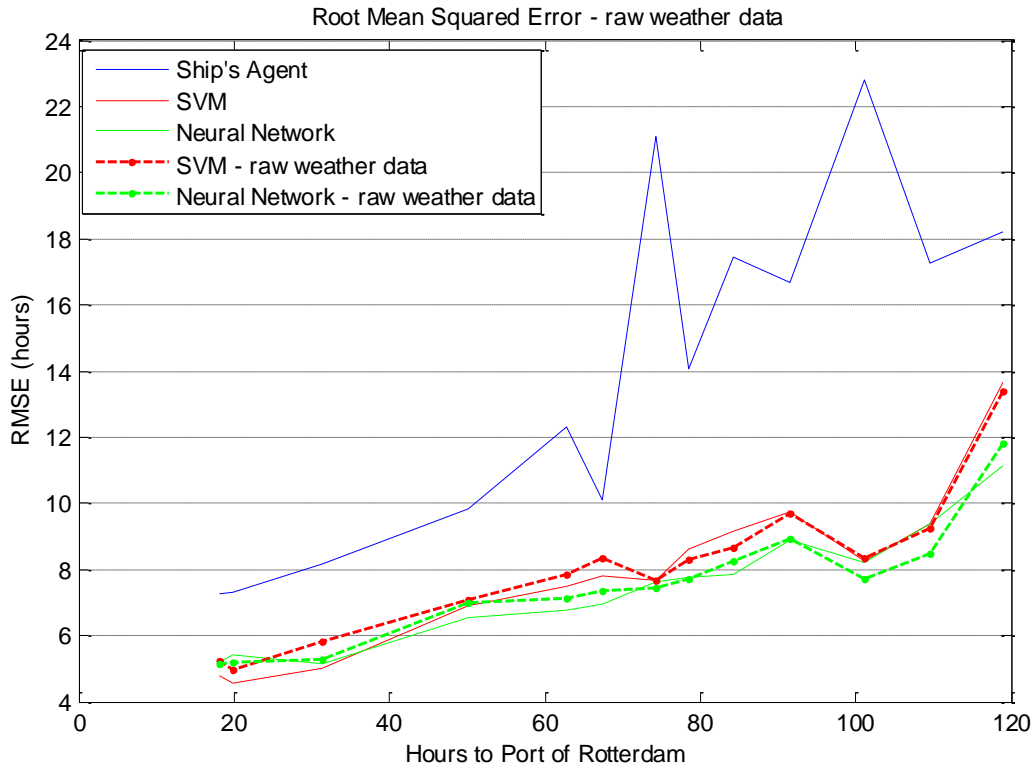

Figure 16: *RMSE on ETA predictions for Ship Agent, SVMs and Neural Networks – comparison between initial case and model without weather and ETA variables*

In general, it can be noticed that the errors of the above graphs are very similar to those where only the ETA variable was missing. This was to be expected, since that variable is of significant importance for having an indication of the captain's intentions regarding the voyage. What is interesting to note, is that when removing both variables, compared to only without ETA, the errors below the 60 hour to the port region are reduced a bit. This is because, as mentioned earlier the weather interpretation does not have an influence on those regions. On the contrary, when removing both variables the error in the regions above 100 hours increased by half an hour (always compared to Figure 11). This means that on the absence of captain's ETA the weather interpretation has some influence on those areas in estimating the time of ship arrival at the Port.

The final simulation that was attempted, was to include the raw weather data obtained from the weather checkpoints, without any clustering preceding them. What can be deduced from the results presented below, is that the raw weather data did not have an influence on the results, since the errors with the clustered approach and with the raw weather data did not differ significantly. In fact, in the areas close to the Port (below 60 hours) the errors with the raw weather data increased compared to the initial case.

Figure 17: *Mean absolute errors on ETA predictions for Ship Agent, SVMs and Neural Networks – comparison between clustered weather model (initial case) and model with raw weather data*

Figure 18: *RMSE on ETA predictions for Ship Agent, SVMs and Neural Networks – comparison between clustered weather model (initial case) and model with raw weather data*

5.2.1 Insight into the neural network predictions

In this section, an attempt towards understanding which variables, from those used for predicting the ETA, were mostly driving the predictions of the machine learning techniques. It has already been identified that the weather variables had minimal impact on making predictions regarding the estimated time of arrival for the sea-vessels at the Port of Rotterdam. On the other hand, it was found that the ETA of the ship's agent was significantly reducing the error of the SVM and neural network, when provided as an input to the models. However, even on its absence as an input, acceptable predictions could still be made.

Having these in mind, the following figure presents the weights assigned to the different inputs of the neural network, with the aim of getting an insight regarding which variables had the larger impact in shaping the predictions. It should be noted however, that this depiction of the weights assigned to the neural network can only serve as an indication regarding which variables were mainly driving the predictions. After all, there are two layers of weights involved in a neural network, which means that the importance of variables can change in the second layer. This is why neural networks are often characterized as "black boxes". Also, a similar intuition regarding the predictions of the SVM is not feasible. However, due to the similar results of the two methods, it can be deducted that SVMs would most probably give approximately the same importance to the input variables as the neural networks.

Neuron Number	Bias unit	Latitude	Longitude	Distance	Ship breadth	Ship length	Speed now	change in speed	average speed	time used for calculating average speed	weathe variable	ETA of ship Agent	Identifier for region
1	-1,18	0,03	0,18	0,18	-0,07	-0,07	-0,20	-0,06	-0,11	0,07	0,04	0,21	-0,08
2	-0,31	0,03	-0,01	0,21	-0,04	0,03	-0,03	-0,04	0,06	-0,03	0,00	0,26	-0,03
3	-0,31	0,03	-0,01	0,21	-0,04	0,03	-0,03	-0,04	0,06	-0,03	0,00	0,26	-0,03
4	-0,03	0,02	-0,08	0,21	0,00	0,05	-0,04	0,05	-0,07	-0,02	-0,01	0,26	-0,02
5	0,31	-0,03	0,01	-0,21	0,04	-0,03	0,03	0,04	-0,06	0,03	0,00	-0,26	0,03
6	-0,33	0,04	0,16	-0,22	0,01	-0,01	-0,03	-0,05	0,09	-0,03	0,01	-0,25	0,05
7	0,14	-0,03	0,06	-0,21	0,01	-0,04	0,06	-0,04	0,04	0,03	0,01	-0,26	0,02

Table 3: *Weights assigned to the neural network, connecting the inputs and the hidden layer. The bias unit is an extra column of value one for each element, applied for statistical correction.*

As it can be observed from Table 3, relatively high weights are placed on connecting the ETA of the ship agent (as expected) and the distance, to the neurons of the hidden layer. This is an indication that these variables were important in predicting the ETA of the sea-vessels to the Port of Rotterdam. Apart from that, some high weights can be noticed connecting the current and average speed of the vessel to the hidden layer, as well as the longitude of the vessel's position. It can also be noticed that the weather variable has weights very close to zero, which means that it was considered of minor importance for making ETA predictions, something which is in alignment with the findings of the previous section. The bias unit is only used as a constant for applying a statistical correction, so no additional insight can be gained from that. The weights connecting the hidden layer to the output, provide no additional insight for understanding which variables drive the ETA predictions and is included in the Appendix for the purpose of completeness (Appendix A).

## 5.3 Different Models for different geographic regions

In order to further improve the predictions compared to the initial case and to better assess the weather impact interpretation for the different geographic areas on the route to Port of Rotterdam, different neural network and SVM models have been used for distinct geographic areas. To be more precise, 5 different SVM models and neural networks have been used. For the regions when a vessel has already passed a weather checkpoint, the variables in that weather checkpoint are excluded. That way in each prediction only the weather conditions ahead are taken into account.

The following graphs present the results both for SVM and neural networks for the initial case and when multiple models have been used (with the dashed lines).
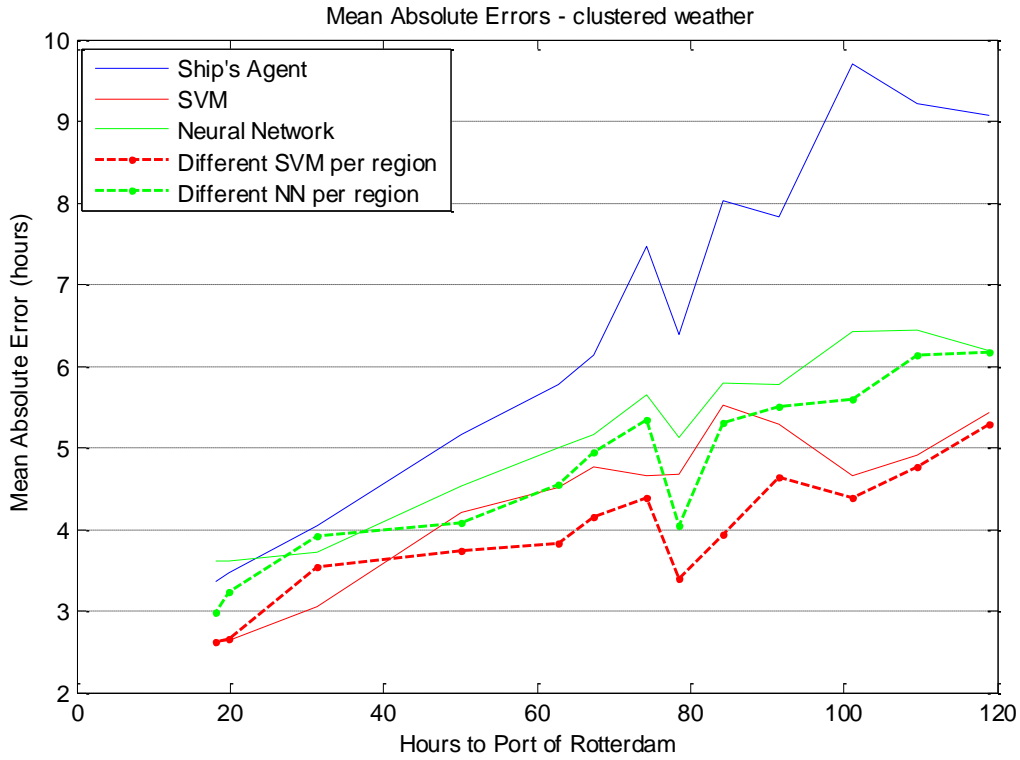
Figure 19: *Mean absolute errors on ETA predictions for Ship Agent, SVMs and Neural Networks – comparison between initial case (normal lines) and case of using different learning algorithms for distinct regions (dashed lines)*
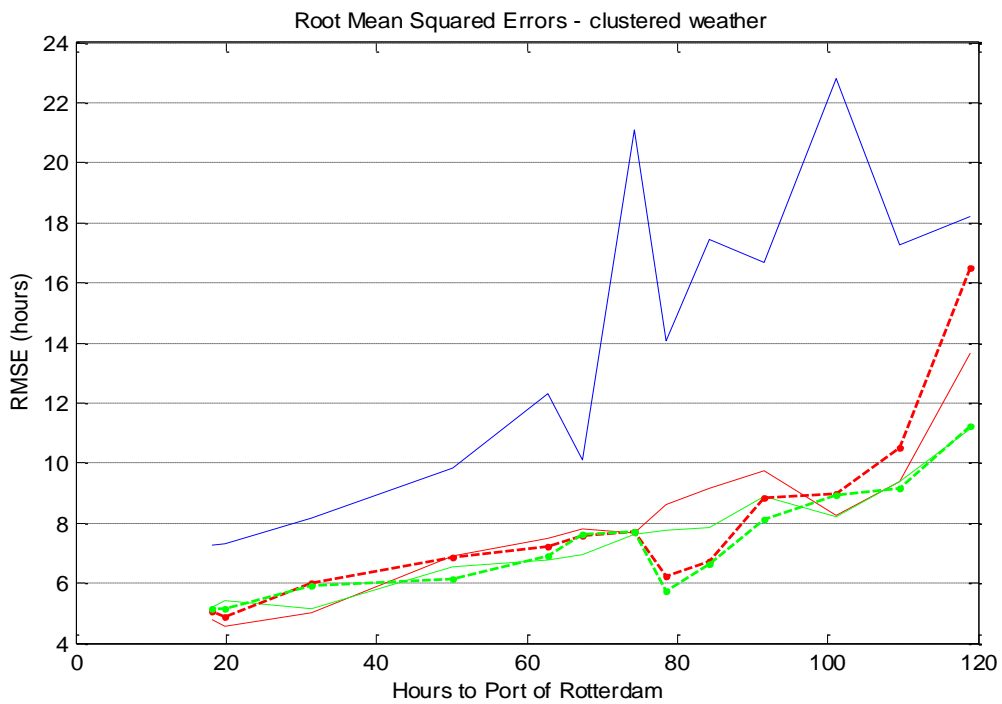
Figure 20: *RMSE on ETA predictions for Ship Agent, SVMs and Neural Networks – comparison between initial case (normal lines) and case of using different learning algorithms for distinct regions (dashed lines)*

As it can be noticed in the previous graphs, in the case when a different model is used for each region the mean absolute error is reduced. This can be attributed to the fact that there is less variance in the ETAs that each model is trying to predict, and thus, generalization becomes easier. It should be noted that if the dataset increases sufficiently, so if there were more voyages available for training the SVM and neural network, the errors of using 1 SVM/Neural Network for the whole problem would be reduced to match the errors of when using multiple models for the different geographic areas.

Furthermore, in every case the support vector machines outperform the neural networks, due to their ability to generalize better when there are relatively few data available for training. One final point to note is that for the cases where the vessels are positioned approximately 110-120 hours from the port, the variance in the predictions with the SVM seem very high, compared to its performance in the other regions. This is mainly attributed to the fact that for those regions very few historical data on voyages were available in the dataset used. Therefore, some cases with big errors could be assigned to the testing set, causing a high variance in the prediction error for those initial points.

Having experimented with different machine learning techniques and determined the influence of the input variables for making predictions the following observations can be made:

- SVMs and Neural Networks achieve significantly better results in predicting the ETA of vessels positioned more than 60 hours away from the port of Rotterdam, compared to the current situation (ETA of ship's agent)
- SVMs outperform Neural Networks for the time horizon investigated
- Having the ETA of the ship agent as input to the prediction algorithms, increases their performance substantially, since it gives an indication of the captain's behavior
- The weather information used for predicting the ETA of containerships was found to be of limited value, since the captains can change the engine power of the vessel to counter any adverse conditions.

Overview of Chapter 5

In this chapter the results obtained while using Neural Networks and support vector machines for predicting the ETA of containerships were presented. The two methods were compared to the ETA of the ship's agent and found to be improving the current situation significantly. That way, the final sub-question, regarding the development of an ETA prediction model, was sufficiently addressed. Also, the value of the different input variables for making the predictions was investigated, with the ETA of the ship's agent proving to improve the performance of the algorithms significantly. The following table presents an overview of the methods used and the results that they achieved.

	MAE (hours)			RMSE (hours)		
Hours to Port of Rotterdam	SVM	Neural Networks	Ship Agent	SVM	Neural Networks	Ship Agent
100-120	5,0	6,3	9,3	10,4	9,6	19,4
80-100	5,2	5,5	7,4	9,2	8,2	16,0
60-80	4,6	5,2	6,5	7,7	7,1	14,5
40-60	4,2	4,5	5,2	6,9	6,5	9,8
20-40	2,7	3,6	3,6	4,8	5,2	7,6

Table 4: *Overview of errors for the prediction methods used (SVMs, Neural Networks) compared to the current situation (Errors of ship Agent)*

Chapter 6 - Practical implications of the results

This chapter provides an interpretation of the results obtained in chapter 5 from the perspective of the different stakeholders involved in container transport. Special attention is given to the value of predictions depending on the time-horizon and accuracy. The benefit of the improved ETA predictions is quantified in section 6.2 for the container terminals. The next steps towards the realization of the proposed ETA information tool are described in section 6.3.

6.1 Value of findings from the Stakeholder's perspective

From the results that have been obtained using support vector machines for estimating the ETA of containerships, it can be observed that the improvement is substantial in the regions that on average are positioned above 60 hours away from the port of Rotterdam. Moderate improvement is noticed in the regions from 60-40 hours to the port. For short distances from the Port of Rotterdam, the SVM errors are still better, but closer to the errors of the ship's agent. This can signify the added value that the predictions can have for the different stakeholders involved in container transport.

Due to the improvement in long time horizon predictions, the party that will mostly be benefited is the hinterland transportation side. The hinterland transportation parties, such as EGS are planning their schedule much in advance from arrival at the Port of Rotterdam. Every week, they have to decide on how much capacity to book for barge and rail. A deviation in the pre-allocated capacity from the actual cargo to be transported incurs additional costs. These costs can be saved through a more accurate estimation of vessel arrivals. They can also plan better for the schedule of barges and trains.

The container terminals on the other hand, are currently more interested in short term horizon predictions, where the improvement of prediction accuracy compared to the current situation is moderate. Therefore, the information tool based on SVMs provides additional value for their planning activities (how to allocate berthing place, manpower, equipment), but of less impact than that compared to the hinterland transportation side. It was also identified that of significant value for terminal operators would be to know the waiting time of vessels, when they previously call in another European port, such as Antwerp or Felixstowe, before going to the Port of Rotterdam. Such a decision support tool is proposed as area for further research at the end of the report.

For the carriers, since the weather interpretation did not add to the improvement of ETA predictions, the current information tool can only serve as an indication for the vessel arrival to the Port based on speed, position data and the technical ship characteristics. Apart from that, there is the indirect benefit of lower handling times for the vessel, due to the optimization in container terminal operations. This means less idle time for the ships and thus, faster delivery of the goods and an opportunity to start earlier on the next voyage. However, the biggest value that the proposed tool, as it currently stands, can offer to carriers, is as a competition monitoring tool. By aggregating the ETAs of all the vessels in the examined area, it is possible to know how many ships will be arriving during the next days at

the Port of Rotterdam and in which time slots. Knowledge of this information can give bargaining power to a carrier company. For instance, if it is known that very few ships are in the area and the company's vessel is of the first to arrive at the port, it can negotiate on high prices for transporting cargo from the Port of Rotterdam to another destination.

As far as the Port of Rotterdam is concerned, there are both direct and indirect benefits involved. On the direct benefits is the ability to plan better for pilot availability to guide the vessels to the terminals, as well as reduction of traffic congestion around the areas of the Port. The latter is because of the fact, that better estimation of arrivals can result in better planning from the truck operators who would therefore arrive closer to the actual arrival time, instead of waiting around the Port. On the indirect benefits, the competitive position of the port compared to other ports in the area will greatly be enhanced, due to the decreased handling time at container terminals, thus benefiting the shippers, as well as the decrease in the cost of hinterland transportation to inland destinations. Due to its leading position in the transportation chain, the Port of Rotterdam should be the main benefactor of this information tool, providing the information to the other stakeholders, such as terminals and hinterland transportation parties, through a common platform.

The following figure summarizes the additional benefit for the stakeholders, depending on the time-horizon which is currently used for the purposes of their planning activities:
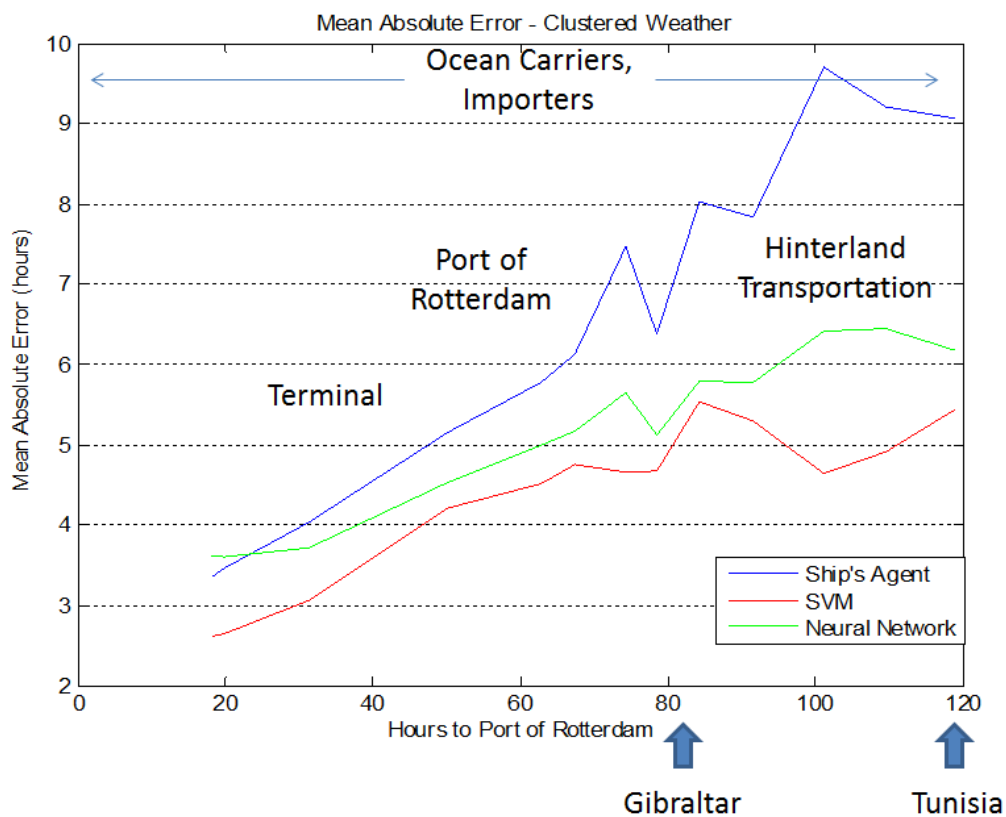
Figure 21: *The time-horizon of ETA predictions for which the stakeholders would mostly be interested, according to their planning activities*

6.2 Estimation of monetary benefits from predictions

There are multiple sources of expenses that can be reduced through improved ETA predictions. For the container terminals within the Port of Rotterdam, handling expenses can be reduced, while from the hinterland transportation parties, intermodal transportation will increase, thus reducing the transportation costs. Also, the terminals are subject to demurrage costs when they need more than 3 days to unload a vessel . These are costs that can be minimized through a more robust planning procedure, which can be achieved through early identification of disturbances regarding the vessel arrivals. For the purposes of this study, the savings for the terminals in the Port of Rotterdam will investigated, by examining the handling costs.

The average handling time of container vessels at the Port of Rotterdam was 30,3 hours in 2013 (Römers, 2013). Through the more accurate predictions obtained in this study regarding the expected arrivals at the Port, it is assumed that the handling time for the vessels following the Asia-Rotterdam route will be decreased by 2 hours. This is the benefit that improving the ETA predictions by 4 hours on average for the vessels positioned more than 80 hours away from the port, is expected to have. Furthermore, it is estimated that those voyages (Asia-Rotterdam directly) account for the 25% of the calls at the Port of Rotterdam (Appendix E). This means that the average decrease of handling time is 2*25%=0,5 hours. Therefore, the average handling time is expected to be 29,8 hours, improved by 1,65 % compared to the current handling time.

Now the handling expenses for the container terminals should be estimated, to understand the impact of improving handling times by 1,65 %. From the Port of Rotterdam balance sheet, the amount spent on wages is 76 million Euros annually (Port of Rotterdam, 2014). It is assumed that 80% of those are variable costs (that can be reduced by lower handling times) and 40% of those expenses are related to container terminal employees, thus $76 * 0,8 * 0,4 = 24,32 \; million \; Euros$ are the handling costs for manpower, annually. By improving handling times by 1,65% the savings realized are $24,32 * 1,65\% \cong 400.000 \; Euros$. Therefore, through the implementation of the ETA information tool, it is estimated that the Port of Rotterdam can reduce its handling expenses by 400 thousand Euros annually.

One final point to note, is that if the ETA information tool is implemented for more routes than those following the Asia - Rotterdam route directly, the benefits will be even greater. Quantifying the positive impact in reducing demurrage and hinterland transportation costs is left out of the scope of the current thesis.

6.3 Next steps towards the implementation of ETA predictions

As indicated in chapter 5, the AIS data alone, are enough for making ETA predictions for the route and time-horizon examined. The weather variables on the other hand, did not seem to improve predictions regarding the ETA of the vessels. Therefore, there is currently no additional cost involved for acquiring the necessary data. What is needed is for the real-time AIS data received by Intertransis, to be fed forward to the data pre-processing algorithm, in order to select the variables that were used for making ETA predictions in this thesis. Then, the input variables can be provided to the already trained SVM algorithm, which will produce the ETA prediction. This procedure requires no additional manpower or cost, other than connecting the data pre-processing algorithm to the AIS real-time data receiver.

Regarding the next steps towards the implementation of the ETA information tool, a real-time testing case should be undertaken for a time period spanning over some months in the future. During this testing phase, real-time AIS data, as acquired by Intertransis, will be used as input to the trained SVM model and the ETA predictions will be produced on a rolling-time horizon. Then, when the ship reaches the Port, by calculating the difference between the prediction and the actual time of arrival, the accuracy of the model can be tested to see if it responds as it is expected. If the errors are indeed within the region specified by the report, the system can go online and start making ETA predictions that can assist in the planning activities of the interested parties.

For the maintenance of the ETA prediction algorithm, there will be a need to update the SVM model, to account for the most recent voyages. The way that this task can be executed, is to retrain the SVM algorithm every year, each time adding to the training set the new voyages of the year that passed. This way, the algorithm will be trained to account for the changing circumstances in the shipping world. For instance, if in the future the market becomes more profitable for the ocean carriers, they will start travelling at higher speeds to maximize their revenue by increasing the number of voyages. An algorithm that is re-trained every year will be able to capture such changes in the market behavior, by recognizing the new patterns. This maintenance work can be carried out by TNO.

Overview of Chapter 6

In this chapter, the added value for the stakeholders in container transport, according to the improvement achieved for the different time horizons, was presented. It was deducted that the ETA predictions obtained were more valuable for the hinterland transportation parties, due to the improvement achieved in the medium-long time horizon. In section 6.2, the benefits in terms of reduction in handling costs were assessed for the Port of Rotterdam, where it was found that the manpower costs for vessel handling can be reduced by 400.000 Euros annually. The chapter ends with a description of the next steps towards the development of the ETA information tool.

Chapter 7- Conclusion

The previous chapter presented the results regarding estimating the ETA of containerships, using neural networks and support vector machines. In this chapter, the importance of the results will be discussed, pointing out which of the stakeholders would be mostly benefited from them. A summary of the answers to the research questions posed in the beginning of the report will also be provided. The report concludes with proposed areas for further research.

7.1 Discussion

7.1.1 Main academic findings

For the purposes of the current research, two models were used for estimating the time of arrival of containerships at the port of Rotterdam, accounting for a middle time horizon, namely the support vector machines and neural networks. The support vector machines were found to outperform the neural networks when comparing the mean absolute error of the models in all cases. This can be attributed to the fact that through the principle of similarity, on which SVMs function, they were better able to generalize on unseen data. The neural networks would need more data available for achieving the same results, since their training requires more examples to generalize well.

Both of the models achieved significantly better results than the current situation, where the ETA is based on the ship's agent estimations. In the optimal results obtained, when using different SVMs for each region, the mean absolute error reduced from 9 hours to less than 5 for areas positioned above 100 hours from the Port of Rotterdam (Western Mediterranean) and for Gibraltar (80 hours from the Port of Rotterdam) the error reduced to 4 hours, while the ship's agent error is 7,5 hours.

Of equal significance is the fact that the root mean squared error (signifying the variance in prediction errors) reduced drastically. For instance in the case of regions that are positioned on average 100 hours from the Port of Rotterdam by ship, prediction errors had an rmse of 9 hours, while for the ship agent the respective value was 20 hours. This means that in 95% of the cases the prediction error of the proposed SVM was less than 18 hours ($2*\sigma_{SVM}=2*9$), while for the ship's agent, 95% of the cases were lying within a time interval of 40 hours ($2*\sigma_{agent}=2*20$). Therefore, the uncertainty over vessel arrivals at the port is significantly reduced through the proposed model.

Regarding the relevance between weather conditions and their importance for predicting the ETA, it was found that they do not play a crucial role for estimating vessel arrival for the route that was examined. This is attributed mainly to two reasons. The first is that the speeds that have been used as inputs to the model are speeds over ground, therefore information regarding the currents and weather conditions in the area are included in these values. The second reason is that the driving behavior of the captain is largely affected by their deadline for delivering the goods and their desire to save on fuel consumption when possible. This results in a speed up for the first leg of the trip and slowing down when arrival at the Port before the deadline, is guaranteed. It also means that even if weather conditions are

unfavorable within a region, the captain can still achieve a high speed by changing the engine power of the vessel. Another possibility is to speed up at a later part of the route. The number of options available to the captain make it extremely difficult to forecast his driving behavior based on the weather conditions ahead.

Nevertheless, it cannot be concluded that weather conditions and currents have no impact on the time of vessel arrival to the Port. What can be deducted, is that for the specific route, from Tunisia to the Port of Rotterdam, the attempted ways to interpret the weather yield no benefit for predicting the ETA of sea-vessels. Moreover, even if there is a way to add additional insight to ETA predictions through a different weather interpretation, there is not enough room for improvement in the prediction error. An error of 5 hours on average in ETA estimation from Tunisia to the Port of Rotterdam is already considered good enough for the planning activities of the stakeholders involved, as assessed through the interviews. Even if a weather modelling approach could yield somewhat better results in predictions, the extra effort of acquiring and using the data for the model development would be more time-consuming than the additional value created. This means that for making ETA predictions for vessel arrivals at the Port of Rotterdam, positional and speed data, as provided in the AIS database, are sufficient.

A final point to note here is that the route that was examined for the purposes of this research did not include extreme weather conditions, since most of the route is taking place close to the shore and no big ocean is crossed. The case may be very different when routes that are crossing over oceans are examined. In those cases, weather conditions may have significant impact on the ETA of the sea-vessel.

Another finding of the research was that the variable regarding the ETA of the ship's agent, when provided as an input, was giving a significant improvement to the prediction accuracy of both the SVMs and neural networks. This was because of the fact that it served as an indication regarding the captain's intentions along the route, information that cannot be extracted from the other variables. However, for areas that are positioned more than 100 hours away from the port of Rotterdam, the error of the ship's agent is very high. Therefore, when it was attempted to make predictions without that variable the error of the SVM did not increase as much as in the areas positioned closer to the Port. This means that when trying to expand to longer time horizons, that the ETA of the ship agent may not have been provided, it is still possible to make predictions based on the positional and speed data provided in the AIS dataset.

7.1.2 Answers to the research questions posed

In this section, the answers to the research questions that were posed in the beginning of the report will be summarized. The main research question of this study was: *How can the big data, provided by marine traffic providers, be used in order for container terminals to improve their business processes by addressing the uncertainty regarding the expected vessel arrival times at the Port of Rotterdam?*

To answer the main research question, three sub-questions were formulated. Each one of them will be addressed here. Combining the answers of these sub-questions, the main research question can sufficiently be addressed.

1. *What is the added value that a more accurate prediction of the estimated time of arrivals for containerships at container terminals would have for the planning activities of the stakeholders involved in container transport?*

The benefits of more accurate ETA predictions, taking into account the results obtained in the current study, have been analyzed in chapter 6. For the purpose of completeness, they are also briefly summarized here.

- Hinterland Transportation parties (e.g. EGS): Cost reduction due to the ability to decide in advance how much capacity to book for barge and rail. Also, better planning of the barge and train schedules can be achieved, since there is greater certainty over vessel arrivals for a time horizon spanning over 5 days.

- The container terminals (e.g. ECT): They are currently more interested in short term horizon predictions, where the improvement of prediction accuracy compared to the current situation is moderate. Therefore, the information tool based on SVMs provides additional value for their planning activities (how to allocate berthing place, manpower, equipment), but of less impact than that compared to the hinterland transportation side.

- Carriers: The ETA information tool can serve as an indication for the vessel arrival to the Port based on speed, position data and the technical ship characteristics. Apart from that, there is the indirect benefit of lower handling times for the vessel, due to the optimization in container terminal operations. This means less idle time for the ships and thus, faster delivery of the goods and an opportunity to start earlier on the next voyage. However, the biggest value that the proposed tool, as it currently stands, can offer to carriers, is as a competition monitoring tool. By aggregating the ETAs of all the vessels in the examined area, it is possible to know how many ships will be arriving during the next days at the Port of Rotterdam and in which time slots. Knowledge of this information can give to a carrier company bargaining power when negotiating for cargo transportation.

- Port of Rotterdam: there are both direct and indirect benefits involved. On the direct benefits is the ability to plan better for pilot availability to guide the vessels to the terminals, as well as reduction of traffic congestion around the areas of the Port. The latter is because of the fact, that better estimation of arrivals can result in better planning from the truck operators who would therefore arrive closer to the actual arrival time, instead of waiting around the Port. On the indirect benefits, the competitive position of the port compared to other ports in the area will greatly be enhanced, due to the decreased handling time at container terminals, thus benefiting

the shippers, as well as the decrease in the cost of hinterland transportation to inland destinations.

2. *Which are the main factors and to what extent are they affecting the average speed of the loaded containerships sailing towards container terminals?*

The second sub-question was initially answered through the literature review. There it was identified that the nominal speed of the vessel (based on the power of the ship's engine in rounds per minute), the impact of the wind on the sailing speed, as well as the impact from currents and waves are important variables for estimating the ETA of the vessel (Szelangiewicz, Wiśniewski, & Żelazny, 2014). The direction of the currents, wind and waves, relative to the vessel's direction were also taken into account.

However, as it was found after the usage of SVMs and neural networks, the impact of the weather variables is of limited value when it comes to estimating the ETA of sea-vessels for the route under consideration. They definitely have an effect on the temporary speed, which can be seen in the AIS data since it is the speed over ground, but the captain can change the engine's power along the route. This means that weather variables can have a large impact when it comes to estimating the fuel consumption of a voyage.

Another factor that was identified to be of high relevance for the vessel speed was related to the market of containerships, suggesting that in periods where demand is relatively low, vessels sail at the lowest attainable speeds (slow steaming). This was also the case for the time period that was examined in the current research (Wright, 2016). To account for this market-related factor in the eta-estimation stage, from the data available, the ship length and breadth were used, as an indication regarding the abilities of the ship engine. Also, the current speed with which the vessel is sailing gives an indication for its engine abilities. These factors were used as a starting point for the ETA-prediction stage.

3. *How can a model be developed in order to accurately predict the estimated times of arrival for the containerships that have the port of Rotterdam as their destination, accounting for a medium-range time horizon?*

Based on the factors affecting the vessel speed that were identified above, two different machine learning techniques were used for estimating the time of vessel arrivals at the Port of Rotterdam. The first approach was based on neural networks and the second on support vector machines. Neural networks are based on the interconnections between layers of neurons to depict difficult relationship functions between the input and output variables. A cost function, based on the divergence between the outputs and actual values, is minimized in order to select the optimal weights connecting the different neurons. On the other hand, support vector machines are based on the concept of similarity, as decided by the Euclidean distance between inputs, to categorize similar kind of inputs. Then, similar inputs give similar prediction values.

Three distinct sets from the historical data of the voyages in 2015 and early 2016 were chosen, the training, validation and testing set. The training set was used for "teaching" the models on how to recognize patterns in the data. Then, the validation set was used for selecting the optimal parameters for the neural network and support vector machine respectively. The testing phase, depicted the real world error that the algorithms are expected to have.

The two methods were evaluated based on their mean absolute error and root mean squared error of their forecasts. It was observed that both methods were performing much better than the current ETA estimations, based on the ship's agent, especially for long time horizons (above 60 hours). Moreover, the support vector machines were found to outperform the neural networks in every case, as signified by the lower mean absolute error that they presented in the testing set. This was due to the fact that given a limited dataset, such as in the case at hand that consisted of 600 voyages in total, SVMs were more capable of generalizing the data patterns presented. The neural networks on the other hand require abundance of data in order to generalize from the patterns presented in the training set. The explanation of the difference between the two algorithms, lies in the fact that neural networks solve equations to decide on the optimal weight connecting their neurons, while SVMs are categorizing inputs according to their similarity in order to predict.

With the information above, the main research question can now sufficiently be addressed from the perspective of the port:

How can the big data, provided by marine traffic providers, be used in order for the stakeholders involved in container transport to improve their business processes, by addressing the uncertainty regarding expected vessel arrival times at the Port of Rotterdam?

Due to its role as information broker, Intertransis should be responsible for communicating the information regarding the predicted ETA to the Port of Rotterdam and the other stakeholders in container transport, such as terminals and hinterland transportation parties. This can be realized through a common platform. Intertransis is already receiving AIS data from marine traffic providers. One of the most important findings of this study was that the AIS data alone, are enough for making ETA predictions for the route and time-horizon examined. Therefore, there is currently no additional cost involved for Intertransis to acquire the necessary data. What is needed is for the real-time AIS data received, to be fed forward to the data pre-processing algorithm to select the variables that were used for making ETA predictions in this thesis. Then, the input variables can be provided to the already trained SVM algorithm, which will produce the ETA prediction. This procedure requires no additional manpower or cost, other than connecting the data pre-processing algorithm to the AIS real-time data receiver. However, for the full realization of the ETA information tool benefits from the perspective of the Port, a common platform should be introduced through which the information can be communicated to the other parties (e.g. ECT, EGS).

7.2 Reflection

Due to the complexity of the problem of predicting the ETA of sea-vessels at the Port of Rotterdam, a lot of choices had to be made and there are many approaches that could be followed. In this section, some of the challenges regarding the research undertaken will be highlighted, alongside with other possible approaches and the next steps towards the full implementation of the ETA information tool.

One of the main choices that had to be made is the way to interpret the weather conditions and their influence of the ETA of containerships. This was a lot of information per voyage since the weather conditions were available over a large number of points along the vessel's route. Therefore, 4 regions where significant changes to the speed of the vessel were usually noticed, were chosen for the analysis. Those areas served as weather checkpoints in the hopes that information about the currents and weather conditions in those areas would give an indication regarding the driving behavior of the captain. However, due to the limited total number of voyages available (600 for the period examined) the number of input variables had to be kept to a minimum. That is why a clustering approach of the weather variables was implemented, to reduce the number of variables.

Also, another aspect of it is that actual weather data were used for the purposes of the research. However, in a real-world setting the data available will be weather predictions for the route ahead. With the clustering approach this does not change anything, since as long as the predictions and the actual data are mapped to the same cluster everything will work the same way. On the other hand, if raw weather data are used, these are susceptible to errors of weather predictions.

Another possible way of choosing the weather-area checkpoints would be to use multilinear regression for selecting the regions that have the larger impact on the ETA of the vessels. Then the weather inputs of the neural network or support vector machine would be provided to them by a multilinear regression, a structure that resembles more deep learning algorithms. On the other hand extra complexity is added. An approach like this would be more worthwhile to test in voyages that cross oceans, such as the Atlantic, where weather impact may have larger impact than the one identified in the present case. In such cases, gathering a large amount of data can enable the usage of multilayer neural networks, which can yield better results than SVMs due to their ability to recognize complex patterns.

The fact that weather data proved not to improve the ETA predictions for the route and time-horizon under consideration, while the errors can be kept at relatively low levels with the other parameters, means that predictions of ETA for the route examined are possible without acquiring those data, something which reduces costs and complexity.

What is more, although the ETA of the ship agent improves significantly the prediction accuracy of the machine-learning techniques used, the algorithms still perform better than the current situation even on its absence as an input variable. This means that it is still possible to make acceptable predictions regarding the ETA of sea-vessels even when the ETA has not

been provided yet. This can be useful when trying to expand the time-horizon of predictions and the ETA of sea-vessels is not available.

On a personal level, I enjoyed the process of doing research in the field of maritime logistics. Working in a challenging environment within TNO, I had the opportunity to enhance my skills in machine learning and programming, while developing a business insight by exploring the value of the research for the relevant stakeholders. I found the project very engaging and the results obtained were rewarding for the effort. When starting with the thesis, I was expecting that the problem would be structured, with straightforward steps towards its realization. However, I soon realized that real-world cases present a lot of challenges that cannot be foreseen, and a lot of creativity is required in order to deal with them. I believe that I successfully responded to the challenge and learnt along the way. Also, cooperating with colleagues in data analytics within TNO was a valuable experience that I can use in the future.

7.3 Areas for further Research

The current study can serve as a starting point for many possible research areas in the field. A few of them are presented here.

Firstly, an interesting topic for further research would be to try and expand the time horizon for predictions, both for the specific route that was examined, as well as for the other routes. It is possible, that if other routes to port of Rotterdam are examined, such as voyages that travel over the Atlantic, the weather variables may have more impact on estimating the ETA compared to the case that was examined in this study. Also, if sufficient data are gathered, the predictions could expand to other vessel-types, such as tankers. The same models that were used in this case can be used for other vessel types with an extra identifier, specifying the vessel class. With enough data gathered from past voyages, it should be possible to train neural networks and support vector machines in predicting vessel arrival times given the same inputs as in the current study. Then, the accuracy of the models can be checked for different time horizons, to determine in which range they perform sufficiently.

Another proposed area for further research, is related to port operations with the aim of forecasting the waiting time of vessels when unloading at previous ports in Europe, such as Antwerp and Felixstowe, before going to the Port of Rotterdam. The uncertainty over the time that a vessel stays in those ports causes huge deviations in their original schedule of arrival at the Port of Rotterdam. This in turn causes changes in the berthing planning of the terminals, hindering their planning activities, as it was identified during the interviews conducted in the present study. The variables that would be of interest for this task are the berth utilization rates of the terminals, as well as the number of containers that the vessel is expected to unload. In the present study, vessels that were stopping at the port of Gibraltar for bunkering were also taken into account. In these cases, the SVMs and Neural Networks, were recognizing that the vessels are in the bunkering area of Gibraltar, through the longitude and

latitude variables, and were adding a waiting time. This can serve as a starting point for the proposed area for further research in Antwerp and Felixstowe.

Also, estimating the waiting time due to Port operations in the Port of Rotterdam, would be of value for the hinterland transportation parties. However, if the ETA information tool is implemented, thus reducing the uncertainty over vessel arrivals, this may have a secondary effect on reducing handling times at the Port. In that case, estimating port operations time in the Port of Rotterdam in the current context may be premature.

Last but not least, further research is proposed for the realization of a decision support system for vessel-speed planning. The proposed system would be estimating the weather impact on the vessel's speed along its route to the Port of Rotterdam. That way, the captains could plan on keeping a relatively steady engine power, that would guarantee on-time arrival at the port, while saving on fuel consumption. Two different approaches can be followed for the realization of such a decision support system. The first, based on machine learning techniques would be to provide as input the rounds per minute of the ship's engine, alongside with the weather conditions in the area, with the aim of predicting the speed over ground with which the vessel will be travelling. Then, aggregating over the whole route until Port of Rotterdam, the weather effect on the vessel's speed can be obtained. The other way, would be to model the weather impact on the ship's nominal speed (based on the rpms of the engine) using a physics-based approach. Such an approach could be undertaken by the company Hermess, due to their expertise in oceanography. In the present study, developing this decision support system was not possible due to absence of information regarding the ship engines.

Overview of Chapter 7

In this final chapter, the academic findings of the study were summarized. The improvement in predictions in the long time-horizon was highlighted, alongside with the benefits that can be realized for the interested parties. The answers to the main research question and sub-questions posed at the beginning of the report were also summarized in section 7.1.2. The added value for the terminals, hinterland transportation parties and carriers in their planning activities was also explained. Special attention was given to how the Port of Rotterdam can implement the proposed system, since adoption of the ETA information tool would enhance the competitive position of the port.

Moreover, further research into three different directions was proposed. The first is that of a vessel speed-planning decision support system, in order to minimize fuel consumption for the shippers, while still delivering the goods in time. Secondly, investigating the possibility of predicting ship waiting times due to Port operations and last but not least, an expansion to the time horizon of the ETA predictions to more than 5 days away from the port.

References

Bakas, T., & Crainic, T. G. (2007). A Brief Overview of Intermodal Transportation. Quebec: CIRRELT.

Chapelle, O., & Vapnik, V. (1999). Model Selection for Support Vector Machines. *Advances in Neural Information Processing Systems*, Vol 12.

Coates, A., & NG, A. (2012). *Learning Feature Representations with K-means.* Stanford CA 94306, USA: Stanford University.

De Boer, T. M. (2010). An analysis of vessel behaviour based on AIS data. *Master Thesis TU Delft*.

Fancello, G., Pani, C., Pisano, M., Serra, P., Zuddas, P., & Fadda, P. (2011). Prediction of arrival times and human resources allocation for container terminal. *Maritime Economics & Logistics 13(2)*, pp. 142-173.

Frawley, W., Piatetsky-Shapiro, G., & Matheus, C. J. (1992). Knowledge discovery in databases: an overview. *AI Magazine 13(3)*, pp. 57-70.

Freeman, D. A. (2005). Statistical Models Theory and Practice. Cambridge University Press.

Gambarella, L. M., Rizzoli, A. E., & Zaffalon, M. (1998). Simulation and planning of an intermodal container terminal. *Simulation 71(2)*, pp. 107-116.

Hakimi, B., Montreuil, D., Sarraj, R., Ballot, E., & Pan, S. (2012). Simulating a Physical Internet mobility web: the case of mass distribution in France. *International conference of Modeling, Optimization and Simulation, Bordeaux, France.* Bordeaux.

Haykin, S. (2008). *Neural Networks and Learning Machines.*

Heaton, J. (2008). *Introduction to Neural Networks with Java, Second Edition.* St. Louis: Heaton Research , Inc.

Heij, R. (2015). *Opportunities for peak-shaving consumption at container Terminals.* Delft: TU Delft.

Hermess. (2013). *Hermess website*. Retrieved July 2016, from http://www.hermess.nl/

Intertransis. (2013). *Intertransis Website*. Retrieved July 2016, from Intertransis Website: http://www.intertransis.com/

Kemme, N. (2013). *Design and operation of automated container storage systems.* Heidelberg: Springer.

Loptien, U., & Axell, L. (2014, July 14). Ice and AIS: ship speed data and sea ice forecasts in the Baltic Sea. *The Cryosphere*, pp. 2409-2418.

Menger, I. (2016). *Information Exchange between deep sea container terminals and hinterland parties.* Delft: TU Delft.

Mokhtari, K., Ren, J., Roberts, C., & Wang, J. (2012). Decision support framework for risk management on sea ports and terminals using fuzzy set theory and evidential reasoning approach. *Expert Systems with Applications 39(5)*, 5087-5103.

Murty, K. G., Wan, Y. -W., & Linn, R. (2005). A decision support system for operations in a container terminal. *Decision Support Systems 39(30)*, pp. 309-332.

Port of Rotterdam. (2014). Retrieved June 8, 2016, from Port of Rotterdam web site: https://www.portofrotterdam.com/sites/default/files/Modal%20split%20maritime%20containers%202014%20-%202011.pdf

Port of Rotterdam. (2014). *Annual Report.* Rotterdam.

Port of Rotterdam. (2016). *Port Statistics.* Rotterdam: Port of Rotterdam.

Römers, I. (2013). *Port Call Optimization in three oil shipping markets.* Rotterdam: Erasmus University Rotterdam.

Schölkopf, B., & Smola, A. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond.* MIT Press.

Stopford, M. (2009). *Maritime Economics 3rd Edition.* New York: Routledge.

Szelangiewicz, T., Wiśniewski, B., & Żelazny, K. (2014). The influence of wind, wave and loading condition on total resistance and speed of the vessel. *POLISH MARITIME RESEARCH 3(83)*, pp. 61-67.

Tavasszy, L. A., & de Jong, G. (2013). Modelling Freight Transport. *Elsevier*.

TNO. (2015). *TNO website*. Retrieved July 2016, from https://www.tno.nl/en/about-tno/mission-and-strategy/

UNCTAD. (2007). Review of Maritime Transport. *United Nations Conference on Trade and Development.* New York.

UNCTAD. (2013). Review of Maritime Transport. *United Nations Conference on Trade and Development.* New York.

United Nations. (2014). *Review of Maritime Transport.* United Nations, UNCTAD.

van Riesen, B. (2013). *Planning of hinterland transportation in the EGS network.* Delft: TU Delft.

Vis, I. F., & De Koster, R. (2003). Transshipment of containers at a container terminal: an overview. *European Journal of Operational Research 147(1)*, pp. 1-16.

Voss, S., Stahlbock, R., & Steenken, D. (2004). Container terminal operation and operations research - a classification and literature review. *OR Spectr 26(1)*, 3-49.

Wanders, G. (2014). *Determining shippers' attribute preference for container transport on the Rotterdam-Venlo Corridor.* Delft: TU Delft.

Wiering, M. A., van der Ree, M., Embrechts, M. J., Stollenga, M. F., Meijster, A., Nolte, A., et al. (2013). *The Neural Support Vector Machine.*

Wright, R. (2016, May 19). *Container shipping lines mired in crisis*. Retrieved June 1, 2016, from Financial Times: http://www.ft.com/intl/cms/s/0/1e98963c-1853-11e6-bb7d-ee563a5a1cc1.html#axzz4AoPLqw1A

# Appendix A- Parameter Selection for Neural Networks and SVMs

The following figures present the training, validation and Testing Errors for the neural network approach, for different choices of the number of neurons in the hidden layer. It can be noticed that the usage of the regularization parameter is negating the effect of choosing the number of neurons and the results obtained from using a slightly different number of neurons than the original case (7 neurons), do not deviate significantly. Nevertheless, in all of the cases, the validation error is not lower than in the original case of 7 neurons for the different values of the regularization parameter lambda. Therefore, using a neural network consisting of 7 neurons in its hidden layer is the optimal choice.
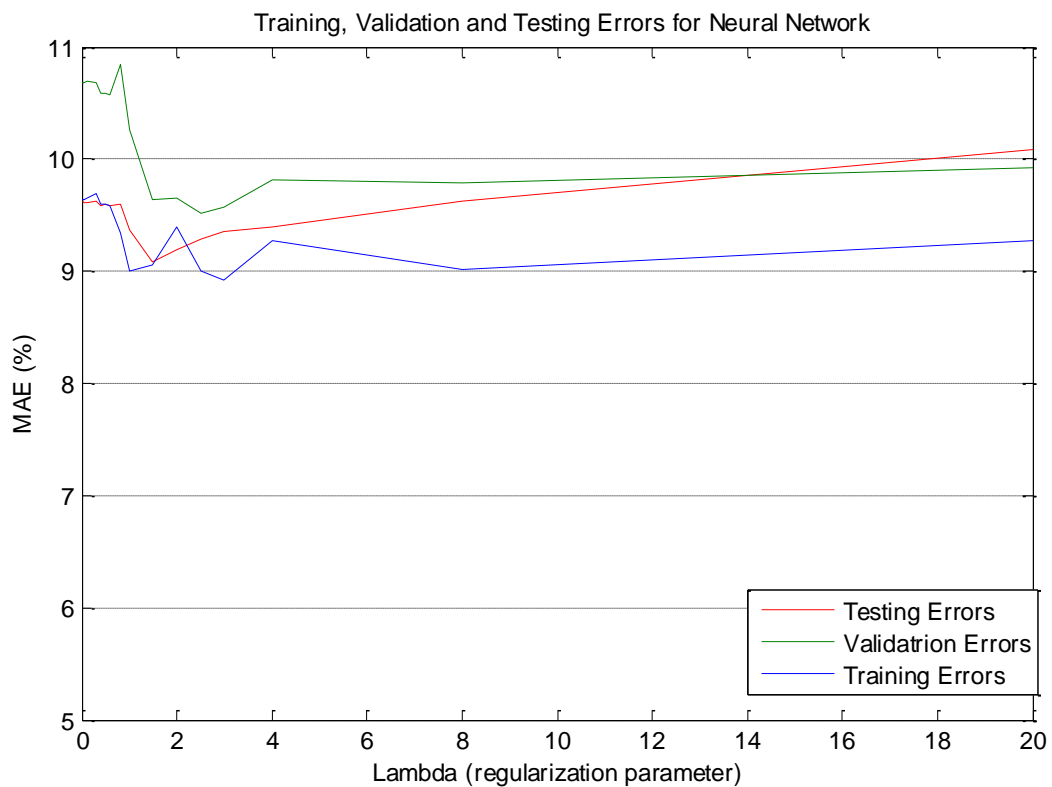
Figure 22: *Training, validation and Testing Errors for different values of the regularization parameter lambda. For lambda=2.5, the validation error becomes minimum and thus, lambda=2.5 is chosen for the training of the optimal neural network.*
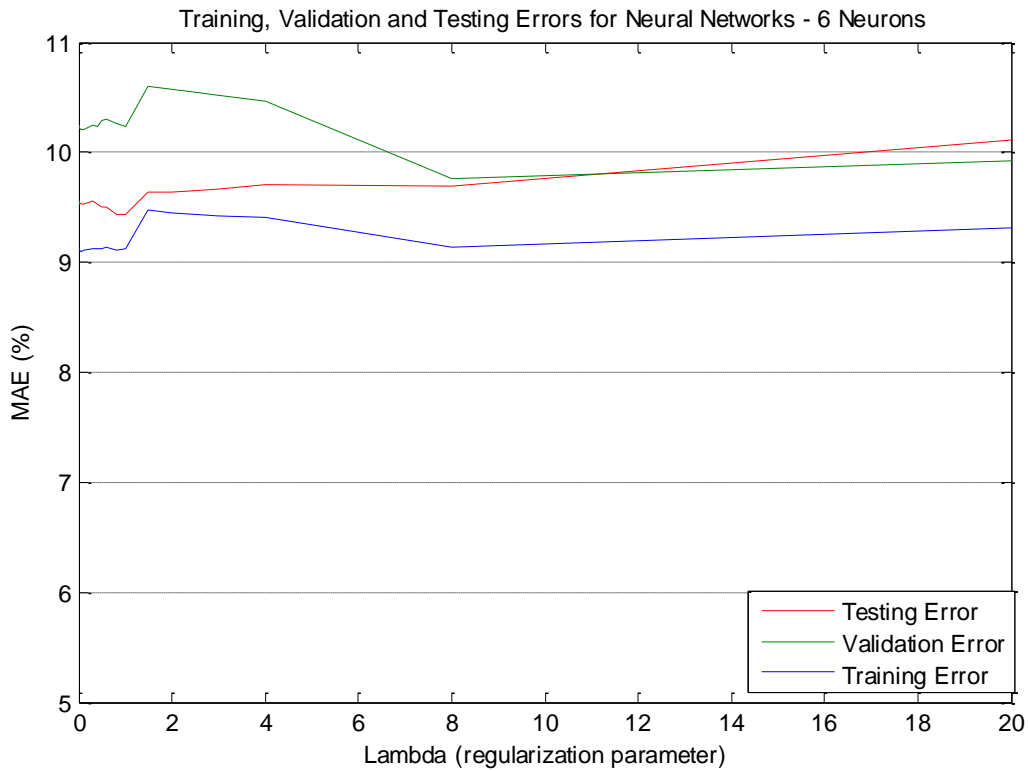
Figure 23: *Training, validation and testing errors for a neural network with 6 neurons in its hidden layer. The validation error is higher than in the case of the original case of 7 neurons, for all the values of the regularization parameter lambda.*
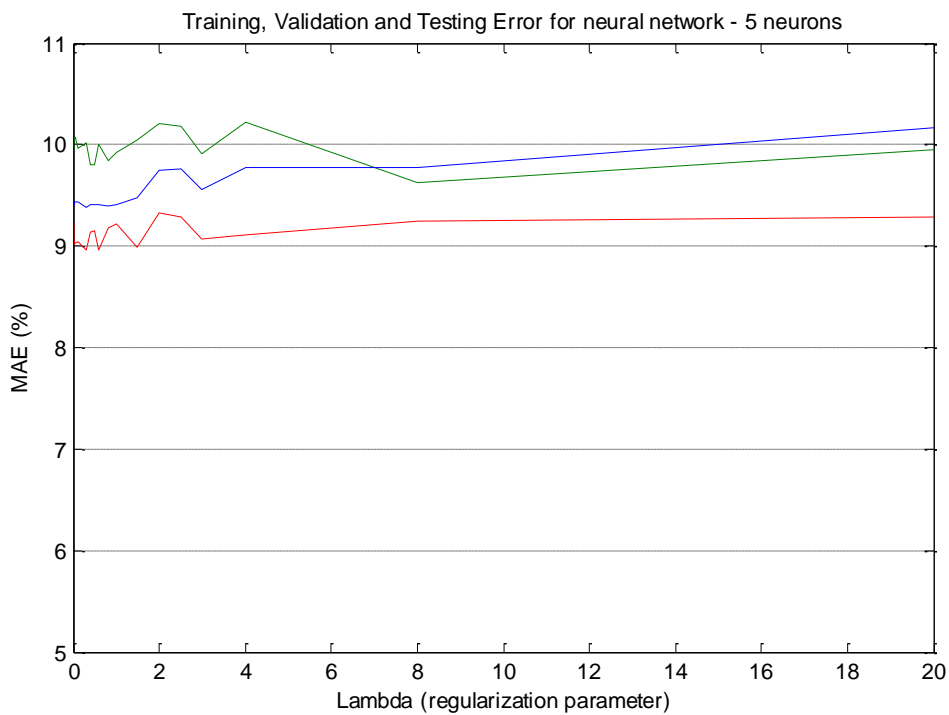
**Figure 24:** *Training, validation and testing errors for a neural network with 5 neurons in its hidden layer*
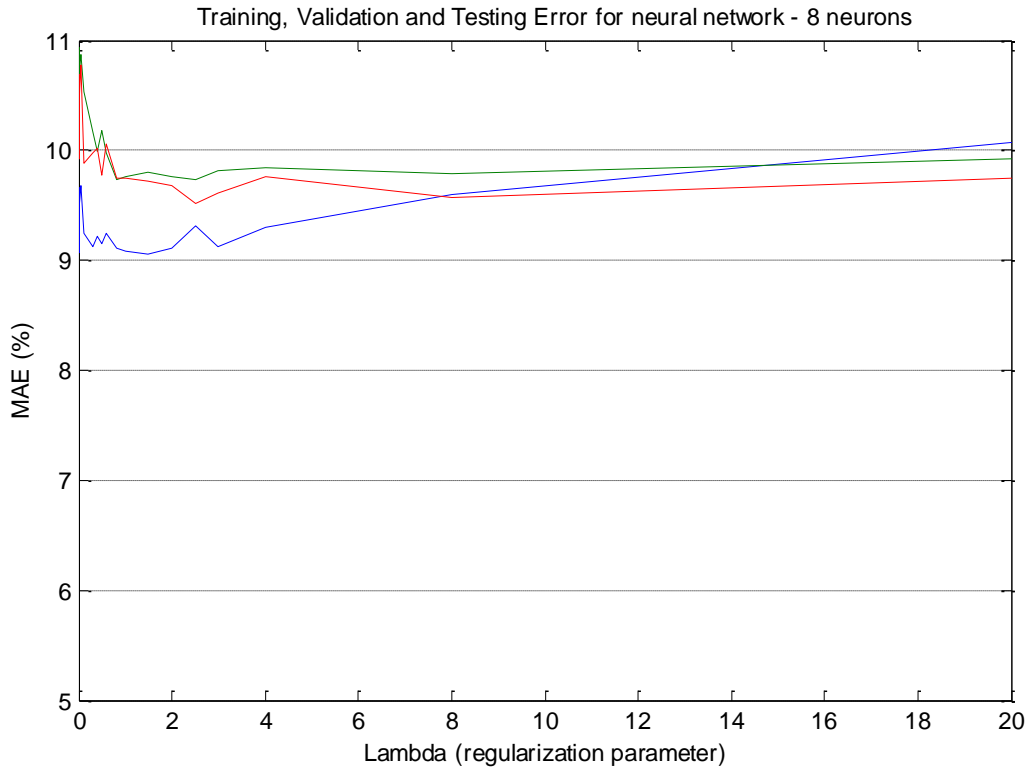
**Figure 25:** *Training, validation and testing errors for a neural network with 8 neurons in its hidden layer*
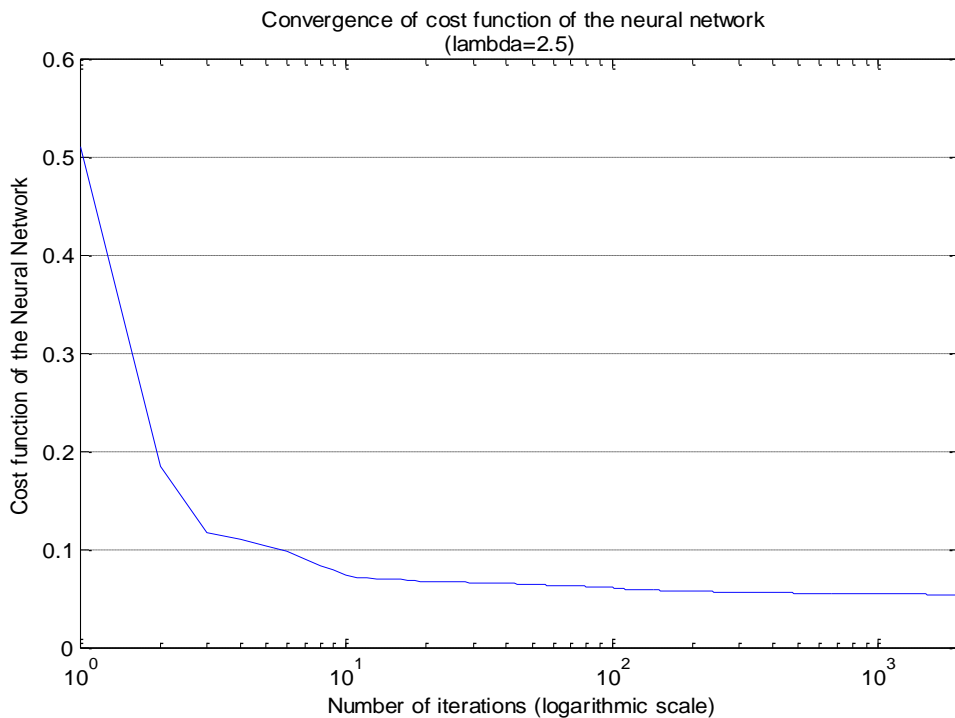
Figure 26: *Convergence of the cost function of the Neural Network through the process of gradient descent. As it can be noticed, the function converges very quickly to its minimal value after the first iterations.*

Selection of the optimal set of parameters C and gamma for the SVM approach. Different values of the regularization parameters were tried. For each one of them, the validation error was calculated for a number of values of the parameter gamma in the Gaussian Kernel. The optimal set of parameters is the point where the validation error is minimized and is obtained for C=100 and gamma=0.001 .
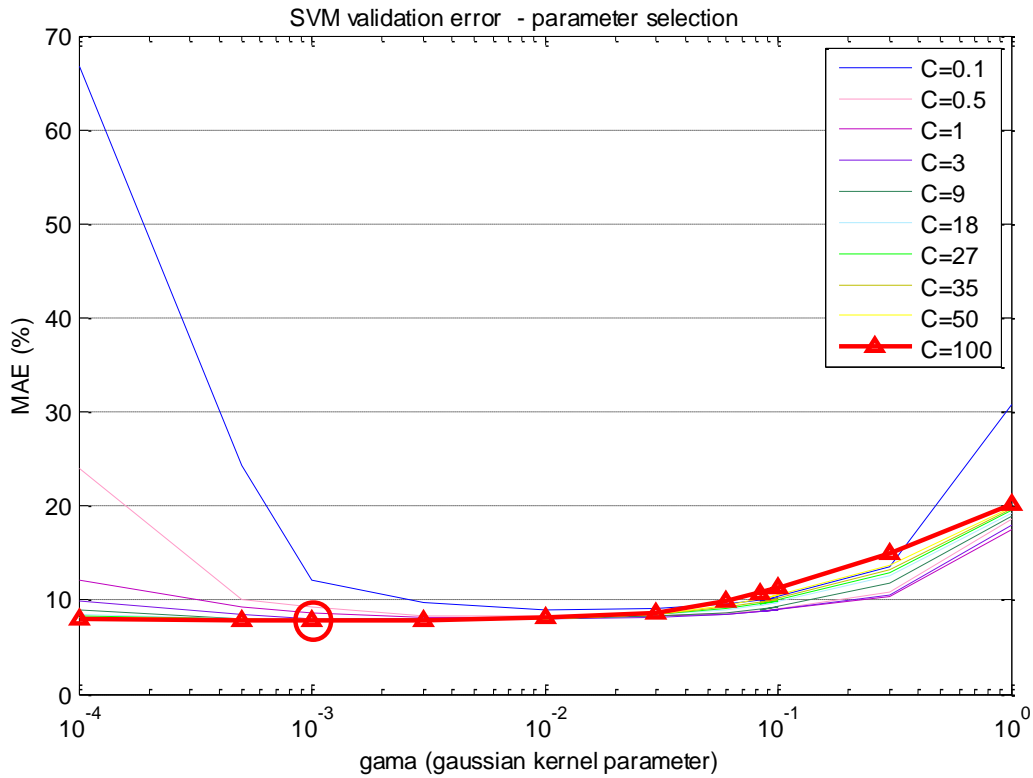
Figure 27: *SVM validation error for different values of the regularization parameter C and the Kernel function parameter gamma. Optimal choice of parameters for C=100 and gamma=0.001*
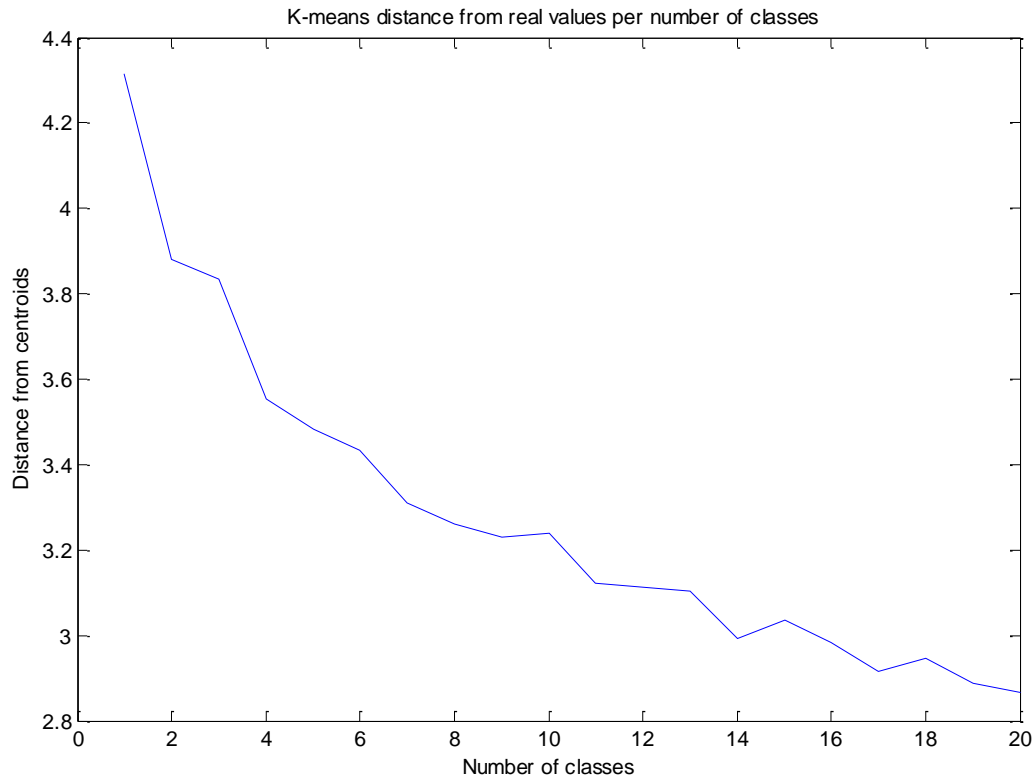
Figure 28*: Distance of weather condition cluster centroids to the weather variables. The most steep decline is until the 5 weather classes, which was the value chosen for the number of clusters.*

Neuron Number	Weight Assigned
0*	3,07
1	0,50
2	0,46
3	0,46
4	0,47
5	-0,46
6	-0,49
7	-0,46

Table 5: *Weights Assigned to the neurons connecting the hidden layer to the output of the neural network. *Neuron number 0 stands for the bias unit.*

Appendix B – Selection of validation and testing set size

Different sizes for the validation and testing sizes were tried to determine the sensitivity of the results on the choice of the sets. It was identified that the results are stable for changes in the validation set, since they do not deviate significantly from the original case (validation set 15% - testing set 20%). For lower testing sets however, the number of testing examples were not enough for the areas further away from the Port of Rotterdam and very low or high training errors were obtained. Therefore, the testing set should be no lower than 20%.
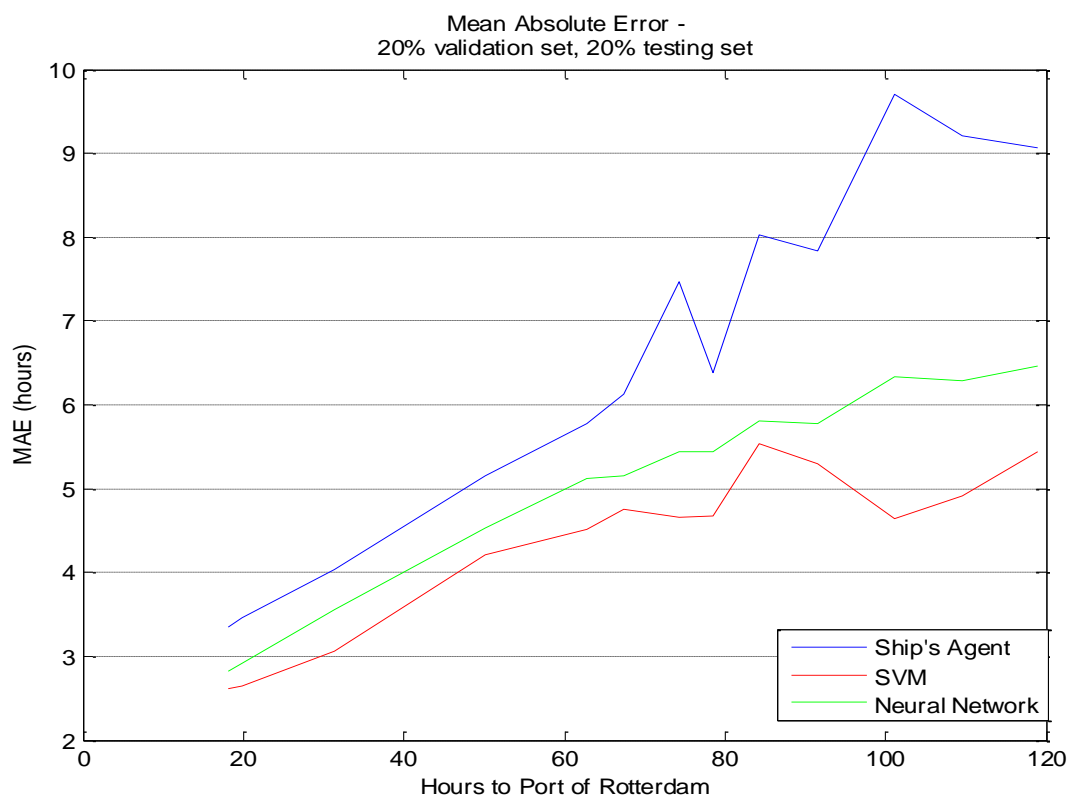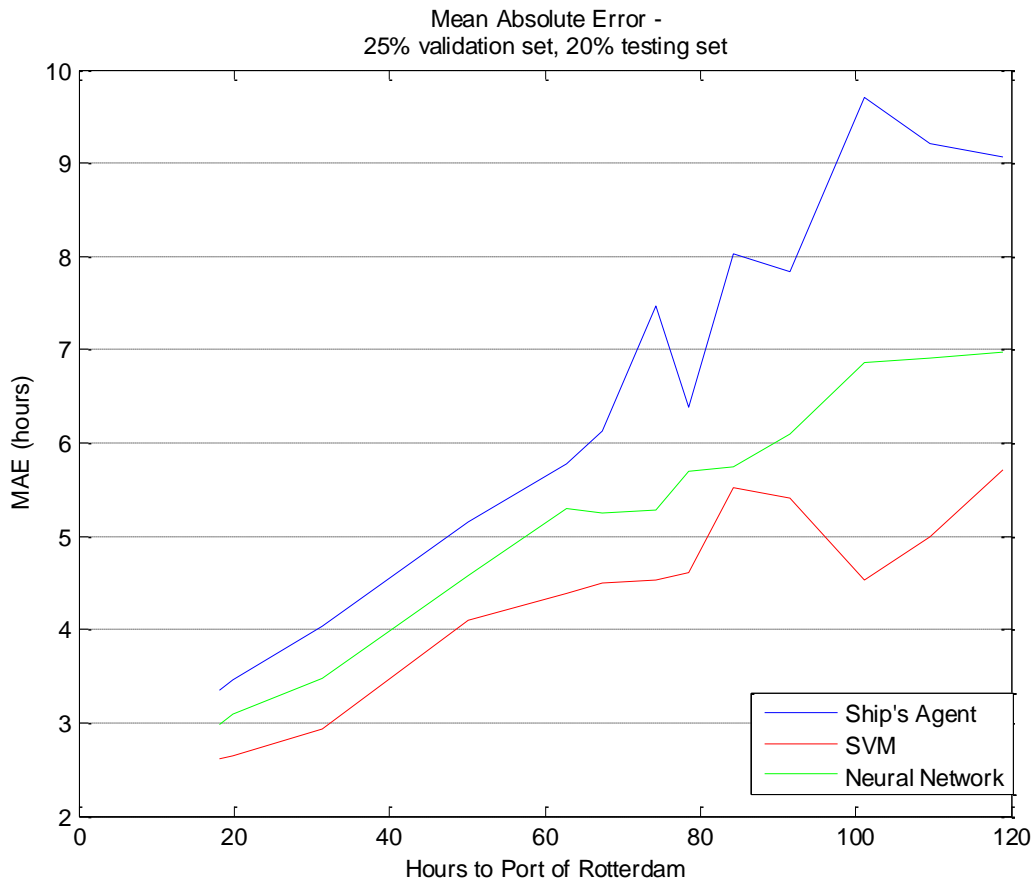

Figure 29: *Neural Network and SVM Mean Absolute errors for 20% validation and 20% Testing set. The results obtained are very similar to the initial case.*
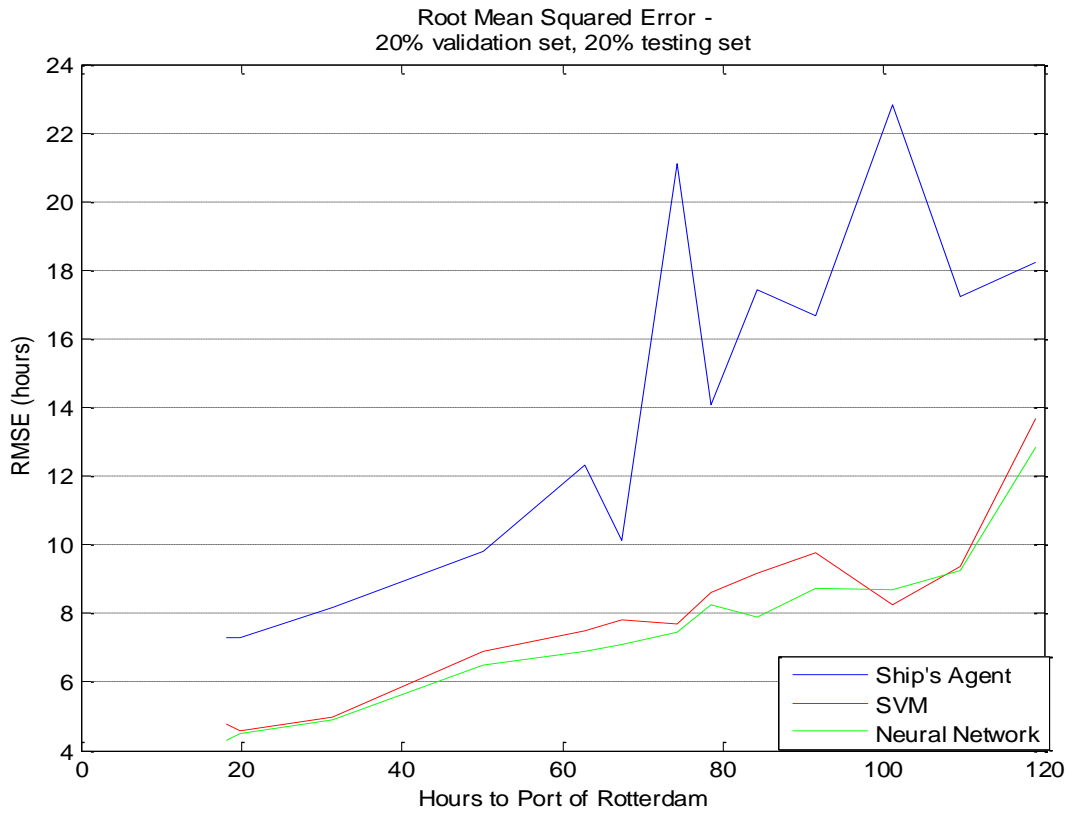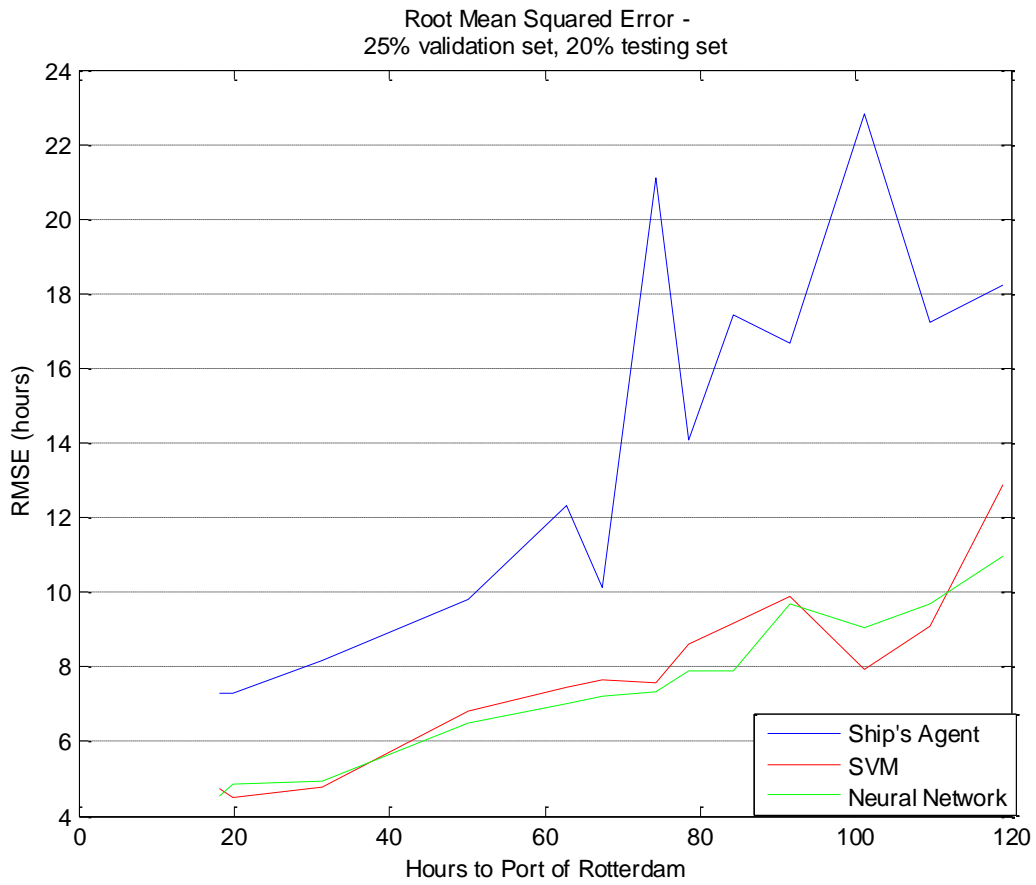
Figure 30: *Neural Network and SVM Root Mean squared errors for 20% validation and 20% Testing set. The results obtained are very similar to the initial case.*

Figure 31: *Neural Network and SVM Mean Absolute errors for 25% validation and 20% Testing set. The results obtained are very similar to the initial case.*

Figure 32: *Neural Network and SVM Root Mean Squared errors for 25% validation and 20% Testing set. The results obtained are very similar to the initial case.*

Appendix C - Gibraltar Pilot ETA predictions

In this part the pilot project of estimating the ETA from Gibraltar to the port of Rotterdam is presented. The first graph shows the mean absolute error without adding a weather variable and ETA of the ship's agent. In the second graph, the weather variable is included and in the third graph, both the weather and the ETA variables as introduced as inputs to the neural network.
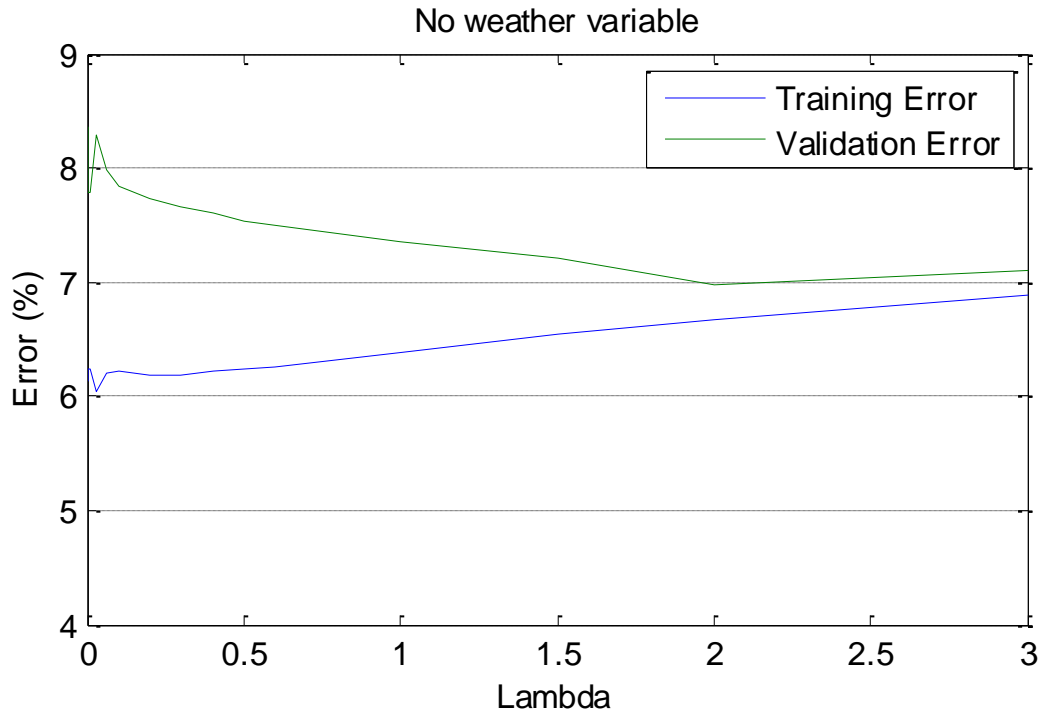
Figure 33: *Training and validation errors for the neural network approach without an ETA or a weather variable as input. Optimal value for the regularization parameter lambda is at lambda=2, with a validation error of 7%.*
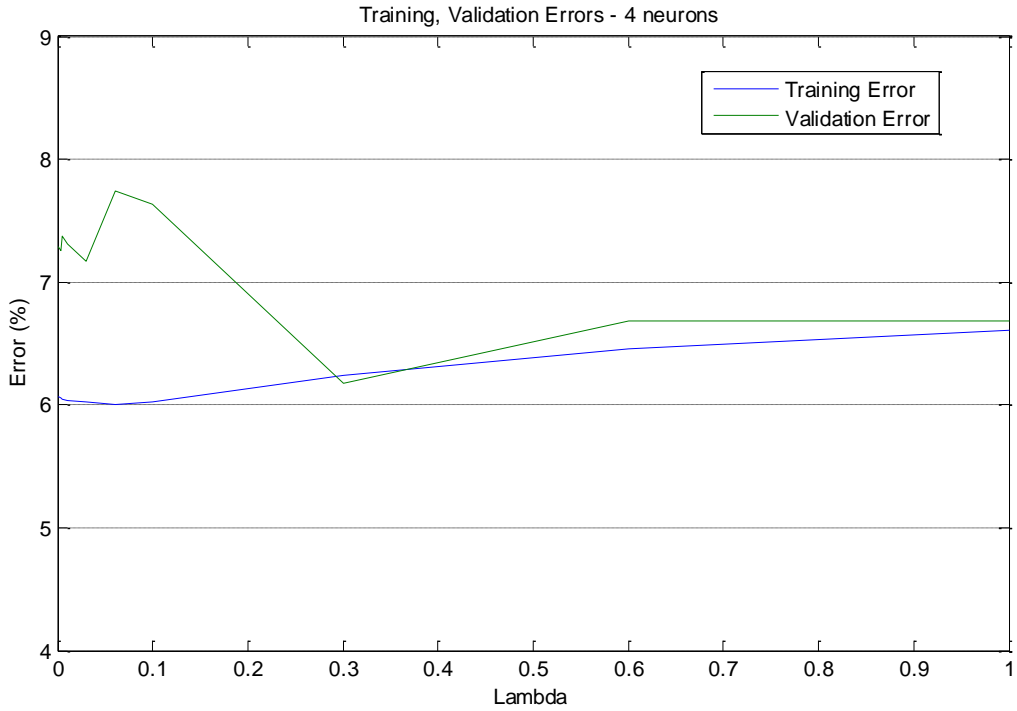
Figure 34: *Training and validation errors for the neural network approach without an ETA, but with a weather variable as input. Optimal value for the regularization parameter lambda is at lambda=0.3, with a validation error of 6,2%.*
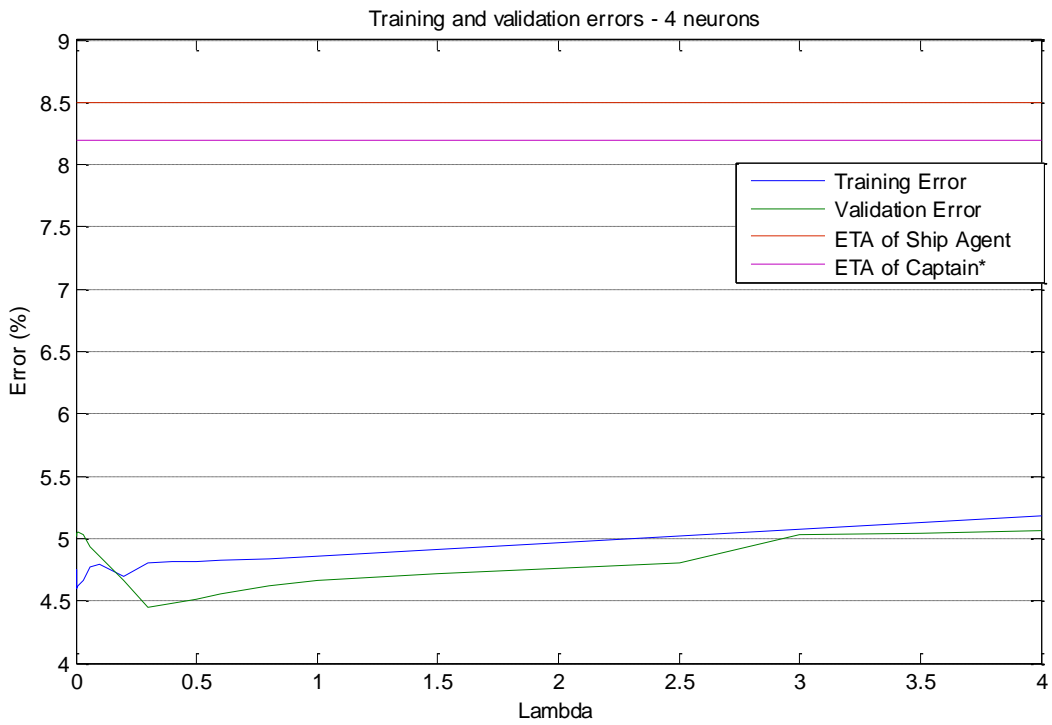
Figure 35: *Training and validation errors for the neural network approach with ETA and weather as input. The results are significantly improved compared to the ETA of the ship's agent.*

The following table presents the error metrics for the different methods used for the case of Gibraltar:

Error Metrics	SVM	Neural Network	ETA of Captain	ETA of ship's agent
MAE $(\frac{ATA-ETA}{ATA} * 100\ \%)$	4,7 %	4,5 %	8,2 %	8,5 %
RMSE (days) $(\sqrt{\frac{\sum(ATA-ETA)^2}{n}}\ )$	0,26	0,33	0,68	0,71

Table 6: *Error metrics for the SVM and Neural network approach, compared to the current situation based on the ETA of the ship agent.*

Appendix D – Interviews

Interview questions

Ed van Dort (Managing Partner at Intertransis):

1) How is currently the Port of Rotterdam making use of the ETA provided by the captain of the ship? Is this information being used to achieve more efficient planning?
2) What are the effects that an early/late arrival of a ship has on the port operation activities?
3) What would the benefit of more accurate ETA predictions be for resolving problems related to Port-Operating activities?
4) What is the level of accuracy required for the ETA of sea-vessels in order to improve the efficiency at the Port?
5) Can you think of other parties that might be interested in such an information system?

Siem van Marriënboer (Senior consultant in Sustainable transport and Logistics at TNO)

1) Is currently information regarding the ETA of sea vessels being used for the on-line planning of hinterland transportation?
2) How would the information regarding the ETA of sea vessels be used in order to improve the hinterland transportation activities in terms of cost and time?
3) What is the accuracy of ETA prediction and the time-window required for the purposes of improving hinterland transportation activities?
4) How would the split between truck/barge and rail as modes of transport, be affected through the integration of ETA prediction in the daily planning activities of hinterland transportation?
5) Which other parties would be interested/affected by such an information tool?

Johan Hoekwater (Manager Logistics Development, Europe Container Terminals (ECT))

1) How is currently ECT making use of the ETA provided by the captain of the ship? Is this information being used to achieve more efficient planning? (Do you know cargo or #number of containers in a ship?)
2) What are the effects that an early/late arrival of a ship has on container terminal activities?
3) What would the benefit of more accurate ETA predictions be for resolving problems related to Container Terminal-Operating activities?
4) What is the level of accuracy required for the ETA of sea-vessels in order to have an impact on container terminal operations?
5) Can you think of other parties that might be interested in such an information system?

- So, you would be interested in a system predicting how long vessels stay at a port in continental Europe? (Do we know in advance if they are going to call in another port first?)
- Which are these ports? (Antwerp, Hamburg or Amsterdam?) Does Port of Gibraltar count as such a port?
- How many are the ships arriving directly (1st port call in Rotterdam) and how many call first in another port?
- What delays are we talking about and how much in advance do you adjust your quay planning?

Benefits of better ETA:

➢ Solving the berthing allocation problem -> Less waiting time for ships (carriers value this service)
➢ Optimize resources allocation (personnel, equipment), also energy saving (cranes, quay to yard transport vehicles) -> reduction in costs
➢ Maintenance schedule planning

Johan Hoekwater(ECT Logistics Manager) findings:

- Static schedule for a whole year. ECT is using the information provided by the ship's Agent to dynamically update the ship's ETA and schedule its operating activities.
- Suez canal has specific time slots, so this means that the vessels are entering and leaving at specific times. Then the ships sail at slow steam to save fuel, usually. The captains try to stick to their ETA. This results in reduced uncertainty regarding the arrival at the PoR.
- Major delays are caused because of Port Operations, however at that time the container terminal is in communication with the vessel and thus updated.
- Berth allocation is a significant problem with huge impact on waiting time for sea-vessels. If a vessel diverges from original schedule, it is assigned in other berthing place. The cargo in the stock goods has to be removed first from the loading position. Then it has to be moved to the new berthing place. This results in increased workload, unnecessary moves and longer waiting time for the vessel.
- Early arrivals are reduced by slowing down and saving on fuel.
- Early arrival is difficult, a lot of waiting time for the carrier. Containers have to be moved from the yard to a position ready to be used. One carrier is continuously arriving early.
- 1 week in advance for the ETA is the scope of interest for ECT.
- Other interested stakeholders: Barge, rail, truck operators, freight forwarders such as DHL. They may need information earlier, due to their planning activities for large time-windows. They get ETA of Ships from ECT, so the ship's agent ETA.
- 50% direct arrivals from Asia, 50% of ships call first to another European Port. Big delays in the second case.
- Need to forecast the waiting time at previous port. Maybe Pilot at the Port of Antwerp. Affecting variables: DWT, ship length, ship breadth, draught, traffic at the Port.
  ECT is trying to take into account these delays in the previous port. The schedule of the ships is fixed, so they know which ships will go to Port of Antwerp in advance.

When vessels stop in many ports before PoR, there is huge uncertainty regarding their arrival at the Port of Rotterdam. For instance when they stop in Antwerp, then in Felixstowe, this introduces a lot of uncertainty regarding vessel arrival at the Port of Rotterdam.

Ed van Dort (Managing Partner at Intertransis) findings:

72 hours mandatory announcement of arrival at the Port

Captain gives information to the ship's agent about the ETA, then he sends it to the Port Authority and they give the information to ECT. Shipper has a problem with early arrival. The containers that he has to load may not be there.

If berth occupation rate is high, then divergence from schedule of arrivals can prove to be a challenge for the planning of terminal operators. Barges also have the same berthing places as sea-vessels so divergence of the sea-vessel schedule causes problems in barge allocation, hindering the planning activities and hampering the fast delivery to inland destinations. Waiting time added.

EGS is mainly interested in 5 days in advance to 8 hours before vessel arrivals, and probably even earlier. This would allow them to plan their hinterland transportation activities.

Slow steam versus arriving on time by captains. Sometimes captains speed up to ensure on time arrival and then they have to slow down (to avoid early arrival at the port). These fluctuations in speed result in higher fuel consumptions, which comprise a big portion of a ship's voyage expenses. An area of investigation would be, that we can guarantee by keeping a certain level of speed they can reach the port, due to weather conditions, so that they can navigate by that constant speed.

Sometimes a captain can even skip a port to ensure on time arrival. However this has a lot of extra expenses, because the containers would then have to be transported by inland waterways or hinterland to the previous destination.

Other interested parties in ETA information tool: Stacks, pilots, port authority interested in the part of 8 hours in advance, for port operations. Pilots are assisting vessels that help the ship enter the port and transfer people on and off board.

Siem van Marriënboer (Senior consultant in Sustainable transport and Logistics at TNO) findings:

BCTN (like EGS): Responsible for hinterland transportation between ECT deep water terminals and Venlo corridor.

First they try to assign barges to transport the incoming containers to inland destinations and then trucks. The barges have a fixed schedule according to the planning of BCTN. Delays of sea-vessels can cause barges to sail relatively empty.

Uncertainty of ship vessels, if it can be reduced to 4 hours it would be helpful for the hinterland transportation activities.

2 to 72 hours takes for a ship to be unloaded in the PoR -> 2 to 72 hours until commercial release of container. Commercial release is when the container will become available for hinterland transportation, after being unloaded from the ship. ECT holds a model developed by TNO that knowing the position of the container within a ship, tries to predict the time for its commercial release.

BCTN would like to receive much more accurate information regarding the commercial release of container. Therefore, they are currently making use of ship position data to track the vessel and estimate when it will be arriving at the port. Prediction of barge arrival 14-18 hours usually, with a variance of 10-52 hours depending on previous stops. 20 minutes for last haulage by truck. The goal is to reduce uncertainty within the whole hinterland transportation process. ETA prediction is the 1st step in the process.

LSP books BCTN to transport the container hinterland. They schedule for a barge at a specific time slot. If the vessel is not there on time, the container will not be on time in the barge, causing delay or less cargo consolidation on barge. The rest will be transported by trucks.

Appendix E

In this part, a calculation of fuel savings in the case of a vessel keeping a steady speed of 12.5 knots over the whole route, compared to the case of travelling at 15 knots for the first half the voyage (in time measures) and at 10 knots for the rest of the route is carried out. The purpose is to show the potential in fuel saving with a decision support system in speed planning of the vessel.

$$Fuel\ consumption\ for\ steady\ speed\ =\ A + B * 12.5^3$$

$$Fuel\ consumption\ for\ changing\ speed\ = (A + B * 15^3) * \frac{1}{2} + (A + B * 10^3) * \frac{1}{2}$$

$$= A + B * (15^3 + 10^3) * \frac{1}{2}$$

Savings in fuel consumption (%)=$\dfrac{Fuel\ consumption\ for\ changing\ speed - Fuel\ consumption\ for\ steady\ speed}{Fuel\ consumption\ for\ changing\ speed}$

$\Rightarrow$ Savings in fuel consumption (%) $= \dfrac{B*[(15^3+10^3)*\frac{1}{2}-12.5^3]}{A+B*(15^3+10^3)*\frac{1}{2}} \cong 6{,}7\%$

Note: A is a constant of relatively small magnitude (Bialystocki & Konovessis, 2016), negligible when compared to the speeds in the third power, therefore for the purposes of the analysis it can be neglected.

Number of voyages following the Asia-Rotterdam route:

In the dataset used, 600 voyages were found to follow the Asia-Rotterdam route for unloading containers for the year 2015, accounting only for ECT terminals. The total number of port calls from containerships was 7398. This however includes empty (ballast) vessels or vessels approaching for maintenance. Assuming that 30% of the vessels fall into these categories, the loaded containerships are estimated at 5178. Therefore the ECT vessels examined, account for 12% of the total port calls. If the other terminals are taken into account, a 25% of port calls originating from the Asia-Rotterdam route seems reasonable.